150 MOST FREQUENTLY ASKED QUESTIONS ON QUANT INTERVIEWS

DAN STEFANICA
RADOŠ RADOIČIĆ

TAI-HO WANG

Baruch College
City University of New York

FE Press
New York

FE PRESS
New York

www.fepress.org

This edition first published 2013

Printed in the United States of America

Cover art and design by Max Rumyantsev

To our beloved families

Contents

Preface

The use of quantitative methods and programming skills in all areas of finance, from trading to risk management has grown tremendously in recent years, and accelerated through the financial crisis and with the advent of the big data era.

A core body of knowledge is required for successfully interviewing for a quant type position. The challenge lies in the fact that this knowledge encompasses finance, programming (in particular C++ programming), and several areas of mathematics (probability and stochastic calculus, numerical methods, linear algebra, and advanced calculus). Moreover, brainteasers are often asked to probe the ingenuity of candidates.

This book contains over 150 questions covering this core body of knowledge, without which it is not possible to advance to a final interview round. These questions are not only frequently, but also currently, asked on interviews for quantitative positions, and cover a vast spectrum, from C++ and data structures, to finance, brainteasers, and stochastic calculus.

The answers to all of these questions are included in the book. These answers are written in the same very practical vein that was used to select the questions: they are complete, but straight to the point – as they would be given in an interview.

Topics:

- Mathematics, calculus, differential equations.
- Covariance and correlation matrices. Linear algebra.
- Financial instruments: options, bonds, swaps, forwards, futures.
- C++, algorithms, data structures.
- Monte Carlo simulations. Numerical methods.
- Probability. Stochastic calculus.
- Brainteasers.

The authors are faculty members of the Baruch College Financial Engineering Masters Program, and have over 20 years of experience educating students who were very successful interviewing for quantitative positions. As such, the authors had the privilege to interact with generations of exceptional students, whose contributions as alumni to the continued success of our students has been tremendous. This book is a tribute to our special Baruch MFE community.

This is the first book in the *Pocket Book Guides for Quant Interviews* Series, to be followed by books on advanced probability and stochastic calculus questions and on challenging brainteasers asked in quant interviews.

New York, 2013

Acknowledgments

As professors in the Financial Engineering Masters Program at Baruch College, it has been a privilege to have the opportunity to contribute to the early career success of so many talented students and alumni. The strong community that developed around the Baruch MFE program is the reward educators truly dream of, and, for us, a reality that is an inspiration. This book would have not been possible without our Baruch MFE community – we are grateful to everyone who is part of it.

Working alongside and learning from our colleagues from the mathematics department and from the financial industry created the perfect context for writing this book; we are thankful to all of them.

Alejandro Cañete is owed a special thank you for his wonderful help on the programming questions, which are much improved due to him. The art for the book cover is, as always, due to the professional help of our alum and friend Max Rumyantsev. Several students spearheaded the proofreading effort, and their help was greatly appreciated: Yu Gan, Jun Hao, Alireza Kashef, Yi Bill Lu, Svetlana Rafailova, Fubo Shi, Yujia Helen Sun, Jun Charlotte Wang, Peng Wu, Yanzhu Wendy Wu, Yongyi Ivan Ye, He Hillary Zhao, and, last but not least, Wenyi Zhou.

It is the tremendous support and the understanding of our families that made this book possible, and we are

forever in debt to them. Thank you, thank you, and thank you!

This book is dedicated to our wives and to our children, from all our hearts.

Dan Stefanica
Radoš Radoičić
Tai-Ho Wang

New York, 2013

Chapter 1

First Look: Ten Questions.

1. Put options with strikes 30 and 20 on the same underlying asset and with the same maturity are trading for \$6 and \$4, respectively. Can you find an arbitrage?

2. The number 2^{29} has 9 digits, all different. Without computing 2^{29}, find the missing digit.

3. Let W_t be a Wiener process, and let

   $$X_t = \int_0^t W_\tau \, d\tau.$$

   What is the distribution of X_t? Is X_t a martingale?

4. Alice and Bob stand at opposite ends of a straight line segment. Bob sends 50 ants towards Alice, one after another. Alice sends 20 ants towards Bob. All ants travel along the straight line segment. Whenever two ants collide, they simply bounce back and start traveling in the opposite direction. How many

1

ants reach Bob and how many ants reach Alice? How many ant collisions take place?

5. Find all the values of ρ such that

$$\begin{pmatrix} 1 & 0.6 & -0.3 \\ 0.6 & 1 & \rho \\ -0.3 & \rho & 1 \end{pmatrix}$$

is a correlation matrix.

6. How many independent random variables uniformly distributed on $[0,1]$ should you generate to ensure that there is at least one between 0.70 and 0.72 with probability 95%?

7. Show that the probability density function of the standard normal integrates to 1.

8. Assume the Earth is perfectly spherical and you are standing somewhere on its surface. You travel exactly 1 mile south, then 1 mile east, then 1 mile north. Surprisingly, you find yourself back at the starting point. If you are not at the North Pole, where can you possibly be?!

9. Solve the Ornstein-Uhlenbeck SDE

$$dr_t = \lambda(\theta - r_t)dt + \sigma dW_t,$$

which is used, e.g., in the Vasicek model for interest rates.

10. Write a C++ function that computes the n-th Fibonacci number.

Solutions

Question 1. Put options with strikes 30 and 20 on the same underlying asset and with the same maturity are trading for \$6 and \$4, respectively. Can you find an arbitrage?

Answer: Since the value of a put option with strike 0 is \$0, we in fact know the prices of put options with three different strikes, i.e.,

$$P(30) = 6; \ P(20) = 4; \ P(0) = 0,$$

where $P(K)$ denotes the value of a put option with strike K.

In the plane $(K, P(K))$, these option values correspond to the points $(30, 6)$, $(20, 4)$, and $(0, 0)$, which are on the line $P(K) = \frac{2}{3}K$.

This contradicts the fact that put options are strictly convex functions of strike price, and creates an arbitrage opportunity.

The arbitrage comes from the fact that the put with strike 20 is overpriced. Using a "buy low, sell high" strategy, we could buy (i.e., go long) $\frac{2}{3}$ put options with strike 30, and sell (i.e., go short) 1 put option with strike 20. To avoid fractions, we set up the following portfolio:

- long 2 puts with strike 30;
- short 3 puts with strike 20.

This portfolio is set up at no initial cost, since the cash flow generated by selling 3 puts with strike 20 and buying 2 puts with strike 30 is \$0:

$$3 \cdot \$4 - 2 \cdot \$6 = \$0.$$

At the maturity T of the options, the value of the portfolio is

$$V(T) = 2 \max(30 - S(T), 0) - 3 \max(20 - S(T), 0).$$

Note that $V(T)$ is nonnegative for any value $S(T)$ of the underlying asset:

If $S(T) \geq 30$, then both put options expire worthless, and $V(T) = 0$.

If $20 \leq S(T) < 30$, then

$$V(T) = 2(30 - S(T)) > 0.$$

If $0 < S(T) < 20$, then

$$
\begin{aligned}
V(T) &= 2(30 - S(T)) - 3(20 - S(T)) \\
&= S(T) \\
&> 0.
\end{aligned}
$$

In other words, we took advantage of the existing arbitrage opportunity by setting up, at no initial cost, a portfolio with nonnegative payoff at T regardless of the price $S(T)$ of the underlying asset, and with a strictly positive payoff if $0 < S(T) < 30$. $\square$

Question 2. The number 2^{29} has 9 digits, all different. Without computing 2^{29}, find the missing digit.

Answer: For any positive integer n, denote by $D(n)$ the sum of the digits of n. Recall that the difference between a number and the sum of its digits is divisible by 9, i.e.,

$$9 \mid n - D(n);$$

see the footnote below[1] for details.

---

[1] If the digits of n are $a_k, a_{k-1}, \ldots a_1, a_0$ (from left to right), then

$$
\begin{aligned}
n &= a_k \cdot 10^k + a_{k-1} \cdot 10^{k-1} + \ldots a_1 \cdot 10 + a_0; \\
D(n) &= a_k + a_{k-1} + \ldots + a_1 + a_0.
\end{aligned}
$$

Hence,

$$n - D(n) = \sum_{i=0}^{k} a_i \cdot (10^i - 1).$$

Since $10^i - 1$ is an i-digit number with all digits equal to 9, it follows that $9 \mid 10^i - 1$, for all $i = 1 : k$, and therefore $9 \mid n - D(n)$.

Thus, for $n = 2^{29}$, it follows that

$$9 \mid 2^{29} - D\left(2^{29}\right). \tag{1.1}$$

We are given that 2^{29} has 9 digits, and that all 9 digits are different. Denote by x the missing digit. Then,

$$D\left(2^{29}\right) = \left(\sum_{j=0}^{9} j\right) - x = 45 - x. \tag{1.2}$$

From (1.1) and (1.2), it follows that

$$9 \mid 2^{29} - (45 - x). \tag{1.3}$$

Note that

$$
\begin{aligned}
2^{29} &= 2^5 \cdot (2^6)^4 = 2^5 \cdot 64^4 \\
&= 2^5 \cdot (63 + 1)^4 \\
&= 2^5 \cdot (63 \cdot k + 1) \\
&= 2^5 \cdot 63 \cdot k + 2^5,
\end{aligned} \tag{1.4}
$$

where k is a positive integer.[2]

From (1.4), we find that

$$2^{29} - 2^5 = 63 \cdot 2^5 \cdot k,$$

and therefore

$$9 \mid 2^{29} - 2^5. \tag{1.5}$$

From (1.3) and (1.5), it follows that

$$
\begin{aligned}
9 \mid & \left(2^{29} - 2^5\right) - \left(2^{29} - (45 - x)\right) \\
& = (45 - x) - 2^5 \\
& = 13 - x.
\end{aligned}
$$

---

[2] It is easy to see that

$$
\begin{aligned}
(63 + 1)^4 &= 63^4 + 4 \cdot 63^3 + 6 \cdot 63^2 + 4 \cdot 63 + 1 \\
&= 63 \cdot (63^3 + 4 \cdot 63^2 + 6 \cdot 63 + 4) + 1 \\
&= 63 \cdot k + 1,
\end{aligned}
$$

where $k = 63^3 + 4 \cdot 63^2 + 6 \cdot 63 + 4$.

Since $9 \mid 13 - x$ and x is a digit, we conclude that $x = 4$. In other words, we identified that x, the missing digit from 2^{29}, must be 4.

Indeed, $2^{29} = 536\,870\,912$, i.e., 2^{29} has 9 digits, all different, and 4 is not a digit of 2^{29}. $\square$

Question 3. Let W_t be a Wiener process, and let

$$X_t = \int_0^t W_\tau d\tau. \tag{1.6}$$

What is the distribution of X_t? Is X_t a martingale?

Answer: Note that we can rewrite (1.6) in differential form as

$$dX_t = W_t dt = W_t dt + 0\, dW_t.$$

Then, X_t is a diffusion process with only drift part W_t, and therefore X_t is not a martingale.

We use integration by parts to find the distribution of X_t; a different solution can be found in Section 3.6.

By applying integration by parts, we obtain that

$$\begin{aligned}
X_t &= \int_0^t W_\tau d\tau \\
&= tW_t - \int_0^t \tau dW_\tau \\
&= t\int_0^t dW_\tau - \int_0^t \tau dW_\tau \\
&= \int_0^t (t - \tau) dW_\tau.
\end{aligned}$$

Recall that, if $f(t)$ is a deterministic square integrable function, then the stochastic integral $\int_0^t f(\tau) dW_\tau$ is a normal random variable of mean 0 and variance $\int_0^t |f(\tau)|^2 d\tau$, i.e.,

$$\int_0^t f(\tau) dW_\tau \sim N\left(0, \int_0^t |f(\tau)|^2 d\tau\right).$$

Thus,

$$\begin{aligned}
X_t &= \int_0^t (t-\tau) dW_\tau \\
&\sim N\left(0, \int_0^t (t-\tau)^2 \, d\tau\right) \\
&= N\left(0, \frac{t^3}{3}\right).
\end{aligned}$$

We conclude that X_t is a normal random variable of mean 0 and variance $\frac{t^3}{3}$.   □

Question 4. Alice and Bob stand at opposite ends of a straight line segment. Bob sends 50 ants towards Alice, one after another. Alice sends 20 ants towards Bob. All ants travel along the straight line segment. Whenever two ants collide, they simply bounce back and start traveling in the opposite direction. How many ants reach Bob and how many ants reach Alice? How many ant collisions take place?

Answer: Imagine that when two ants meet, they switch identities. Hence, even after a collision, two ants are traveling in two opposite directions. It follows that 20 ants reach Bob, while 50 ants reach Alice.

To calculate the number of ant collisions, imagine that each ant carries a message. In other words, Bob sends 50 messages to Alice, one message per ant. Similarly, Alice sends 20 messages to Bob, one message per ant. Furthermore, imagine that the two ants swap messages when they collide. Then a message always makes forward progress. Each of Alice's messages goes through 50 ant collisions. Each of Bob's messages goes through 20 ant collisions. The total number of collisions is 50 times 20, which is 1000 collisions.   □

Question 5. Find all the values of ρ such that

$$\begin{pmatrix} 1 & 0.6 & -0.3 \\ 0.6 & 1 & \rho \\ -0.3 & \rho & 1 \end{pmatrix}$$

is a correlation matrix.

Answer: A symmetric matrix with diagonal entries equal to 1 is a correlation matrix if and only if the matrix is symmetric positive semidefinite. Thus, we need to find all the values of ρ such that the matrix

$$\Omega = \begin{pmatrix} 1 & 0.6 & -0.3 \\ 0.6 & 1 & \rho \\ -0.3 & \rho & 1 \end{pmatrix} \tag{1.7}$$

is symmetric positive semidefinite.

We give a short solution using Sylvester's criterion. Two more solutions, one using the Cholesky decomposition, and another one based on the definition of symmetric positive semidefinite matrices will be given in Section 3.2.

Recall from Sylvester's criterion that a matrix is symmetric positive semidefinite if and only if all its principal minors are greater than or equal to 0. Also, recall that the principal minors of a matrix are the determinants of all the square matrices obtained by eliminating the same rows and columns from the matrix. In particular, the matrix Ω from (1.7) has the following principal minors:

$$\det(1) = 1; \quad \det(1) = 1; \quad \det(1) = 1;$$

$$\det \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix} = 0.64;$$

$$\det \begin{pmatrix} 1 & -0.3 \\ -0.3 & 1 \end{pmatrix} = 0.91;$$

$$\det \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = 1 - \rho^2;$$

$$\det(\Omega) = 1 - 0.36\rho - 0.09 - 0.36 - \rho^2$$
$$= 0.55 - 0.36\rho - \rho^2.$$

Thus, it follows from Sylvester's criterion that Ω is a symmetric positive semidefinite matrix if and only if

$$1 - \rho^2 \geq 0;$$
$$0.55 - 0.36\rho - \rho^2 \geq 0,$$

which is equivalent to $-1 \leq \rho \leq 1$ and

$$\rho^2 + 0.36\rho - 0.55 \leq 0. \tag{1.8}$$

Since the roots of the quadratic equation corresponding to (1.8) are -0.9432 and 0.5832, we conclude that the matrix Ω is symmetric positive semidefinite, and therefore a correlation matrix, if and only if

$$-0.9432 \leq \rho \leq 0.5832. \quad \square \tag{1.9}$$

Question 6. How many independent random variables uniformly distributed on $[0,1]$ should you generate to ensure that there is at least one between 0.70 and 0.72 with probability 95%?

Answer: Denote by N the smallest number of random variables you should generate such that

$$P(\text{at least one r.v. in } [0.70, 0.72]) \geq 0.95. \tag{1.10}$$

The probability that a random variable uniformly distributed on $[0,1]$ is not in the interval $[0.70, 0.72]$ is 0.98. Thus, the probability that none of the N independent variables are in $[0.70, 0.72]$ is 0.98^N, i.e.,

$$P(\text{no r.v. in } [0.70, 0.72]) = 0.98^N.$$

Note that

$$P(\text{at least one r.v. in } [0.70, 0.72])$$
$$= 1 - P(\text{no r.v. in } [0.70, 0.72])$$
$$= 1 - (0.98)^N. \tag{1.11}$$

From (1.10) and (1.11), we find that N is the smallest integer such that

$$1 - (0.98)^N \geq 0.95,$$

which is equivalent to

$$(0.98)^N \leq 0.05$$
$$\iff N \ln(0.98) \leq \ln(0.05)$$
$$\iff N \geq \frac{\ln(0.05)}{\ln(0.98)} \approx 148.28$$
$$\iff N = 149.$$

We conclude that at least 149 uniform random variables on $[0, 1]$ must be generated in order to have 95% confidence that at least one of the random variables is between 0.70 and 0.72.   □

Question 7. Show that the probability density function of the standard normal integrates to 1.

Answer: The probability density function of the standard normal variable is $\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$. We want to show that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} \, dt = 1,$$

which, using the substitution $t = \sqrt{2}x$, can be written as

$$I = \int_{-\infty}^{\infty} e^{-x^2} \, dx = \sqrt{\pi}. \tag{1.12}$$

We prove (1.12) by using polar coordinates. Since x is just an integrating variable, we can also write the integral I in terms of another integrating variable, denoted by y, as $I = \int_{-\infty}^{\infty} e^{-y^2} dy$.

Then,[3]

$$
\begin{aligned}
I^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2} \, dx \right) \cdot \left(\int_{-\infty}^{\infty} e^{-y^2} \, dy \right) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2} e^{-y^2} \, dx dy \qquad (1.13) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2 + y^2)} \, dx dy.
\end{aligned}
$$

We use the polar coordinates transformation $x = r\cos\theta$ and $y = r\sin\theta$, with $r \in [0,\ \infty)$ and $\theta \in [0,\ 2\pi)$, to evaluate the last integral. Since $dx dy = r d\theta dr$, we obtain that

$$
\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2 + y^2)} \, dx dy \\
&= \int_{0}^{\infty} \int_{0}^{2\pi} r \, e^{-\left(r^2 \cos^2 \theta + r^2 \sin^2 \theta \right)} \, d\theta dr \\
&= \int_{0}^{\infty} \int_{0}^{2\pi} r \, e^{-r^2} \, d\theta dr \qquad (1.14) \\
&= \int_{0}^{\infty} 2\pi \, r \, e^{-r^2} \, dr \\
&= 2\pi \lim_{t \to \infty} \int_{0}^{t} r \, e^{-r^2} \, dr \\
&= 2\pi \lim_{t \to \infty} \left(-\frac{1}{2} e^{-r^2} \right) \Big|_{0}^{t} \\
&= \pi;
\end{aligned}
$$

note that (1.14) follows from the equality $\cos^2 \theta + \sin^2 \theta = 1$ for any real number θ.

---

[3]Note that Fubini's theorem is needed for a rigorous derivation of the equality (1.13); this technical step is rarely required by the interviewer.

Since $I > 0$, $I = \sqrt{\pi}$, which is what we wanted to prove; see (1.12). $\square$

Question 8. Assume the Earth is perfectly spherical and you are standing somewhere on its surface. You travel exactly 1 mile south, then 1 mile east, then 1 mile north. Surprisingly, you find yourself back at the starting point. If you are not at the North Pole, where can you possibly be?!

Answer: There are infinitely many locations, aside from the North Pole, that have this property.

Somewhere near the South Pole, there is a latitude that has a circumference of one mile. In other words, if you are at this latitude and start walking east (or west), in one mile you will be back exactly where you started from. If you instead start at some point one mile north of this latitude, your journey will take you one mile south to this special latitude, then one mile east "around the globe" and finally one mile north right back to wherever you started from. Moreover, there are infinitely many points on the Earth that are one mile north of this special latitude, where you could start your journey and eventually end up exactly where you started.

We are still not finished! There are infinitely many special latitudes as well; namely, you could start at any point one mile north of the latitude that has a circumference of $1/k$ miles, where k is a positive integer. Your journey will take you one mile south to this special latitude, then one mile east looping "around the globe" k times, and finally one mile north right back to where you started from. $\square$

Question 9. Solve the Ornstein-Uhlenbeck SDE

$$dr_t = \lambda(\theta - r_t)dt + \sigma dW_t, \qquad (1.15)$$

which is used, e.g., in the Vasicek model for interest rates.

Answer: We can rewrite (1.15) as

$$dr_t + \lambda r_t dt = \lambda \theta dt + \sigma dW_t. \tag{1.16}$$

By multiplying (1.16) on both sides by the integrating factor $e^{\lambda t}$, we obtain that

$$e^{\lambda t} dr_t + \lambda e^{\lambda t} r_t dt = \lambda \theta e^{\lambda t} dt + \sigma e^{\lambda t} dW_t,$$

which is equivalent to

$$d\left(e^{\lambda t} r_t\right) = \lambda \theta e^{\lambda t} dt + \sigma e^{\lambda t} dW_t. \tag{1.17}$$

By integrating (1.17) from 0 to t, it follows that

$$
\begin{aligned}
e^{\lambda t} r_t - r_0 &= \lambda \theta \int_0^t e^{\lambda s} ds + \sigma \int_0^t e^{\lambda s} dW_s \\
&= \theta \left(e^{\lambda t} - 1\right) + \sigma \int_0^t e^{\lambda s} dW_s.
\end{aligned}
$$

By solving for r_t, we find that the solution to the Ornstein-Uhlenbeck SDE is

$$
\begin{aligned}
r_t &= e^{-\lambda t} r_0 + e^{-\lambda t} \theta \left(e^{\lambda t} - 1\right) + \sigma e^{-\lambda t} \int_0^t e^{\lambda s} dW_s \\
&= e^{-\lambda t} r_0 + \theta \left(1 - e^{-\lambda t}\right) + \sigma \int_0^t e^{-\lambda(t-s)} dW_s.
\end{aligned}
$$

Note that the process r_t is mean reverting to θ, regardless of the starting point r_0. To see this, recall that the expected value of the stochastic integral $\int_0^t f(s) dW_s$ of a non-random function $f(s)$ is 0. Then,

$$E\left[\int_0^t e^{-\lambda(t-s)} dW_s\right] = 0,$$

and therefore

$$E[r_t] = e^{-\lambda t} r_0 + \theta \left(1 - e^{-\lambda t}\right).$$

Thus,

$$\lim_{t \to \infty} E[r_t] = \theta. \quad \square$$

Question 10. Write a C++ function that computes the n-th Fibonacci number.

Answer: The Fibonacci numbers $(F_n)_{n \geq 0}$ are given by the following recurrence:

$$F_{n+2} = F_{n+1} + F_n, \quad \forall \, n \geq 0,$$

with $F_0 = 0$ and $F_1 = 1$.

```cpp
//recursive implementation
int fib(int n) {
    if (n == 0 || n == 1) return n;
    else {
        return fib(n-1) + fib(n-2);
    }
}
```

```cpp
//iterative implementation
int fib(int n ){
    if (n == 0 || n == 1) return n;
    int prev = 0, last = 1, temp;
    for (int i = 2; i <= n; ++i) {
        temp = last;
        last = prev + last;
        prev = temp;
    }
    return last;
}
```

```cpp
//tail recursive implementation
int fib(int n, int last = 1, int prev = 0)
{
    if (n == 0) return prev;
```

```
    if (n == 1) return last;
        return fib(n-1, last+prev, last);
}
```

Chapter 2

Questions

1. Mathematics, calculus, differential equations.

2. Covariance and corelation matrices. Linear algebra.

3. Financial instruments: options, bonds, swaps, forwards, futures.

4. C++. Data structures.

5. Monte Carlo simulations. Numerical methods.

6. Probability. Stochastic calculus.

7. Brainteasers.

2.1 Mathematics, calculus, differential equations.

1. What is the value of i^i, where $i = \sqrt{-1}$?

2. Which number is larger, π^e or e^π?

3. Show that
$$\frac{e^x + e^y}{2} \geq e^{\frac{x+y}{2}}, \quad \forall\, x, y \in \mathbb{R}.$$

4. Solve $x^6 = 64$.

5. What is the derivative of x^x?

6. Calculate
$$\sqrt{2 + \sqrt{2 + \sqrt{2 + \ldots}}}$$

7. Find x such that
$$x^{x^{x^{\cdot^{\cdot^{\cdot}}}}} = 2.$$

8. Which of the following series converge:
$$\sum_{k=1}^{\infty} \frac{1}{k}; \quad \sum_{k=1}^{\infty} \frac{1}{k^2}; \quad \sum_{k=2}^{\infty} \frac{1}{k \ln(k)}?$$

9. Compute

$$\int \frac{1}{1+x^2} \, dx.$$

10. Compute

$$\int x \ln(x) dx \quad \text{and} \quad \int xe^x dx.$$

11. Compute

$$\int x^n \ln(x) \, dx.$$

12. Compute

$$\int (\ln(x))^n \, dx.$$

13. Solve the ODE

$$y'' - 4y' + 4y = 1.$$

14. Find $f(x)$ such that

$$f'(x) = f(x)(1 - f(x)).$$

15. Derive the Black-Scholes PDE.

2.2 Covariance and correlation matrices. Linear algebra.

1. Show that any covariance matrix is symmetric positive semidefinite. Show that the same is true for correlation matrices.

2. Find the correlation matrix of three random variables with covariance matrix

$$\Sigma_X = \begin{pmatrix} 1 & 0.36 & -1.44 \\ 0.36 & 4 & 0.80 \\ -1.44 & 0.80 & 9 \end{pmatrix}.$$

3. Assume that all the entries of an $n \times n$ correlation matrix which are not on the main diagonal are equal to ρ. Find upper and lower bounds on the possible values of ρ.

4. How many eigenvalues does an $n \times n$ matrix with real entries have? How many eigenvectors?

5. Let

$$A = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}.$$

(i) Find a 2×2 matrix M such that $M^2 = A$;

(ii) Find a 2×2 matrix M such that $A = MM^t$.

6. The 2×2 matrix A has eigenvalues 2 and -3 with corresponding eigenvectors $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 3 \end{pmatrix}$. If $v = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$, find Av.

7. Let A and B be square matrices of the same size. Show that the traces of the matrices AB and BA are equal.

8. Can you find $n \times n$ matrices A and B such that

$$AB - BA = I_n,$$

where I_n is the identity matrix of size n?

9. A probability matrix is a matrix with nonnegative entries such that the sum of the entries in each row of the matrix is equal to 1. Show that the product of two probability matrices is a probability matrix.

10. Find all the values of ρ such that

$$\begin{pmatrix} 1 & 0.6 & -0.3 \\ 0.6 & 1 & \rho \\ -0.3 & \rho & 1 \end{pmatrix}$$

is a correlation matrix.[1]

---

[1] A solution to this question was given in Chapter 1 using Sylvester's criterion; two different solutions will be given herein.

2.3 Financial instruments: options, bonds, swaps, forwards, futures.

1. The prices of three put options with strikes 40, 50, and 70, but otherwise identical, are $10, $20, and $30, respectively. Is there an arbitrage opportunity present? If yes, how can you make a riskless profit?

2. The price of a stock is $50. In three months, it will either be $47 or $52, with 50% probability. How much would you pay for an at–the–money put? Assume for simplicity that the stock pays no dividends and that interest rates are zero.

3. A stock worth $50 today will be worth either $60 or $40 in three months, with equal probability. The value of a three months at–the–money put on this stock is $4. Does the value of the three months ATM put increase or decrease, and by how much, if the probability of the stock going up to $60 were 75% and the probability of the stock going down to $40 were 25%?

4. What is risk–neutral pricing?

5. Describe briefly how you arrive at the Black–Scholes formula.

6. How much should a three months at–the–money put on an asset with spot price $40 and volatility 30% be worth? Assume, for simplicity, that interest rates are zero and that the asset does not pay dividends.

7. If the price of a stock doubles in one day, by how much will the value of a call option on this stock change?

8. What are the smallest and largest values that Delta can take?

9. What is the Delta of an at–the–money call? What is the Delta of an at–the–money put?

10. What is the Put–Call parity? How do you prove it?

11. Show that the time value of a European call option is highest at–the–money.

12. What is implied volatility? What is a volatility smile? How about a volatility skew?

13. What is the Gamma of an option? Why is it preferable to have small Gamma? Why is the Gamma of plain vanilla options positive?

14. When are a European call and a European put worth the same? (The options are written on the same asset and have the same strike and maturity.) What is the intuition behind this result?

15. What is the two year volatility of an asset with 30% six months volatility?

16. How do you value an interest rate swap?

17. By how much will the price of a ten year zero coupon bond change if the yield increases by ten basis points?

18. A five year bond with 3.5 years duration is worth 102. What is the value of the bond if the yield decreases by fifty basis points?

19. What is a forward contract? What is the forward price?

20. What is the forward price for treasury futures contracts? What is the forward price for commodities futures contracts?

21. What is a Eurodollar futures contract?

22. What are the most important differences between forward contracts and futures contracts?

23. What is the ten–day 99% VaR of a portfolio with a five–day 98% VaR of $10 million?

2.4 C++. Data structures.

1. How do you declare an array?

2. How do you get the address of a variable?

3. How do you declare an array of pointers?

4. How do you declare a const pointer, a pointer to a const and a const pointer to a const?

5. How do you declare a dynamic array?

6. What is the general form for a function signature?

7. How do you pass-by-reference?

8. How do you pass a read only argument by reference?

9. What are the important differences between using a pointer and a reference?

10. How do you set a default value for a parameter?

11. How do you create a template function?

12. How do you declare a pointer to a function?

13. How do you prevent the compiler from doing an implicit conversion with your class?

14. Describe all the uses of the keyword **static** in C++.

15. Can a static member function be const?

16. C++ constructors support the initialization of member data via an initializer list. When is this preferable to initialization inside the body of the constructor?

17. What is a copy constructor, and how can the default copy constructor cause problems when you have pointer data members?

18. What is the output of the following code:

```
#include <iostream>
using namespace std;

class A
{
public:
    int * ptr;
    ~A()
    {
        delete(ptr);
    }
};
```

```cpp
void foo(A object_input)
{
    ;
}

int main()
{
    A aa;
    aa.ptr = new int(2);
    foo(aa);
    cout<<(*aa.ptr)<<endl;
    return 0;
}
```

19. How do you overload an operator?

20. What are smart pointers?

21. What is encapsulation?

22. What is a polymorphism?

23. What is inheritance?

24. What is a virtual function? What is a pure virtual function and when do you use it?

25. Why are virtual functions use for destructors? Can they be used for constructors?

26. Write a function that computes the factorial of a positive integer.

27. Write a function that takes an array and returns the subarray with the largest sum.

28. Write a function that returns the prime factors of a positive integer.

29. Write a function that takes a 64-bit integer and swaps the bits at indices i and j.

30. Write a function that reverses a single linked list.

31. Write a function that takes a string and returns true if its parenthesis are balanced.

32. Write a function that returns the height of an arbitrary binary tree.

2.5 Monte Carlo simulations. Numerical methods.

1. How would you compute π using Monte Carlo simulations? What is the standard deviation of this method?

2. What methods do you know for generating independent samples of the standard normal distribution?

3. How do you generate a geometric Brownian motion stock path using random numbers from a normal distribution?

4. How do you generate a sample of the standard normal distribution from 12 independent samples of the uniform distribution on $[0, 1]$?

5. What is the rate of convergence for Monte Carlo methods?

6. What variance reduction techniques do you know?

7. How do you generate samples of normal random variables with correlation ρ?

8. What is the order of convergence of the Newton's method?

9. Which finite difference method corresponds to trinomial trees?

2.6 Probability. Stochastic calculus.

1. What is the exponential distribution? What are the mean and the variance of the exponential distribution?

2. If X and Y are independent exponential random variables with mean 6 and 8, respectively, what is the probability that Y is greater than X?

3. What are the expected value and the variance of the Poisson distribution?

4. A point is chosen uniformly from the unit disk. What is the expected value of the distance between the point and the center of the disk?

5. Consider two random variables X and Y with mean 0 and variance 1, and with joint normal distribution. If $\text{cov}(X, Y) = \frac{1}{\sqrt{2}}$, what is the conditional probability $P(X > 0 | Y < 0)$?

6. If X and Y are lognormal random variables, is their product XY lognormally distributed?

7. Let X be a normal random variable with mean μ and variance σ^2, and let Φ be the cumulative distribution function of the standard normal distribution. Find the expected value of $Y = \Phi(X)$.

8. What is the law of large numbers?

9. What is the central limit theorem?

10. What is a martingale? How is it related to option pricing?

11. Explain the assumption $(dW_t)^2 = dt$ used in the informal derivation of Itô's Lemma.

12. If W_t is a Wiener process, find $E[W_t W_s]$.

13. If W_t is a Wiener process, what is $\text{var}(W_t + W_s)$?

14. Let W_t be a Wiener process. Find

$$\int_0^t W_s \, dW_s \quad \text{and} \quad E\left[\int_0^t W_s \, dW_s\right].$$

15. Find the distribution of the random variable

$$X = \int_0^1 W_t \, dW_t.$$

16. Let W_t be a Wiener process. Find the mean and the variance of

$$\int_0^t W_s^2 \, dW_s.$$

17. If W_t is a Wiener process, find the variance of

$$X = \int_0^1 \sqrt{t} e^{\frac{W_t^2}{8}} dW_t.$$

18. If W_t is a Wiener process, what is $E\left[e^{W_t}\right]$?

19. If W_t is a Wiener process, find the variance of

$$\int_0^t s \, dW_s.$$

20. Let W_t be a Wiener process, and let

$$X_t = \int_0^t W_\tau d\tau.$$

What is the distribution of X_t? Is X_t a martingale?[2]

21. What is an Itô process?

22. What is Itô's lemma?

23. If W_t is a Wiener process, is the process $X_t = W_t^2$ a martingale?

---

[2] A solution to this question was given in Chapter 1 using integration by parts; a different solution will be given herein.

24. If W_t is a Wiener process, is the process

$$N_t \ = \ W_t^3 - 3tW_t$$

a martingale?

25. What is Girsanov's theorem?

26. What is the martingale representation theorem, and how is it related to option pricing and hedging?

27. Solve $dY_t = Y_t \, dW_t$, where W_t is a Wiener process.

28. Solve the following SDEs:

    (i) $dY_t = \mu Y_t dt + \sigma Y_t dW_t$;

    (ii) $dX_t = \mu dt + (aX_t + b)dW_t$.

29. What is the Heston model?

2.7 Brainteasers.

1. A flea is going between two points which are 100 inches apart by jumping (always in the same direction) either one inch or two inches at a time. How many different paths can the flea travel by?

2. I have a bag containing three pancakes: one golden on both sides, one burnt on both sides, and one golden on one side and burnt on the other. You shake the bag, draw a pancake at random, look at one side, and notice that it is golden. What is the probability that the other side is golden?

3. Alice and Bob are playing heads and tails, Alice tosses $n + 1$ coins, Bob tosses n coins. The coins are fair. What is the probability that Alice will have strictly more heads than Bob?

4. Alice is in a restaurant trying to decide between three desserts. How can she choose one of three desserts with equal probability with the help of a fair coin? What if the coin is biased and the bias is unknown?

5. What is the expected number of times you must flip a fair coin until it lands on head? What if the coin is biased and lands on head with probability p?

6. What is the expected number of coin tosses of a fair coin in order to get two heads in a row? What if

the coin is biased with 25% probability of getting heads?

7. A fair coin is tossed n times. What is the probability that no two consecutive heads appear?

8. You have two identical Fabergé eggs, either of which would break if dropped from the top of a building with 100 floors. Your task is to determine the highest floor from which the an egg could be dropped without breaking. What is the minimum number of drops required to achieve this? You are allowed to break both eggs in the process.

9. An ant is in the corner of a $10 \times 10 \times 10$ room and wants to go to the opposite corner. What is the length of the shortest path the ant can take?

10. A $10 \times 10 \times 10$ cube is made of $1,000$ unit cubes. How many unit cubes can you see on the outside?

11. Fox Mulder is imprisoned by aliens in a large circular field surrounded by a fence. Outside the fence is a vicious alien that can run four times as fast as Mulder, but is constrained to stay near the fence. If Mulder can contrive to get to an unguarded point on the fence, he can quickly scale the fence and escape. Can he get to a point on the fence ahead of the alien?

12. At your subway station, you notice that of the two trains running in opposite directions which are supposed to arrive with the same frequency, the train

going in one direction comes first 80% of the time, while the train going in the opposite direction comes first only 20% of the time. What do you think could be happening?

13. You start off with one amoeba. Every minute, this amoeba can either die, do nothing, split into two amoebas, or split into three amoebas; all these scenarios being equally likely to happen. All further amoebas behave the same way. What is the probability that the amoebas eventually die off?

14. Given a set X with n elements, choose two subsets A and B at random. What is the probability of A being a subset of B?

15. Alice writes two distinct real numbers between 0 and 1 on two sheets of paper. Bob selects one of the sheets randomly to inspect it. He then has to declare whether the number he sees is the bigger or smaller of the two. Is there any way Bob can expect to be correct more than half the times Alice plays this game with him?

16. How many digits does the number 125^{100} have? You are not allowed to use values of $\log_{10} 2$ or $\log_{10} 5$.

17. For every subset of $\{1, 2, 3, \ldots, 2013\}$, arrange the numbers in the increasing order and take the sum with alternating signs. The resulting integer is called

the weight of the subset.[3]   Find the sum of the weights of all the subsets of $\{1, 2, 3, \ldots, 2013\}$.

18. Alice and Bob alternately choose one number from one of the following nine numbers: 1/16, 1/8, 1/4, 1/2, 1, 2, 4, 8, 16, without replacement. Whoever gets three numbers that multiply to one wins the game. Alice starts first. What should her strategy be? Can she always win?

19. Mr. and Mrs. Jones invite four other couples over for a party. At the end of the party, Mr. Jones asks everyone else how many people they shook hands with, and finds that everyone gives a different answer. Of course, no one shook hands with his or her spouse and no one shook the same person's hand twice. How many people did Mrs. Jones shake hands with?

20. The New York Yankees and the San Francisco Giants are playing in the World Series (best of seven format). You would like to bet $100 on the Yankees winning the World Series, but you can only place bets on individual games, and every time at even odds. How much should you bet on the first game?

21. We have two red, two green and two yellow balls. For each color, one ball is heavy and the other is light. All heavy balls weigh the same. All light balls weigh the same. How many weighings on a scale are necessary to identify the three heavy balls?

---

[3]For example, the weight of the subset $\{3\}$ is 3. The weight of the subset $\{2, 5, 8\}$ is $2 - 5 + 8 = 5$.

22. There is a row of 10 rooms and a treasure in one of them. Each night, a ghost moves the treasure to an adjacent room. You are trying to find the treasure, but can only check one room per day. How do you find it?

23. How many comparisons do you need to find the maximum in a set of n distinct numbers? How many comparisons do you need to find both the maximum and minimum in a set of n distinct numbers?

24. Given a cube, you can jump from one vertex to a neighboring vertex with equal probability. Assume you start from a certain vertex (does not matter which one). What is the expected number of jumps to reach the opposite vertex?

25. Select numbers uniformly distributed between 0 and 1, one after the other, as long as they keep decreasing; i.e. stop selecting when you obtain a number that is greater than the previous one you selected.

26. To organize a charity event that costs $100K, an organization raises funds. Independent of each other, one donor after another donates some amount of money that is exponentially distributed with a mean of $20K. The process is stopped as soon as $100K or more has been collected. Find the distribution, mean, and variance of the number of donors needed until at least $100K has been collected.

27. Consider a random walk starting at 1 and with equal probability of moving to the left or to the right by one unit, and stopping either at 0 or at 3.

(i) What is the expected number of steps to do so?

(ii) What is the probability of the random walk ending at 3 rather than at 0?

28. A stick of length 1 drops and breaks at a random place uniformly distributed across the length. What is the expected length of the smaller part?

29. You are given a stick of unit length.

(i) The stick drops and breaks at two places. What is the probability that the three pieces could form a triangle?

(ii) The stick drops and breaks at one place. Then the larger piece is taken and dropped again, breaking at one place. What is the probability that the three pieces could form a triangle?

30. Why is a manhole cover round?

31. When is the first time after 12 o'clock that the hour and minute hands of a clock meet again?

32. Three light switches are in one room, and they turn three light bulbs in another. How do you figure out which switch turns on which bulb in one shot?

Chapter 3

Solutions

3.1 Mathematics, calculus, differential equations.

Question 1. What is the value of i^i, where $i = \sqrt{-1}$?

Answer: Recall that $e^{i\theta} = \cos\theta + i\sin\theta$. Then,

$$i = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2} = e^{i\frac{\pi}{2}},$$

and therefore

$$i^i = \left(e^{i\frac{\pi}{2}}\right)^i = e^{(i\frac{\pi}{2})i} = e^{i^2\frac{\pi}{2}} = e^{-\frac{\pi}{2}},$$

since $i^2 = -1$. $\square$

Question 2. Which number is larger, π^e or e^π?

Answer: We will show that $\pi^e < e^\pi$. By taking the natural logarithm, we find that

$$\pi^e < e^\pi \iff \ln(\pi^e) < \ln(e^\pi) \iff e\ln(\pi) < \pi$$
$$\iff \frac{\ln(\pi)}{\pi} < \frac{1}{e}, \tag{3.1}$$

41

which can be written as

$$\frac{\ln(\pi)}{\pi} < \frac{\ln(e)}{e}.$$

Let $f : (0, \infty) \to \mathbb{R}$ given by $f(x) = \frac{\ln(x)}{x}$. Then,

$$f'(x) = \frac{1 - \ln(x)}{x^2}.$$

Note that $f'(x) = 0$ has one solution, $x = e$. Also, $f'(x) > 0$ for $0 < x < e$, and $f'(x) < 0$ for $x > e$, and therefore $f(x)$ is increasing on the interval $(0, e)$ and is decreasing on the interval (e, ∞).

Thus, the function $f(x) = \frac{\ln(x)}{x}$ has a global maximum point at $x = e$, i.e., $f(x) < f(e) = \frac{1}{e}$ for all $x > 0$ with $x \neq e$, and therefore

$$f(\pi) = \frac{\ln(\pi)}{\pi} < \frac{\ln(e)}{e} = \frac{1}{e},$$

which is equivalent to $\pi^e < e^\pi$; cf. (3.1). $\square$

Question 3. Show that

$$\frac{e^x + e^y}{2} \geq e^{\frac{x+y}{2}}, \quad \forall \, x, y \in \mathbb{R}. \tag{3.2}$$

Answer: Let $e^x = a$ and $e^y = b$. Note that $a, b > 0$, and that

$$e^{\frac{x+y}{2}} = \sqrt{e^{x+y}} = \sqrt{e^x \cdot e^y} = \sqrt{ab}.$$

Then, (3.2) can be written as

$$\frac{a + b}{2} \geq \sqrt{ab} \iff a + b - 2\sqrt{ab} \geq 0$$

$$\iff \left(\sqrt{a} - \sqrt{b}\right)^2 \geq 0,$$

which is what we wanted to show. $\square$

Question 4. Solve $x^6 = 64$.

Answer: Recall that the six unit roots of $z^6 = 1$ are

$$z_k = \exp\left(\frac{2k\pi i}{6}\right) = \exp\left(\frac{k\pi i}{3}\right)$$
$$= \cos\left(\frac{k\pi}{3}\right) + i\sin\left(\frac{k\pi}{3}\right), \quad (3.3)$$

for $k = 0:5$. Since $\sqrt[6]{64} = 2$, we obtain from (3.3) that the solutions of $x^6 = 64$ are

$$x_k = 2\cos\left(\frac{k\pi}{3}\right) + 2i\sin\left(\frac{k\pi}{3}\right), \quad \forall\, k = 0:5. \quad \square$$

Question 5. What is the derivative of x^x?

Answer: Note that

$$x^x = e^{\ln(x^x)} = e^{x\ln(x)}. \quad (3.4)$$

Using Chain Rule and (3.4), we find that

$$(x^x)' = \left(e^{x\ln(x)}\right)' = e^{x\ln(x)}\left(x\ln(x)\right)'$$
$$= x^x(\ln(x) + 1). \quad \square$$

Question 6. Calculate

$$\sqrt{2 + \sqrt{2 + \sqrt{2 + \ldots}}} \quad (3.5)$$

Answer: Assume that the limit from (3.5) exists, and denote that limit by l. Then, $l = \sqrt{2 + l}$, which can be written as

$$l^2 - l - 2 = (l - 2)(l + 1) = 0.$$

Since $l > 0$, we obtain that $l = 2$, i.e.,

$$\sqrt{2 + \sqrt{2 + \sqrt{2 + \ldots}}} = 2. \qquad (3.6)$$

Thus, proving (3.6) is equivalent to showing that the sequence $(x_n)_{n \geq 0}$ given by $x_0 = \sqrt{2}$ and

$$x_{n+1} = \sqrt{2 + x_n}, \quad \forall\, n \geq 0,$$

is convergent.

We can see by induction that the sequence $(x_n)_{n \geq 0}$ is bounded from above by 2, since $x_0 = \sqrt{2} < 2$, and, if we assume that $x_n < 2$, then

$$x_{n+1} = \sqrt{2 + x_n} < \sqrt{4} = 2.$$

Moreover, the sequence $(x_n)_{n \geq 0}$ is increasing, since

$$\begin{aligned}
x_n < x_{n+1} &\iff x_n < \sqrt{2 + x_n} \\
&\iff x_n^2 - x_n - 2 < 0 \\
&\iff (x_n - 2)(x_n + 1) < 0,
\end{aligned}$$

which holds true since $x_n > 0$, and since, as shown above, $x_n < 2$ for all $n \geq 0$.

Thus, the sequence $(x_n)_{n \geq 0}$ is convergent since it is increasing and bounded from above, which is what we needed to show in order to prove (3.6). $\square$

Question 7. Find x such that

$$x^{x^{x^{\cdot^{\cdot^{\cdot}}}}} = 2. \qquad (3.7)$$

Answer: If x exists such that (3.7) holds true, then

$$x^{x^{x^{x^{\cdot^{\cdot^{\cdot}}}}}} = x^2 = 2,$$

and therefore the only possible solution to (3.7) is $x = \sqrt{2}$. We prove that $x = \sqrt{2}$ is, indeed, the solution to (3.7), by showing that

$$\sqrt{2}^{\sqrt{2}^{\sqrt{2}^{\cdots}}} = 2. \qquad (3.8)$$

Consider the sequence $(x_n)_{n \geq 0}$ with $x_0 = \sqrt{2}$ and satisfying the following recursion:

$$x_{n+1} = \left(\sqrt{2}\right)^{x_n} = 2^{x_n/2}, \ \forall \, n \geq 0. \qquad (3.9)$$

We can see by induction that $(x_n)_{n \geq 0}$ is an increasing sequence, since $x_0 = \sqrt{2} < \sqrt{2}^{\sqrt{2}} = x_1$, and, if we assume that $x_{n-1} < x_n$, then

$$x_n = 2^{x_{n-1}/2} < 2^{x_n/2} = x_{n+1}.$$

Also by induction, we can prove that the sequence $(x_n)_{n \geq 0}$ is bounded from above by 2, since $x_0 = \sqrt{2} < 2$, and, if $x_n < 2$, then

$$x_{n+1} = 2^{x_n/2} < 2.$$

Thus, the sequence $(x_n)_{n \geq 0}$ is convergent since it is increasing and bounded from above.

Let $l = \lim_{n \to \infty} x_n$. From (3.9), we find that $l = 2^{l/2}$, which is equivalent to

$$l^{1/l} = 2^{1/2}. \qquad (3.10)$$

The function $f : (0, \infty) \to (0, \infty)$ given by

$$f(t) = t^{1/t} = \exp\left(\ln(t^{1/t})\right)$$

$$= \exp\left(\frac{\ln(t)}{t}\right)$$

is increasing for $t < e$ and decreasing for $t > e$, since

$$f'(t) = \frac{1 - \ln(t)}{t^2} \exp\left(\frac{\ln(t)}{t}\right)$$

and $f'(t) > 0$ for $t < e$ and $f'(t) < 0$ for $t > e$.

Thus, there are two values of l such that (3.10) is satisfied, i.e., such that $l^{1/l} = 2^{1/2}$, one value being equal to 2, and the other one greater than e. Since we showed that $x_n < 2$ for all $n \geq 0$, we conclude that $l = \lim_{n \to \infty} x_n = 2$, which is what we needed to show to complete the proof of (3.8). $\square$

Question 8. Which of the following series converge:

$$\sum_{k=1}^{\infty} \frac{1}{k}; \quad \sum_{k=1}^{\infty} \frac{1}{k^2}; \quad \sum_{k=2}^{\infty} \frac{1}{k\ln(k)}?$$

Answer: We show that

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \quad \text{is convergent;}$$

$$\sum_{k=1}^{\infty} \frac{1}{k} \quad \text{and} \quad \sum_{k=2}^{\infty} \frac{1}{k\ln(k)} \quad \text{are divergent.}$$

Since all the terms of the series $\sum_{k=1}^{\infty} \frac{1}{k^2}$ are positive, it is enough to show that the partial sums $\sum_{k=1}^{n} \frac{1}{k^2}$ are uniformly bounded, in order to conclude that the series is convergent. This can be seen as follows:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{1}{k^2} &= 1 + \sum_{k=2}^{n} \frac{1}{k^2} \\
&\leq 1 + \sum_{k=2}^{n} \frac{1}{k(k-1)} \\
&= 1 + \sum_{k=2}^{n} \left(\frac{1}{k-1} - \frac{1}{k} \right) \\
&= 1 + \left(1 - \frac{1}{n} \right) \\
&< 2, \quad \forall \ n \geq 2.
\end{aligned}$$

To show that the series $\sum_{k=1}^{\infty} \frac{1}{k}$ is divergent, we will prove that

$$\sum_{k=1}^{n} \frac{1}{k} > \ln(n) + \frac{1}{n}, \ \ \forall \, n \geq 1. \tag{3.11}$$

Since $\frac{1}{x}$ is a decreasing function, it follows that

$$\frac{1}{x} < \frac{1}{k}, \ \ \forall \, k < x < k+1.$$

Then,

$$\begin{aligned}
\int_{1}^{n} \frac{1}{x} \, dx &= \sum_{k=1}^{n-1} \int_{k}^{k+1} \frac{1}{x} \, dx \\
&< \sum_{k=1}^{n-1} \int_{k}^{k+1} \frac{1}{k} \, dx \\
&= \sum_{k=1}^{n-1} \frac{1}{k} \\
&= -\frac{1}{n} + \sum_{k=1}^{n} \frac{1}{k}. \tag{3.12}
\end{aligned}$$

From (3.12), we find that

$$\begin{aligned}
\sum_{k=1}^{n} \frac{1}{k} &> \int_{1}^{n} \frac{1}{x} \, dx + \frac{1}{n} \\
&= \ln(n) + \frac{1}{n}, \ \ \forall \, n \geq 1,
\end{aligned}$$

which is what we wanted to prove; see (3.11).

Similarly, note that

$$\frac{1}{x \ln(x)} < \frac{1}{k \ln(k)}, \ \ \forall \, k < x < k+1,$$

and therefore

$$\int_2^{n+1} \frac{1}{x\ln(x)}\, dx = \sum_{k=2}^{n} \int_k^{k+1} \frac{1}{x\ln(x)} dx$$
$$< \sum_{k=2}^{n} \int_k^{k+1} \frac{1}{k\ln(k)} dx$$
$$= \sum_{k=2}^{n} \frac{1}{k\ln(k)}. \qquad (3.13)$$

Since

$$\int_2^{n+1} \frac{1}{x\ln(x)}\, dx = \ln(\ln(n+1)) - \ln(\ln(2)),$$

we obtain from (3.13) that

$$\sum_{k=2}^{n} \frac{1}{k\ln(k)} > \ln(\ln(n+1)) - \ln(\ln(2)),$$

and we conclude that the series $\sum_{k=2}^{\infty} \frac{1}{k\ln(k)}$ is divergent.

Note: Although not needed to answer this question, it can be shown that

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6},$$

and

$$\lim_{n\to\infty} \left(\sum_{k=1}^{\infty} \frac{1}{k} - \ln(n) \right) = \gamma,$$

where $\gamma \approx 0.57721$ is Euler's constant. $\qquad \square$

Question 9. Compute

$$\int \frac{1}{1+x^2}\, dx.$$

Answer: Use the substitution $x = \tan(z)$. Then, $dx = \frac{1}{\cos^2(z)} dz$ and

$$
\begin{aligned}
\int \frac{1}{1+x^2} \, dx &= \int \frac{1}{(1+\tan^2(z))\cos^2(z)} \, dz \\
&= \int \frac{1}{\cos^2(z)+\sin^2(z)} \, dz \\
&= \int 1 \, dz \\
&= z + C,
\end{aligned}
$$

where C is a real constant, since $\cos^2(z) + \sin^2(z) = 1$ for any z. Solving $x = \tan(z)$ for z, we obtain that $z = \arctan(x)$, and therefore

$$
\int \frac{1}{1+x^2} \, dx = \arctan(x) + C. \quad \square
$$

Question 10. Compute

$$
\int x \ln(x)dx \quad \text{and} \quad \int xe^x dx.
$$

Answer: By integration by parts,

$$
\begin{aligned}
\int x \ln(x) \, dx &= \frac{x^2}{2}\ln(x) - \int \frac{x^2}{2} \cdot \frac{1}{x} \, dx \\
&= \frac{x^2}{2}\ln(x) - \frac{1}{2}\int x \, dx \\
&= \frac{x^2 \ln(x)}{2} - \frac{x^2}{4} + C; \\
\int xe^x \, dx &= xe^x - \int 1 \cdot e^x \, dx \\
&= xe^x - e^x + C. \quad \square
\end{aligned}
$$

Question 11. Compute

$$\int x^n \ln(x) \ dx.$$

Answer: If $n \neq -1$, we use integration by parts and find that

$$
\begin{aligned}
\int x^n \ln(x) \ dx &= \frac{x^{n+1}}{n+1} \ln(x) - \int \frac{x^{n+1}}{n+1} \cdot \frac{1}{x} \ dx \\
&= \frac{x^{n+1} \ln(x)}{n+1} - \frac{1}{n+1} \int x^n \ dx \\
&= \frac{x^{n+1} \ln(x)}{n+1} - \frac{x^{n+1}}{(n+1)^2} + C.
\end{aligned}
$$

For $n = -1$, we obtain that

$$\int \frac{\ln(x)}{x} \ dx = \frac{(\ln(x))^2}{2} + C,$$

where C is a real constant, since

$$\left(\frac{(\ln(x))^2}{2} \right)' = \frac{1}{2} \cdot 2\ln(x) \cdot (\ln(x))' = \frac{\ln(x)}{x}. \quad \square$$

Question 12. Compute

$$\int (\ln(x))^n \ dx.$$

Answer: For every integer $n \geq 0$, let

$$f_n(x) = \int (\ln(x))^n \ dx.$$

By using integration by parts, we find that, for any $n \geq 1$,

$$\int (\ln(x))^n \, dx$$

$$= x(\ln(x))^n - \int x \left((\ln(x))^n\right)' \, dx$$

$$= x(\ln(x))^n - \int x \cdot n(\ln(x))^{n-1} \cdot (\ln(x))' \, dx$$

$$= x(\ln(x))^n - \int x \cdot n(\ln(x))^{n-1} \cdot \frac{1}{x} \, dx$$

$$= x(\ln(x))^n - n \int (\ln(x))^{n-1} \, dx,$$

and therefore

$$f_n(x) = x(\ln(x))^n - n f_{n-1}(x), \quad \forall \, n \geq 1. \qquad (3.14)$$

Note that

$$f_0(x) = \int 1 \, dx = x + C.$$

Thus, the recursion (3.14) can be used to find the values of $f_n(x)$ for all n. For example:

$$f_1(x) = x \ln(x) - f_0(x) = x(\ln(x) - 1) + C;$$
$$f_2(x) = x(\ln(x))^2 - 2 f_1(x)$$
$$= x \left((\ln(x))^2 - 2 \ln(x) + 2\right) + C.$$

The following general formula can be obtained by induction:

$$\int (\ln(x))^n \, dx = x \sum_{k=0}^{n} \frac{(-1)^{n-k} n!}{k!} (\ln(x))^k + C,$$

for all $n \geq 0$. $\quad \square$

Question 13. Solve the ODE

$$y'' - 4y' + 4y = 1. \qquad (3.15)$$

Answer: Note that (3.15) is a second order non homogeneous linear ODE with constant coefficients. The homogeneous ODE associated to (3.15) is

$$y'' - 4y' + 4y = 0, \qquad (3.16)$$

whose characteristic equation, $z^2 - 4z + 4 = 0$, has a double root $z_1 = z_2 = 2$. Thus, the solution to the homogeneous ODE (3.16) is

$$y(x) = c_1 e^{2x} + c_2 x e^{2x}, \qquad (3.17)$$

where c_1 and c_2 are constants.

Since the constant function $y_0(x) = \frac{1}{4}$ is a solution to the non homogeneous ODE (3.15), we conclude from (3.17) that the general form of the solution of (3.15) is

$$y(x) = c_1 e^{2x} + c_2 x e^{2x} + \frac{1}{4}. \ \square$$

Question 14. Find $f(x)$ such that

$$f'(x) = f(x)(1 - f(x)). \qquad (3.18)$$

Answer: Note that (3.18) is as an ODE with separable variables and can be written as follows:

$$\frac{y'}{y(1 - y)} = 1, \qquad (3.19)$$

where $y = f(x)$. By integrating (3.19) with respect to x we obtain that

$$\int \frac{y'}{y(1 - y)} \, dx = \int 1 \, dx = x + C_1, \qquad (3.20)$$

where $C_1 \in \mathbb{R}$ is a real constant.

Note that $dy = y' dx$, and therefore

$$
\begin{aligned}
\int \frac{y'}{y(1-y)} \, dx &= \int \frac{1}{y(1-y)} \, dy \\
&= \int \left(\frac{1}{y} + \frac{1}{1-y} \right) \, dy \\
&= \int \frac{1}{y} \, dy + \int \frac{1}{1-y} \, dy \\
&= \ln(|y|) - \ln(|1-y|) \\
&= \ln \left| \frac{y}{1-y} \right|.
\end{aligned}
\tag{3.21}
$$

From (3.20) and (3.21), it follows that

$$
\ln \left| \frac{y}{1-y} \right| = x + C_1,
$$

and therefore

$$
\left| \frac{y}{1-y} \right| = e^{x+C_1} = C_2 e^x,
$$

where $C_2 = e^{C_1} > 0$ is a positive real constant.

Thus, either $\frac{y}{1-y} = C_2 e^x$, or $\frac{y}{1-y} = -C_2 e^x$, which can be written as

$$
\frac{y}{1-y} = C e^x,
\tag{3.22}
$$

where C is a real constant.

From (3.22), we obtain that $y = \frac{C e^x}{1+C e^x}$. We conclude that the ODE (3.18) has the following solution:

$$
f(x) = \frac{C e^x}{1 + C e^x},
$$

where $C \in \mathbb{R}$ is a fixed constant. $\quad\square$

Question 15. Derive the Black-Scholes PDE.

Answer: Consider an asset with spot price S following a lognormal distribution with drift μ and volatility σ and paying dividends continuously at rate q. Then,

$$dS \; = \; (\mu - q)Sdt \; + \; \sigma SdW_t,$$

where W_t, $t \geq 0$, is a Wiener process.

Let $V = V(S,t)$ be the value at time t of a replicable non path dependent derivative security on this asset, when the underlying is priced at S. Set up a portfolio Π made of a long position in the derivative security V and a short position in

$$\Delta \; = \; \frac{\partial V}{\partial S} \qquad\qquad (3.23)$$

units of the asset. Then,

$$\Pi \; = \; V - \Delta S.$$

Denote by dS, dV, and $d\Pi$ the changes in the values of S, V, and Π, respectively, over an infinitesimally small time period dt. Then,

$$d\Pi \; = \; dV - \Delta dS - \Delta q Sdt, \qquad\qquad (3.24)$$

where $\Delta q Sdt$ is the dividend payment owed over the time dt on the short Δ units asset position. From (3.23) and (3.24), we find that

$$d\Pi \; = \; dV - \frac{\partial V}{\partial S}dS - qS\frac{\partial V}{\partial S}dt. \qquad\qquad (3.25)$$

From Itô's formula, it follows that

$$dV \; = \; \left(\frac{\partial V}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 V}{\partial S^2} \right) dt$$
$$+ \; \frac{\partial V}{\partial S}dS, \qquad\qquad (3.26)$$

and, from (3.25) and (3.26), we obtain that

$$d\Pi \; = \; \left(\frac{\partial V}{\partial t} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 V}{\partial S^2} - qS\frac{\partial V}{\partial S} \right) dt, \qquad (3.27)$$

which means that the value of the portfolio Π is deterministic over the small time period dt. For no–arbitrage, the value of the portfolio Π must grow at the risk–free rate over the time period dt, i.e., $d\Pi = r\Pi dt$, where r denotes the risk–free rate. Thus,

$$
\begin{aligned}
d\Pi &= r\Pi dt = r(V - \Delta S)dt \\
&= \left(rV - rS\frac{\partial V}{\partial S} \right) dt.
\end{aligned}
\tag{3.28}
$$

From (3.27) and (3.28), we find that

$$
\frac{\partial V}{\partial t} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 V}{\partial S^2} - qS\frac{\partial V}{\partial S} = rV - rS\frac{\partial V}{\partial S},
$$

and therefore

$$
\frac{\partial V}{\partial t} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 V}{\partial S^2} + (r - q)S\frac{\partial V}{\partial S} - rV = 0,
$$

which is the Black-Scholes PDE for $V(S, t)$. $\quad\square$

3.2   Covariance and correlation matrices. Linear algebra.

Question 1. Show that any covariance matrix is symmetric positive semidefinite. Show that the same is true for correlation matrices.

Answer: Let Σ_X and Ω_X be the covariance matrix and the correlation matrix of n random variables $X_1, X_2, \ldots, X_n$. It is easy to see that Σ_X and Ω_X are symmetric matrices:

$$
\begin{aligned}
\Sigma_X(j,k) &= \operatorname{cov}(X_j, X_k) = \operatorname{cov}(X_k, X_j) \\
&= \Sigma_X(k,j), \quad \forall\, 1 \le j, k \le n; \\
\Omega_X(j,k) &= \operatorname{corr}(X_j, X_k) = \operatorname{corr}(X_k, X_j) \\
&= \Omega_X(k,j), \quad \forall\, 1 \le j, k \le n.
\end{aligned}
$$

Let $c_1, c_2, \ldots, c_n$ be real numbers, and let $C = (c_i)_{i=1:n}$ be a column vector of size n. Recall that

$$
\operatorname{var}\left(\sum_{i=1}^{n} c_i X_i \right) = C^t \Sigma_X C. \qquad (3.29)
$$

Since the variance of any random variable is nonnegative, it follows that

$$
C^t \Sigma_X C \ge 0, \quad \forall\, C \in \mathbb{R}^n, \qquad (3.30)
$$

and we conclude that Σ_X is a symmetric positive semidefinite matrix.

For completeness, we include a proof of (3.29) here. Let $Y = \sum_{i=1}^{n} c_i X_i$. Then,

$$
Y - E[Y] = \sum_{i=1}^{n} c_i (X_i - \mu_i),
$$

where $\mu_i = E[X_i]$, for $i = 1 : n$, and therefore

$$\text{var}\left(\sum_{i=1}^{n} c_i X_i\right)$$

$$= \text{var}(Y) = E\left[(Y - E[Y])^2\right]$$

$$= E\left[\left(\sum_{i=1}^{n} c_i(X_i - \mu_i)\right)^2\right] \quad (3.31)$$

$$= E\left[\sum_{1 \leq j,k \leq n} c_j c_k (X_j - \mu_j)(X_k - \mu_k)\right]$$

$$= \sum_{1 \leq j,k \leq n} c_j c_k E[(X_j - \mu_j)(X_k - \mu_k)]$$

$$= \sum_{1 \leq j,k \leq n} c_j c_k \text{cov}(X_j, X_k)$$

$$= \sum_{1 \leq j,k \leq n} c_j c_k \Sigma_X(j, k)$$

$$= C^t \Sigma_X C, \quad (3.32)$$

where, for (3.31) and for (3.32), we used the following facts, respectively:

$$\left(\sum_{i=1}^{n} z_i\right)^2 = \sum_{1 \leq j,k \leq n} z_j z_k, \quad \forall\, z_i \in \mathbb{R},\ i = 1 : n;$$

$$x^t A x = \sum_{1 \leq j,k \leq n} x_j x_k A(j, k),$$

for any $n \times 1$ vector $x = (x_i)_{i=1:n} \in \mathbb{R}^n$, and for any $n \times n$ matrix A.

To show that Ω_X is a symmetric positive semidefinite matrix, recall the following correspondence between covariance matrices and correlation matrices:

$$\Sigma_X = D_{\sigma_X} \Omega_X D_{\sigma_X}, \quad (3.33)$$

where $D_{\sigma_X} = \text{diag}\,(\sigma_i)_{i=1:n}$ is the diagonal matrix with entries equal to the standard deviations of the n random variables, i.e., $\sigma_i^2 = \text{var}(X_i)$, for $i = 1 : n$.

Note that $(D_{\sigma_X})^{-1} = \text{diag}\left(\frac{1}{\sigma_i}\right)_{i=1:n}$. Let $v \in \mathbb{R}^n$, and let

$$w = (D_{\sigma_X})^{-1} v.$$

Then,

$$w^t = v^t \left((D_{\sigma_X})^{-1}\right)^t = v^t (D_{\sigma_X})^{-1}, \qquad (3.34)$$

since $(D_{\sigma_X})^{-1}$ is a diagonal matrix and therefore symmetric, i.e., $\left((D_{\sigma_X})^{-1}\right)^t = (D_{\sigma_X})^{-1}$.

From (3.33), (3.34), and (3.30), we find that

$$\begin{aligned}
w^t \Sigma_X w &= w^t (D_{\sigma_X} \Omega_X D_{\sigma_X}) w \\
&= v^t (D_{\sigma_X})^{-1} (D_{\sigma_X} \Omega_X D_{\sigma_X}) (D_{\sigma_X})^{-1} v \\
&= v^t \Omega_X v \\
&\geq 0.
\end{aligned}$$

Thus, $v^t \Omega_X v \geq 0$ for all $v \in \mathbb{R}^n$, and we conclude that Ω_X is a symmetric positive semidefinite matrix. $\square$

Question 2. Find the correlation matrix of three random variables with covariance matrix

$$\Sigma_X = \begin{pmatrix} 1 & 0.36 & -1.44 \\ 0.36 & 4 & 0.80 \\ -1.44 & 0.80 & 9 \end{pmatrix}. \qquad (3.35)$$

Answer: If Ω_X is the correlation matrix of the three random variables, then

$$\Sigma_X = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \Omega_X \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix},$$

where σ_i, $i = 1 : 3$, are the standard deviations of the three random variables, and therefore

$$\Omega_X = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 \\ 0 & 0 & \frac{1}{\sigma_3} \end{pmatrix} \Sigma_X \begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 \\ 0 & 0 & \frac{1}{\sigma_3} \end{pmatrix}.$$

Since the standard deviations of the random variables with covariance matrix Σ_X given by (3.35) are

$$\sigma_1 = \sqrt{\Sigma_X(1,1)} = 1;$$
$$\sigma_2 = \sqrt{\Sigma_X(2,2)} = 2;$$
$$\sigma_3 = \sqrt{\Sigma_X(2,2)} = 3,$$

we obtain that

$$\Omega_X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \Sigma_X \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0.18 & -0.48 \\ 0.18 & 1 & 0.1333 \\ -0.48 & 0.1333 & 1 \end{pmatrix}. \quad \square$$

Question 3. Assume that all the entries of an $n \times n$ correlation matrix which are not on the main diagonal are equal to ρ. Find upper and lower bounds on the possible values of ρ.

Answer: Recall that a symmetric matrix with diagonal entries equal to 1 is a correlation matrix if and only if the matrix is symmetric positive semidefinite, i.e., if and only if all the eigenvalues of the matrix are nonnegative.

Let

$$\Omega = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}.$$

We include two ways to compute the eigenvalues of Ω, which are then used to find the necessary and sufficient conditions for the matrix Ω to be a correlation matrix.

Solution 1: Note that

$$
\Omega = (1 - \rho)I + \begin{pmatrix} \rho & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & \rho \end{pmatrix}
$$

$$
= (1 - \rho)I + \rho M,
$$

where M is the $n \times n$ matrix with all entries equal to 1, and I is the $n \times n$ identity matrix.

Let λ and $v = (v_i)_{i=1:n}$ be an eigenvalue and a corresponding eigenvector of M, i.e., $Mv = \lambda v$, with $v \neq 0$.

Then, $Mv = \lambda v$ can be written as

$$
\begin{cases} v_1 + v_2 + \cdots + v_n = \lambda v_1; \\ v_1 + v_2 + \cdots + v_n = \lambda v_2; \\ \quad\vdots \qquad\qquad\qquad \vdots \\ v_1 + v_2 + \cdots + v_n = \lambda v_n, \end{cases}
$$

and therefore

$$
\lambda v_1 = \lambda v_2 = \ldots = \lambda v_n
$$

Thus, either $\lambda = 0$, or $v_1 = v_2 = \cdots = v_n$, in which case $n v_1 = \lambda v_1$, and therefore $\lambda = n$, since $v = (v_i)_{i=1:n} \neq 0$.

In other words, the eigenvalues of M are $\lambda = 0$ and $\lambda = n$.

Note that, if $Mv = \lambda v$, then,

$$
\begin{aligned} \Omega v &= (1 - \rho)v + \rho M v = (1 - \rho)v + \rho \lambda v \\ &= (1 - \rho + \rho \lambda)v. \end{aligned}
$$

Thus, $\mu = 1 - \rho + \rho \lambda$ and v are an eigenvalue and corresponding eigenvector of Ω.

Since the eigenvalues of M are $\lambda = 0$ and $\lambda = n$, it follows that the eigenvalues[1] of Ω are $\mu = 1 - \rho$, corresponding to $\lambda = 0$, and $\mu = (1 - \rho) + n\rho = 1 + (n - 1)\rho$, corresponding to $\lambda = n$.

Since Ω is a correlation matrix if and only if all its eigenvalues are nonnegative, we conclude that the matrix Ω is a correlation matrix if and only if

$$0 \leq 1 + (n - 1)\rho \quad \text{and} \quad 0 \leq 1 - \rho,$$

which is equivalent to

$$-\frac{1}{n - 1} \leq \rho \leq 1. \tag{3.36}$$

Solution 2: Note that

$$
\begin{aligned}
\Omega &= (1 - \rho)I + \rho \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \ldots & 1 & 1 \end{pmatrix} \\
&= (1 - \rho)I + \rho \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} (1 \ldots 1) \\
&= (1 - \rho)I + \rho w w^t \\
&= (1 - \rho)I + \rho A,
\end{aligned}
$$

where I is the $n \times n$ identity matrix and A is the $n \times n$ matrix given by $A = w w^t$, where w is the $n \times 1$ column vector of size n with all entries equal to 1.

Recall that an $n \times n$ matrix of the form $u u^t$, where $u = (u_i)_{i=1:n}$ is an $n \times 1$ column vector, has an eigenvalue equal to $\sum_{i=1}^{n} u_i^2$ with multiplicity 1 and another eigenvalue equal to 0 with multiplicity $n - 1$.

---

[1] The eigenvalue $1 - \rho$ has multiplicity $n - 1$, and the eigenvalue $1 + (n - 1)\rho$ has multiplicity 1; see Solution 2 of this question.

Then, the eigenvalues of the matrix A are:
$\lambda = \sum_{i=1}^{n} w_i^2 = \sum_{i=1}^{n} 1 = n$ with multiplicity 1;
$\lambda = 0$ with multiplicity $n - 1$.

Note that, if λ and v are an eigenvalue and a corresponding eigenvector of A, then $Av = \lambda v$, and therefore

$$\begin{aligned} \Omega v &= (1 - \rho)v + \rho Av = (1 - \rho)v + \rho \lambda v \\ &= (1 - \rho + \rho \lambda)v. \end{aligned}$$

Thus, $1 - \rho + \rho \lambda$ and v are an eigenvalue and corresponding eigenvector of Ω. and we obtain that the matrix Ω has the following eigenvalues:

- $(1 - \rho) + n\rho = 1 + (n - 1)\rho$ with multiplicity 1;
- $1 - \rho$ with multiplicity $n - 1$.

As before, since Ω is a correlation matrix if and only if all its eigenvalues are nonnegative, we conclude that the matrix Ω is a correlation matrix if and only if

$$0 \leq 1 + (n - 1)\rho \quad \text{and} \quad 0 \leq 1 - \rho,$$

which is equivalent to

$$-\frac{1}{n - 1} \leq \rho \leq 1,$$

which is the same as (3.36). $\square$

Question 4. How many eigenvalues does an $n \times n$ matrix with real entries have? How many eigenvectors?

Answer: Any $n \times n$ matrix with real entries has n eigenvalues, counted with their multiplicities; some of the eigenvalues may be complex numbers. Any $n \times n$ matrix has at most n eigenvectors.

Let A be an $n \times n$ matrix. Let λ be an eigenvalue of A with corresponding eigenvector $v \neq 0$, and let $P_A(x) = \det(xI_n - A)$ be the characteristic polynomial of A, where

I_n is the $n \times n$ identity matrix. Note that

$$Av = \lambda v, \; v \neq 0 \iff (\lambda I_n - A)v = 0, \; v \neq 0$$
$$\iff \lambda I_n - A \text{ singular matrix}$$
$$\iff \det(\lambda I_n - A) = 0$$
$$\iff P_A(\lambda) = 0.$$

In other words, λ is an eigenvalue of A if and only if λ is a root of the corresponding characteristic polynomial $P_A(x)$. Since $P_A(x)$ is a polynomial of degree n, it follows from the Fundamental Theorem of Algebra that $P_A(x)$ has exactly n (complex) roots when counted with their multiplicities. We conclude that any $n \times n$ matrix has n eigenvalues, counted with their multiplicities.

An eigenvalue of multiplicity m has at least one eigenvector and at most m linearly independent corresponding eigenvectors, but it may have less than m linearly independent eigenvectors.[2] Thus, an $n \times n$ matrix has at most n eigenvectors, and at least as many eigenvectors as the number of distinct eigenvalues of the matrix. $\square$

Question 5. Let

$$A = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}.$$

(i) Find a 2×2 matrix M such that $M^2 = A$;

(ii) Find a 2×2 matrix M such that $A = MM^t$.

Answer: (i) Recall that any symmetric matrix has the diagonal form

$$A = Q\Lambda Q^t, \tag{3.37}$$

---

[2]For example, the matrix $\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ has eigenvalue 2 with multiplicity 3 and only one eigenvector, $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

where Λ is the diagonal matrix whose entries on the main diagonal are the eigenvalues of A and Q is the orthogonal matrix whose columns are the corresponding eigenvectors of A of norm 1, i.e.,

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}; \quad Q = (v_1 \ v_2), \qquad (3.38)$$

where $Av_1 = \lambda_1 v_1$ and $Av_2 = \lambda_2 v_2$, with $||v_1|| = ||v_2|| = 1$.

If the matrix A has nonnegative eigenvalues, i.e., if $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, then the matrix

$$M = Q\Lambda^{1/2}Q^t \qquad (3.39)$$

with

$$\Lambda^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} \qquad (3.40)$$

has the property that $M^2 = A$:

$$\begin{aligned} M^2 &= \left(Q\Lambda^{1/2}Q^t\right)\left(Q\Lambda^{1/2}Q^t\right) \\ &= Q\Lambda^{1/2}(Q^tQ)\Lambda^{1/2}Q^t \\ &= Q\Lambda^{1/2}\Lambda^{1/2}Q^t \\ &= Q\Lambda Q^t \\ &= A, \end{aligned}$$

since Q is an orthogonal matrix and therefore $Q^tQ = I$, and since, from (3.40), it follows that $\Lambda^{1/2}\Lambda^{1/2} = \Lambda$.

We now proceed to compute the eigenvalues and the eigenvectors of the matrix A. The eigenvalues of the matrix A are the roots of the characteristic polynomial $P_A(x)$

of the matrix A given by[3]

$$
\begin{aligned}
P_A(x) &= \det(xI - A) = \det\begin{pmatrix} x-2 & 2 \\ 2 & x-5 \end{pmatrix} \\
&= (x-2)(x-5) - 4 = x^2 - 7x + 6 \\
&= (x-1)(x-6).
\end{aligned}
$$

The roots of $P_A(x)$ are 1 and 6, and therefore the eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = 6$. The corresponding eigenvectors of norm 1 are

$$
v_1 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix}.
$$

For example, if $\lambda_2 = 6$, any corresponding eigenvector $v_2 = \begin{pmatrix} a \\ b \end{pmatrix} \neq 0$ is a solution to $Av = 6v$, which can be written as

$$
\begin{cases} 2a - 2b = 6a \\ -2a + 5b = 6b \end{cases} \iff \begin{cases} b = -2a \\ b = -2a \end{cases}
$$

Thus, any eigenvector corresponding to the eigenvalue $\lambda_2 = 6$ is of the form

$$
v_2 = \begin{pmatrix} a \\ -2a \end{pmatrix} = a \begin{pmatrix} 1 \\ -2 \end{pmatrix}.
$$

By choosing $a = \frac{1}{\sqrt{5}}$, we obtain that an eigenvector of norm 1 corresponding to the eigenvalue $\lambda_2 = 6$ is

$$
v_2 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix}.
$$

---

[3]The characteristic polynomial of the matrix A can also be obtained as follows:

$$
P_A(x) = x^2 - \operatorname{tr}(A)x + \det(A) = x^2 - 7x + 6,
$$

where $\operatorname{tr}(A) = 2 + 5 = 7$ and $\det(A) = 2 \cdot 5 - (-2) \cdot (-2) = 6$.

Then, it follows from (3.39) and (3.40) that the matrix M given by

$$
\begin{aligned}
M &= \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \sqrt{1} & 0 \\ 0 & \sqrt{6} \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{pmatrix} \\
&= \frac{1}{5} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{6} \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix} \\
&= \frac{1}{5} \begin{pmatrix} 4+\sqrt{6} & 2-2\sqrt{6} \\ 2-2\sqrt{6} & 1+4\sqrt{6} \end{pmatrix}
\end{aligned}
$$

has the property that $M^2 = A$.

(ii) We found that the eigenvalues of A are 1 and 6, i.e., positive. Then, A is a symmetric positive definite matrix, and therefore has a Cholesky decomposition. Recall that, if U is the Cholesky factor of the matrix A, then, by definition, $A = U^t U$. Thus, in order to find a matrix M such that $A = MM^t$, it is enough to compute the Cholesky factor

$$
U = \begin{pmatrix} U(1,1) & U(1,2) \\ 0 & U(2,2) \end{pmatrix}
$$

of the matrix A and let $M = U^t$.

Note that $A = U^t U$ can be written as

$$
\begin{pmatrix} A(1,1) & A(1,2) \\ A(2,1) & A(2,2) \end{pmatrix} \tag{3.41}
$$

$$
= \begin{pmatrix} U(1,1) & 0 \\ U(1,2) & U(2,2) \end{pmatrix} \begin{pmatrix} U(1,1) & U(1,2) \\ 0 & U(2,2) \end{pmatrix}.
$$

In the first step of the Cholesky decomposition, the first row of U is computed as follows:

$$
\begin{aligned}
U(1,1) &= \sqrt{A(1,1)} = \sqrt{2}; \\
U(1,2) &= \frac{A(1,2)}{U(1,1)} = \frac{-2}{\sqrt{2}} = -\sqrt{2};
\end{aligned}
$$

see also (3.41).

In the next step of the Cholesky decomposition, the entry $A(2,2)$ of A is updated to $A(2,2) - U(1,2)^2 = 3$, and therefore

$$\begin{aligned} U(2,2) &= \sqrt{A(2,2) - (U(1,2))^2} \\ &= \sqrt{5 - (-\sqrt{2})^2} \\ &= \sqrt{3} \end{aligned}$$

Thus, the Cholesky factor of the matrix A is

$$U = \begin{pmatrix} \sqrt{2} & -\sqrt{2} \\ 0 & \sqrt{3} \end{pmatrix},$$

and therefore the matrix $M = U^t$ given by

$$M = \begin{pmatrix} \sqrt{2} & 0 \\ -\sqrt{2} & \sqrt{3} \end{pmatrix},$$

has the property that $A = MM^t$. $\quad\square$

Question 6. The 2×2 matrix A has eigenvalues 2 and -3 with corresponding eigenvectors $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 3 \end{pmatrix}$. If $v = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$, find Av.

Answer: Let $\lambda_1 = 2$, $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, and $\lambda_2 = -3$, $v_2 = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$. We first find constants $c_1, c_2 \in \mathbb{R}$ such that $v = c_1 v_1 + c_2 v_2$, i.e., such that

$$\begin{cases} 3 &= c_1 - c_2 \\ 1 &= 2c_1 + 3c_2 \end{cases}$$

The solution of this linear system is $c_1 = 2$ and $c_2 = -1$. Thus,

$$v = 2v_1 - v_2. \tag{3.42}$$

Since $Av_1 = \lambda_1 v_1 = 2v_1$ and $Av_2 = \lambda_2 v_2 = -3v_2$, we find from (3.42) that

$$
\begin{aligned}
Av &= 2Av_1 - Av_2 = 2(2v_1) - (-3v_2) \\
&= 4v_1 + 3v_2 \\
&= \begin{pmatrix} 1 \\ 17 \end{pmatrix}. \quad \square
\end{aligned}
$$

Question 7. Let A and B be square matrices of the same size. Show that the traces of the matrices AB and BA are equal.

Answer: Recall that, for any two square matrices A and B of the same size, the matrices AB and BA have the same characteristic polynomial, i.e.,

$$
\begin{aligned}
P_{AB}(x) &= \det(xI - AB) = \det(xI - BA) \\
&= P_{BA}(x), \ \forall \, x \in \mathbb{R}, \quad (3.43)
\end{aligned}
$$

where I is the identity matrix of the same size as the matrices A and B.

Also, recall the following connection between the characteristic polynomial $P_M(x)$ of an $n \times n$ matrix M and the trace $\mathrm{tr}(M)$ of the matrix:

$$
\begin{aligned}
P_M(x) \\
= \ \det(xI - M) \\
= \ x^n - \mathrm{tr}(M) + \cdots + (-1)^n \det(M). \quad (3.44)
\end{aligned}
$$

Since $P_{AB}(x) = P_{BA}(x)$, see (3.43), we conclude from (3.44) that

$$
\mathrm{tr}(AB) = \mathrm{tr}(BA). \quad (3.45)
$$

For completeness, we include a proof of (3.43). If the matrix B is nonsingular, then

$$
xI - AB = B^{-1}(xI - BA)B,
$$

and therefore

$$\begin{aligned} \det(xI - AB) &= \det(B^{-1})\det(xI - BA)\det(B) \\ &= \det(xI - BA), \qquad (3.46) \end{aligned}$$

since

$$\det(B^{-1})\det(B) = \det(B^{-1}B) = \det(I) = 1.$$

If the matrix B is singular, let ϵ be a real number, and note that the matrix $B - \epsilon I$ is singular if and only if ϵ is equal to an eigenvalue of B. Since the $n \times n$ matrix B has at most n eigenvalues, it follows that, except for a finite number of values of ϵ, the matrix $B - \epsilon I$ is nonsingular, in which case we obtain from (3.46) that

$$\det(xI - A(B - \epsilon I)) = \det(xI - (B - \epsilon I)A). \quad (3.47)$$

Since both sides of (3.47) are polynomials of degree n in ϵ, and therefore continuous functions of ϵ, we can let $\epsilon \to 0$ in (3.47) and obtain that

$$\begin{aligned} \lim_{\epsilon \to 0} &\left(\det(xI - A(B - \epsilon I)) \right) \\ &= \lim_{\epsilon \to 0} \left(\det(xI - (B - \epsilon I)A) \right) \\ \Longleftrightarrow \quad \det(xI - AB) &= \det(xI - BA). \quad (3.48) \end{aligned}$$

From (3.46) and (3.48), we conclude that $\det(xI - AB) = \det(xI - BA)$ regardless of whether the matrix B is nonsingular or singular, which concludes the proof of (3.43). $\square$

Question 8. Can you find $n \times n$ matrices A and B such that

$$AB - BA = I_n,$$

where I_n is the identity matrix of size n?

Answer: We give a proof by contradiction. If it were possible to find $n \times n$ matrices A and B such that $AB - BA = I$, then

$$\text{tr}(AB - BA) = \text{tr}(I_n) = n. \qquad (3.49)$$

However,

$$\text{tr}(AB - BA) = \text{tr}(AB) - \text{tr}(BA) = 0, \qquad (3.50)$$

since, if A and B are square matrices, then $\text{tr}(AB) = \text{tr}(BA)$; cf. (3.45).

Since (3.49) and (3.50) contradict each other, we conclude that there are no matrices A and B such that $AB - BA = I$. $\square$

Question 9. A probability matrix is a matrix with non-negative entries such that the sum of the entries in each row of the matrix is equal to 1. Show that the product of two probability matrices is a probability matrix.

Answer: We first establish the following equivalent definition for a probability matrix:

The $n \times n$ matrix M is a probability matrix if and only if all the entries of M are nonnegative and

$$M\mathbf{1} = \mathbf{1}, \qquad (3.51)$$

where $\mathbf{1}$ is the $n \times 1$ column vector with all entries equal to 1.

To see this, let $M = \text{row}\,(r_j)_{j=1:n}$ be the row form of the matrix M, where r_j is an $1 \times n$ row vector, for $j = 1 : n$. The sum of all the entries in the j-th row r_j of M can be written as follows:[4]

$$\sum_{k=1}^{n} r_j(k) = r_j\mathbf{1}. \qquad (3.52)$$

---

[4]Note that r_j is an $1 \times n$ vector and $\mathbf{1}$ is an $n \times 1$ vector, and therefore the expression $r_j\mathbf{1}$ from (3.52) is consistent.

Thus, the definition of a probability matrix as a matrix with the sum of the entries in each row equal to 1 can be written as

$$\sum_{k=1}^{n} r_j(k) = 1, \quad \forall \, j = 1:n$$
$$\iff r_j \mathbf{1} = 1, \quad \forall \, j = 1:n$$
$$\iff (r_j \mathbf{1})_{j=1:n} = \mathbf{1}$$
$$\iff M\mathbf{1} = \mathbf{1},$$

since $M\mathbf{1} = (r_j \mathbf{1})_{j=1:n}$ if $M = \text{row}\,(r_j)_{j=1:n}$ is the row form of M.

In other words, we established that (3.51) is an equivalent condition for M to be a probability matrix.

Let A and B be probability matrices. Then all the entries of A and B are nonnegative, and therefore all the entries of AB are also nonnegative. From (3.51), it follows that

$$A\mathbf{1} = \mathbf{1} \quad \text{and} \quad B\mathbf{1} = \mathbf{1},$$

and therefore

$$(AB)\mathbf{1} = A(B\mathbf{1}) = A\mathbf{1} = \mathbf{1}.$$

Then, from (3.51), we conclude that AB is a probability matrix. $\quad\square$

Question 10. Find all the values of ρ such that

$$\begin{pmatrix} 1 & 0.6 & -0.3 \\ 0.6 & 1 & \rho \\ -0.3 & \rho & 1 \end{pmatrix}$$

is a correlation matrix.

Answer: Recall that a solution to this question based on was Sylvester's criterion was included in Chapter 1. We give two more solutions to this question here, one using

the Cholesky decomposition, and another one based on
the definition of symmetric positive semidefinite matrices.

A symmetric matrix with diagonal entries equal to 1 is
a correlation matrix if and only if the matrix is symmetric
positive semidefinite. Thus, we need to find all the values
of ρ such that the matrix

$$\Omega = \begin{pmatrix} 1 & 0.6 & -0.3 \\ 0.6 & 1 & \rho \\ -0.3 & \rho & 1 \end{pmatrix} \tag{3.53}$$

is symmetric positive semidefinite.

Solution 1: To identify the values of ρ such that the ma-
trix Ω is symmetric positive semidefinite, we apply the
first step of the Cholesky algorithm to Ω, and obtain the
following 2×2 matrix:

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} - \begin{pmatrix} 0.6 \\ -0.3 \end{pmatrix} (0.6 \quad -0.3)$$

$$= \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} - \begin{pmatrix} 0.36 & -0.18 \\ -0.18 & 0.09 \end{pmatrix}$$

$$= \begin{pmatrix} 0.64 & \rho + 0.18 \\ \rho + 0.18 & 0.91 \end{pmatrix}.$$

Thus, the matrix Ω is symmetric positive semidefinite
if and only if the matrix

$$M = \begin{pmatrix} 0.64 & \rho + 0.18 \\ \rho + 0.18 & 0.91 \end{pmatrix}$$

is symmetric positive semidefinite. Since $M(1,1) = 0.64 > 0$, it follows that M is symmetric positive semidefinite if
and only if $\det(M) \geq 0$, i.e., if and only if

$$\det(M) = 0.5824 - (\rho + 0.18)^2 \geq 0, \tag{3.54}$$

which is equivalent to

$$|\rho + 0.18| \leq \sqrt{0.5824} = 0.7632.$$

We conclude that Ω is a symmetric positive semidefinite matrix, and therefore a correlation matrix, if and only if

$$-0.7632 \leq \rho + 0.18 \leq 0.7632,$$

which can be written as

$$-0.9432 \leq \rho \leq 0.5832. \tag{3.55}$$

Note that condition (3.55) is the same as condition (1.9) obtained when solving the same question using Sylvester's Criterion; see Chapter 1.

Solution 2: By definition, the matrix Ω is symmetric positive semidefinite if and only if $x^t \Omega x \geq 0$ for all $x = (x_i)_{i=1:3} \in \mathbb{R}^3$. Note that

$$
\begin{aligned}
& x^t \Omega x \\
=\ & (x_1\ x_2\ x_3) \begin{pmatrix} 1 & 0.6 & -0.3 \\ 0.6 & 1 & \rho \\ -0.3 & \rho & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\
=\ & x_1^2 + x_2^2 + x_3^2 + 1.2 x_1 x_2 - 0.6 x_1 x_3 + 2 \rho x_2 x_3.
\end{aligned}
$$

By completing the square, we obtain that

$$
\begin{aligned}
& x^t \Omega x \\
=\ & x_1^2 + 2 x_1 (0.6 x_2 - 0.3 x_3) + x_2^2 + x_3^2 + 2 \rho x_2 x_3 \\
=\ & (x_1 + 0.6 x_2 - 0.3 x_3)^2 \\
& - (0.6 x_2 - 0.3 x_3)^2 + x_2^2 + x_3^2 + 2 \rho x_2 x_3 \\
=\ & (x_1 + 0.6 x_2 - 0.3 x_3)^2 \\
& + 0.64 x_2^2 + 2 x_2 x_3 (\rho + 0.18) + 0.91 x_3^2.
\end{aligned}
$$

By completing the square once again, we find that

$$0.64x_2^2 + 2x_2x_3(\rho + 0.18) + 0.91x_3^2$$

$$= \left(0.8x_2 + x_3\frac{\rho + 0.18}{0.8}\right)^2$$

$$-x_3^2\frac{(\rho + 0.18)^2}{0.64} + 0.91x_3^2$$

$$= \left(0.8x_2 + x_3\frac{\rho + 0.18}{0.8}\right)^2$$

$$+\frac{x_3^2}{0.64}\left(0.5824 - (\rho + 0.18)^2\right),$$

and therefore

$$x^t\Omega x$$

$$= (x_1 + 0.6x_2 - 0.3x_3)^2$$

$$+\left(0.8x_2 + x_3\frac{\rho + 0.18}{0.8}\right)^2$$

$$+\frac{x_3^2}{0.64}\left(0.5824 - (\rho + 0.18)^2\right).$$

Thus, $x^t\Omega x \geq 0$ for all $x = (x_i)_{i=1:3} \in \mathbb{R}^3$ if and only if

$$(x_1 + 0.6x_2 - 0.3x_3)^2$$

$$+\left(0.8x_2 + x_3\frac{\rho + 0.18}{0.8}\right)^2$$

$$+\frac{x_3^2}{0.64}\left(0.5824 - (\rho + 0.18)^2\right)$$

$$\geq 0, \quad \forall \, x_1, x_2, x_3 \in \mathbb{R}.$$

The last inequality holds if and only if

$$0.5824 - (\rho + 0.18)^2 \geq 0, \qquad (3.56)$$

which is the same as (3.54).

We conclude that Ω is a correlation matrix if and only if

$$-0.9432 \leq \rho \leq 0.5832. \quad \square$$

3.3   Financial instruments: options, bonds, swaps, forwards, futures.

Question 1. The prices of three put options with strikes 40, 50, and 70, but otherwise identical, are \$10, \$20, and \$30, respectively. Is there an arbitrage opportunity present? If yes, how can you make a riskless profit?

Answer: If an arbitrage exists, it will be due to the fact that the convexity of put option values with respect to the strike price is violated.

In the plane (K, y), the line passing through the points $(K = 40, P(40) = 10)$ and $(K = 70, P(70) = 30)$ is given by

$$y = \frac{70 - K}{30} \cdot 10 + \frac{K - 40}{30} \cdot 30. \qquad (3.57)$$

The point on this line corresponding to strike 50 is obtained by substituting $K = 50$ in (3.57), and has y–coordinate equal to

$$\frac{2}{3} \cdot 10 + \frac{1}{3} \cdot 30 = \frac{50}{3}.$$

Since $P(K)$ is a strictly convex function of K, a no–arbitrage value of the put option with strike 50 should be below the line passing through the price points of the options with strikes 40 and 70. However, $P(50) = 20 > \frac{50}{3}$. Thus, the put option with strike 50 is overpriced, and an arbitrage exists.

Using a "buy low, sell high" strategy, we can take advantage of this arbitrage opportunity as follows: buy 2 put options with strike 40, buy 1 put option with strike 70, and sell 3 put options with strike 50. There is a \$10 positive cash flow when setting up this portfolio, since

$$3 \cdot \$20 - 2 \cdot \$10 - \$30 = \$10.$$

The value $V(T)$ of the portfolio at the maturity T of the options is

$$V(T) = 2\max(40 - S(T), 0)$$
$$+ \max(70 - S(T), 0)$$
$$- 3\max(50 - S(T), 0).$$

Note that $V(T)$ is nonnegative for any value $S(T)$ of the underlying asset at T:

If $70 \leq S(T)$, then all options expire out of the money and

$$V(T) = 0.$$

If $50 \leq S(T) < 70$, then

$$V(T) = 70 - S(T) \geq 0.$$

If $40 \leq S(T) < 50$, then

$$V(T) = (70 - S(T)) - 3(50 - S(T))$$
$$= 2S(T) - 80$$
$$\geq 0.$$

If $S(T) < 40$, then

$$V(T) = 2S(T) - 80 + 2(40 - S(T)) = 0.$$

In other words, we set up a portfolio with positive cash flow at inception which does not lose money regardless of the value of the underlying asset at time T. The risk–free profit is equal to the future value at time T of the \$10 cash flow from setting up the portfolio. $\square$

Question 2. The price of a stock is \$50. In three months, it will either be \$47 or \$52, with 50% probability. How much would you pay for an at the money put? Assume for simplicity that the stock pays no dividends and that interest rates are zero.

Answer: Recall first that real world probabilities do not play any role in valuing an option in a (one period) binomial tree model. Thus, the 50% probability stated in the question is only meant to throw you off–course.

Solution 1: The value of the option is the discounted expected value of the payoff of the option in the risk–neutral probability measure. Since interest rates are zero, this can be written as

$$P(0) = p_{RN,up}P_{up} + p_{RN,down}P_{down}. \qquad (3.58)$$

The up and down factors are $u = \frac{52}{50} = 1.04$ and $d = \frac{47}{50} = 0.94$, respectively. The risk–neutral probabilities of going up and down are

$$p_{RN,up} = \frac{1-d}{u-d} = 0.6; \quad p_{RN,down} = \frac{u-1}{u-d} = 0.4.$$

The ATM put pays \$3 if the stock price goes down to \$47, i.e., $P_{down} = 3$, and expires worthless if the stock price goes up to \$52, i.e., $P_{up} = 0$. From (3.58), we find that the value of the ATM put is

$$P(0) = 0.6 \cdot 0 + 0.4 \cdot 3 = 1.2. \qquad (3.59)$$

In other words, you should pay at most \$1.20 for an at the money put.

Solution 2: An insightful solution can be given by setting up a hedged portfolio. The Delta of the put option is

$$\Delta_P = \frac{P_{up} - P_{down}}{S_{up} - S_{down}} = \frac{0-3}{52-47} = -0.6.$$

A portfolio which is long one ATM put and short Δ_P shares will be long the put and long 0.6 shares, and will have the same value at maturity regardless of whether the stock price goes down to \$47 or up to \$52:

$$\Pi(T) = P(T) + 0.6S(T)$$
$$= \begin{cases} 0 + 0.6 \cdot 52 = 31.20, & \text{if } S(T) = 52; \\ 3 + 0.6 \cdot 47 = 31.20, & \text{if } S(T) = 47. \end{cases}$$

For no–arbitrage, the value of the portfolio at inception must be the discounted value of its payoff. Since the interest rates are zero, we obtain that $\Pi(0) = P(0) + 0.6S(0) = 31.20$, and therefore

$$P(0) = 31.20 - 0.6 \cdot 50 = 1.20,$$

which is the same value of the put, $1.20, obtained above; see (3.59). $\square$

Question 3. A stock worth $50 today will be worth either $60 or $40 in three months, with equal probability. The value of a three months at the money put on this stock is $4. Does the value of the three months ATM put increase or decrease, and by how much, if the probability of the stock going up to $60 were 75% and the probability of the stock going down to $40 were 25%?

Answer: In a one period binomial tree model, the actual probabilities of the asset going up or down do not play any role in the valuation of a plain vanilla option. Thus, the value of the three months at the money put would be the same, $4, even if the probability of the stock going up to $60 were 75%. $\square$

Question 4. What is risk–neutral pricing?

Answer: Risk–neutral pricing, or valuation, refers to valuing derivative securities as discounted expected values of their payoffs at maturity, under the assumption that the underlying asset has lognormal distribution with a drift equal to the risk-free rate.

More precisely, if the price of the underlying asset has a lognormal distribution with volatility σ and pays dividends continuously at the rate q, then the value of a derivative security on this asset with payoff $V(T)$ at maturity T given by risk–neutral valuation is

$$V(0) = e^{-rT} E_{RN}[V(S(T))],$$

where r is the risk–free rate assumed to be constant, and the expected value is computed with respect to the log-normal random variable $S(T)$ given by

$$S(T) \; = \; S(0) \, e^{\left(r - q - \frac{\sigma^2}{2}\right) T \, + \, \sigma \sqrt{T} z}.$$

Risk–neutral valuation can be used for derivative securities which can be perfectly hedged dynamically using cash and the underlying asset. Plain vanilla European options, as well as European options with other payoffs at maturity (such as asset–or–nothing and cash–or–nothing options) can be priced using risk–neutrality. Risk–neutral valuation cannot be used for path dependent options such as American options, barrier options, and Asian options. □

Question 5. Describe briefly how you arrive at the Black–Scholes formula.

Answer: The Black–Scholes formulas give the values of plain vanilla European put and call options on an underlying asset with lognormal distribution. Several methods for deriving the Black-Scholes formulas are:

• Risk neutral pricing: the expected value of the payoff of the option at maturity computed under the assumption that the price of the underlying asset has a lognormal distribution with drift equal to the risk free rate gives the Black–Scholes value of the option.

• Black–Scholes PDE solution: the Black–Scholes value of the option satisfies the Black–Scholes PDE with boundary conditions given by the payoff of the option at maturity. The Black–Scholes PDE is transformed into the heat PDE using a lognormal change of variables, and the closed form solution of the heat PDE is then used to derive the closed form solution of the Black–Scholes PDE, which is the Black–Scholes value of the option.

• Binomial tree model pricing: the evolution of the underlying asset is modeled using a binomial tree calibrated to converge in the limit to a lognormal distribution with drift equal to the risk free rate. For every tree, an approximate option value is obtained from the binomial tree model. The limit of these binomial tree option values as the number of time intervals in the tree goes to infinity is the Black–Scholes value of the option.

Note that twelve different ways to derive the Black–Scholes formula can be found in Wilmott [4]. □

Question 6. How much should a three months at the money put on an asset with spot price \$40 and volatility 30% be worth? Assume, for simplicity, that interest rates are zero and that the asset does not pay dividends.

Answer: The following approximation for the value of an at the money put option on a non dividend paying underlying asset and assuming zero risk–free interest rates is easy to estimate and very accurate if the total variance is small (e.g., if $\sigma^2 T \leq 0.25$):

$$P_{ATM} \approx 0.4\sigma S_0 \sqrt{T}; \qquad (3.60)$$

see Stefanica [3] for a derivation of formula (3.60).

For $S_0 = 40$, $\sigma = 0.3$, and $T = \frac{1}{4}$, we obtain that the value of the at the money put is approximately 2.40.

For comparison purposes, note that the value of the put option computed using the Black–Scholes formula would be 2.3914; the approximate formula (3.60) is very accurate in this case. □

Question 7. If the price of a stock doubles in one day, by how much will the value of a call option on this stock change?

Answer: The value of a deep in the money call on a non dividend paying asset can be approximated, e.g., by using

the Put-Call parity, as $C \approx S - Ke^{-rT}$, where K and T are the strike and the maturity of the option, and r is the constant risk free rate. Thus, if the spot price S doubles, the call option will be even deeper in the money, and therefore its value will be approximately $2S - Ke^{-rT}$. In other words, the value of deep in the money calls roughly doubles if the spot price doubles.

If the option is around at the money, the percentage change generated by the doubling of the stock price is about one order of magnitude larger since the option will become deep in the money.

If the option is deep out of the money, then it trades for fraction of cents. The doubling of the spot price would result in changing of the price of the option by several orders of magnitude.

As a numerical example, consider a six months call option with strike 20 on a non dividend paying underlying asset with volatility 25%. Assume that the risk free rate is constant at 5%. The Black–Scholes values of the call option corresponding to several spot prices of the underlying asset can be found below:

Spot Price	Option Price
10	0.000045
20	1.65
40	20.49
80	60.49
400	380.49
800	780.49

If the call option is deep out of the money and the spot price doubles from \$10 to \$20, the value of the call increases from \$0.000045 to \$1.65, i.e., by more than four orders of magnitude.

If the call option is at the money and the spot price doubles from \$20 to \$40, the value of the call increases

from \$1.65 to \$20.49, i.e., more than tenfold.

If the call option is deep in the money, and the spot price doubles from \$40 to \$80, the value of the call increases from \$20.49 to \$60.49, i.e., by a factor of 2.95; if the call option is even deeper in the money[5] and the spot price doubles from \$400 to \$800, the value of the call increases from \$380.49 to \$780.49, i.e., the call approximately doubles in value.

Moreover, if the call option is deep in the money, its value is very close to $S - Ke^{-rT}$, i.e., the value of the spot price of the underlying asset minus the present value of the strike. For all the spot prices greater than \$40, the estimate $C \approx S - Ke^{-rT}$ is very accurate. Thus, if the call option is at the money and the spot price doubles, the value of the call option increases by the same amount as the increase in the spot price. For example, if the spot price doubles from \$40 to \$80, the value of the call increases by \$40, from \$20.49 to \$60.49, which is exactly the increase in the spot price. □

Question 8. What are the smallest and largest values that Delta can take?

Answer: Assume, for simplicity, that the underlying asset does not pay dividends.

The Delta (Δ) of a long position in a plain vanilla call option is between 0 and 1 (and therefore the Delta of a short plain vanilla call position is between -1 and 0). The Delta of a long call position increases with the spot price of the underlying asset, and goes from 0, when the asset is worthless (i.e., when the call option is deep out of the money), to 1, when the spot price of the asset is very large (i.e., when the call option is deep in the money).

---

[5] Of course, call options with strike ten times smaller than the spot price of the underlying asset never occur in practice; this part of the example is for illustration purposes.

The Delta of a long position in a plain vanilla put option is between -1 and 0; the Delta of a short position in a plain vanilla put option is between 0 and 1. The Delta of a long put position also increases with the spot price of the underlying asset, and goes from -1, when the asset is worthless (i.e., when the put option is deep in the money), to 0, when the spot price of the asset is very large (i.e., when the put option is deep out of the money).

Note that, from the Put–Call parity, i.e., $C - P = S - Ke^{-rT}$, it follows that $\Delta(P) = \Delta(C) - 1$, which is consistent with the bounds above.   □

Question 9. What is the Delta of an at–the–money call? What is the Delta of an at–the–money put?

Answer: The Delta of an at–the–money call is approximately 0.5; the Delta of an at–the–money put is approximately -0.5.

Assume, for simplicity, a Black–Scholes framework with zero risk–free rates and an underlying asset paying no dividends. Then, $\Delta(C_{BS}) = N(d_1)$, where $N(x)$ denotes the cumulative distribution of the standard normal variable and

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - q + \frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}} = \frac{\sigma\sqrt{T}}{2} \quad (3.61)$$

for $K = S$ and for $r = q = 0$.

The linear Taylor approximation of $N(x)$ around 0 is

$$N(x) \approx 0.5 + \frac{x}{\sqrt{2\pi}} \approx 0.5 + 0.4x, \quad (3.62)$$

since $\frac{1}{\sqrt{2\pi}} = 0.3989 \approx 0.4$. Thus,

$$\Delta(C_{BS}) = N(d_1) = N\left(\frac{\sigma\sqrt{T}}{2}\right) \approx 0.5 + 0.2\sigma\sqrt{T}.$$

This is roughly estimated as $\Delta(C_{BS}) \approx 0.5$, since $0.2\sigma\sqrt{T}$ is small for most options (e.g., $0.2\sigma\sqrt{T} \leq 0.1$ for volatility less than 50% and for maturity less than one year).

For put options, $\Delta(P_{BS}) = -N(-d_1)$, where $d_1 = \frac{\sigma\sqrt{T}}{2}$; cf. (3.61). From (3.62), it follows that

$$\Delta(P_{BS}) = -N\left(-\frac{\sigma\sqrt{T}}{2}\right) \approx -0.5 + 0.2\sigma\sqrt{T},$$

which, using the same rationale outlined above, is often stated as $\Delta(P_{BS}) \approx -0.5$. □

Question 10. What is the Put–Call parity? How do you prove it?

Answer: The Put–Call parity is a model independent no–arbitrage relationship between the prices of European call and put options with the same strike and maturity.

In a nutshell, the Put–Call parity states that being long a call option and short a put option on the same underlying asset, and with the same strike and maturity, is the same as being long a forward contract on the asset, with the same maturity as the maturity of the options, and with delivery price equal to the strike of the options. Equivalently, being long a call and short a put is the same as being long one unit of the underlying asset (for non dividend paying assets) and short the present value of the strike of the options.

More precisely, if $C(t)$ and $P(t)$ are the values at time t of a European call and put option with maturity T and strike K, on the same non dividend paying asset with spot price $S(t)$, then the Put–Call parity states that

$$C(t) - P(t) = S(t) - Ke^{-r(T-t)}, \qquad (3.63)$$

where r denotes the risk–free rate, assumed to be constant.

If the underlying asset pays dividends continuously at the rate q, the Put–Call parity has the form

$$C(t) - P(t) = S(t)e^{-q(T-t)} - Ke^{-r(T-t)}. \qquad (3.64)$$

For simplicity, we restrict our attention to non dividend paying underlying assets and prove formula (3.63).

Consider a portfolio made of the following assets:
- long 1 put option;
- long 1 unit of the underlying asset;
- short 1 call option.

The value of the portfolio at time t is

$$V_{portfolio}(t) = P(t) + S(t) - C(t). \qquad (3.65)$$

The values of the call and put option at maturity are

$$C(T) = \max(S(T) - K, 0);$$
$$P(T) = \max(K - S(T), 0).$$

Then, regardless of whether $S(T) < K$ or $S(T) \geq K$, the value of the portfolio at time T will be equal to K:

If $S(T) < K$, then $P(T) = K - S(T)$ and $C(T) = 0$, and therefore

$$\begin{aligned} V_{portfolio}(T) &= P(T) + S(T) - C(T) \\ &= (K - S(T)) + S(T) - 0 \\ &= K. \end{aligned} \qquad (3.66)$$

If $S(T) \geq K$, then $P(T) = 0$ and $C(T) = S(T) - K$, and therefore

$$\begin{aligned} V_{portfolio}(T) &= P(T) + S(T) - C(T) \\ &= 0 + S(T) - (S(T) - K) \\ &= K. \end{aligned} \qquad (3.67)$$

From (3.66) and (3.67), we find that

$$V_{portfolio}(T) = P(T) + S(T) - C(T) = K, \qquad (3.68)$$

regardless of the value $S(T)$ of the underlying asset at the maturity of the option.

Then, for no–arbitrage, the value of the portfolio at time t must be equal to the present value of K at time t, i.e.,

$$V_{portfolio}(t) = Ke^{-r(T-t)}. \tag{3.69}$$

From (3.65) and (3.69), we obtain that

$$P(t) + S(t) - C(t) = Ke^{-r(T-t)}.$$

This can be written as

$$C(t) - P(t) = S(t) - Ke^{-r(T-t)},$$

which is the Put–Call parity formula (3.63).

The Put–Call parity formula (3.64) corresponding to an underlying asset paying dividends continuously at the rate q can be obtained similarly using a portfolio with a long put position, a short call position, and a long position in $e^{-q(T-t)}$ units of the underlying asset. All dividends payed by the long asset position between time t and time T are used to purchase additional fractions of the asset. Doing so continuously results in an asset position at time T equal to long one unit of the asset. Thus, the value of the portfolio at time T will be equal to K regardless of the value of $S(T)$, and the Put–Call parity formula (3.64) for underlying assets paying dividends continuously can be obtained as before. $\square$

Question 11. Show that the time value of a European call option is highest at the money.

Answer: The time value of a call option is the difference between the value $C(S)$ of the option and the intrinsic value $\max(S - K, 0)$ of the call option. In other words, the time value of the option is

$$C(S) - \max(S - K, 0).$$

We want to establish that the time value of the option is highest at the money, i.e., when $S = K$. To do so, we show that the function

$$f(S) = C(S) - \max(S - K, 0)$$

attains its maximum for $S = K$.

Note that[6]

$$f(S) = \begin{cases} C(S), & \text{if } S \leq K; \\ C(S) - S + K, & \text{if } S > K. \end{cases}$$

For $S \leq K$, the function $f(S)$ is the value of a call with strike K, and therefore is increasing.

For $S > K$, we find that

$$f'(S) = \Delta(C) - 1 < 0,$$

since the Delta of a call option is always less than 1,[7] and therefore the function $f(S)$ is decreasing.

We conclude that $f(S)$ has a global maximum point at $S = K$, which is what we wanted to show.

Note that a similar reasoning shows that the time value of a put option, given by

$$P(S) - \max(K - S, 0),$$

is largest at the money, i.e., for $S = K$. If $g(S) = P(S) - \max(K - S, 0)$, then

$$g(S) = \begin{cases} P(S) - K + S, & \text{if } S \leq K; \\ P(S), & \text{if } S > K. \end{cases}$$

[6]The function $f(S)$ is a continuous function, but it is not differentiable at $S = K$.

[7]The Delta of a long call position is an increasing function going from 0 when the spot price of the underlying asset is 0 and the call option is deep out of the money, to 1 when the spot price of the underlying asset goes to ∞ and the call option is deep in the money. For example, in the Black–Scholes model,

$$\Delta(C_{BS}) = e^{-qT} N(d_1) < N(d_1) < 1.$$

For $S \leq K$,

$$g'(S) = \Delta(P) + 1 > 0,$$

and therefore the function $g(S)$ is increasing, since the Delta of a put option is between -1 (when the put is deep in the money) and 0 (when the put is deep out of the money), and therefore is always greater than -1.

For $S > K$, the function $g(S)$ is the value of a put with strike K, and therefore is decreasing.

Thus, $g(S)$ has a global maximum point at $S = K$, and therefore the time value of the put option is largest at the money. $\square$

Question 12. What is implied volatility? What is a volatility smile? How about a volatility skew?

Answer: By definition, implied volatility is the unique value of the volatility parameter σ from the lognormal model for the evolution of the price of an underlying that makes the Black–Scholes value of an option equal to the market price of the option. Implied volatility exists and is unique[8] for any arbitrage–free market value of the option.

On the same asset, prices of options with multiple strikes and maturities are quoted, and implied volatilities can be computed for each of these options. If the price of the asset had a lognormal distribution as assumed in the Black–Scholes model, then the resulting plots of implied volatility vs strike for the same maturity should be flat. In practice, they are not flat, and are often shaped as "smiles" or "skews".

An implied volatility smile occurs when the implied volatilities of deep in the money options and of deep out of the money options are higher than the implied volatilities

---

[8]The uniqueness of the implied volatility comes from the fact that the Black–Scholes value of the option is a strictly increasing function of the volatility parameter (or, equivalently, the vega of the option is strictly positive).

of options close to at the money. Volatility smiles are typical for currency options.

An implied volatility skew occurs when the implied volatilities of options with large strikes are lower than the implied volatilities of at the money options (reverse skew), or when the implied volatilities of options with small strikes are lower than the implied volatilities of at the money options (forward skew). Reverse skews are typical for long dated equity and index options. Forward skews are typical for commodities options. ☐

Question 13. What is the Gamma of an option? Why is it preferable to have small Gamma? Why is the Gamma of plain vanilla options positive?

Answer: The Gamma (Γ) of an option measures the sensitivity of the Delta of the option with respect to the price of the underlying asset, i.e.,

$$\Gamma(V) = \frac{\partial \Delta}{\partial S} = \frac{\partial^2 V}{\partial S^2},$$

where V denotes the value of the option.

It is often important to immunize a portfolio with respect to changes in the price of the underlying asset, i.e., to make the portfolio Delta–neutral. A portfolio with small Gamma would need to be rebalanced less often in order to be kept Delta–neutral, since the change in the Delta of the portfolio is proportional to Gamma for small changes in the value of the underlying asset. Thus, a Delta–neutral and Gamma–neutral portfolio is well hedged against small changes in the price of the underlying asset (although not against jumps in the price of the underlying asset).

The Delta of plain vanilla options (calls or puts) increases as the spot price of the underlying asset increases. Thus, Gamma is positive, since Gamma is the rate of

change of Delta. Moreover, Gamma is asymptotically going to 0 for deep out of the money options and for deep in the money options, and the highest value of Gamma corresponds to options with strike close to the spot price (at the money options). $\square$

Question 14. When are a European call and a European put worth the same? (The options are written on the same asset and have the same strike and maturity.) What is the intuition behind this result?

Answer: Recall from the Put–Call parity that

$$C - P = Se^{-qT} - Ke^{-rT}, \qquad (3.70)$$

where C and P are the values of a call and put option, respectively, with strike K and maturity T on an underlying asset with spot price S and paying dividends continuously at rate q. If $C = P$, it follows from (3.70) that $K = Se^{(r-q)T}$. Since the forward price of the underlying asset is $F = Se^{(r-q)T}$, we conclude that a call and a put are worth the same if their strike is equal to the forward price of the asset; these options are called at–the–money–forward options.

Note that this result is independent of any assumption on the evolution of the price of the underlying asset, since Put–Call parity is model independent.

The fact that an at–the–money–forward call and an at–the–money–forward put are worth the same may seem counterintuitive at first glance: call options have unlimited upside since their payoff at maturity,

$$C(T) = \max(S(T) - K, 0),$$

can be infinitely large, while put options have limited upside since their payoff at maturity,

$$P(T) = \max(K - S(T), 0),$$

is bounded by the strike price, i.e., $P(T) \leq K$.

However, the value of an option is equal to the risk–neutral expected value of the option payoff at maturity. In every model for the evolution of the price of the underlying asset (including in the geometric Brownian motion model underlying the Black–Scholes framework), the probability density of the underlying asset at maturity decreases exponentially for large values of the spot price. This renders the expected value of large payoffs negligible, and makes it possible for at–the–money–forward put options to be worth the same as at–the–money–forward call options.   □

Question 15. What is the two year volatility of an asset with 30% six months volatility?

Answer: Asset volatility scales with the square root of time: if σ denotes the annualized volatility of an asset, the volatility $\sigma(t)$ of the asset over a time horizon t is given by $\sigma(t) = \sigma\sqrt{t}$. Then,

$$\frac{\sigma(t_2)}{\sigma(t_1)} = \sqrt{\frac{t_2}{t_1}}.$$

For $t_2 = 2$, $t_1 = 0.5$, and $\sigma(t_1) = \sigma(0.5) = 0.3$, we find that $\sigma(t_2) = \sigma(2) = 0.6$, i.e., the two year volatility of the asset is 60%.   □

Question 16. How do you value an interest rate swap?

Answer: For valuation purposes only, add payments at maturity equal to the notional of the swap both for the fixed leg and for the floating leg of the swap. Then, the value of the swap for the party receiving fixed payments is $V_{swap} = V_{fix} - V_{float}$, where V_{fix} is the value of a coupon bond with coupon rate equal to the fixed rate of the swap, and V_{float} is the value of an instrument making

the floating payments of the swap, plus a payment equal to the notional at maturity.

The value V_{fix} of the fixed rate coupon bond is the sum of the present values of its future payments discounted using risk–free zero rates (most frequently, LIBOR rates).

To compute V_{float}, note that, right after a payment is made, the remaining payments of the floating leg (including the notional at maturity) are equivalent to rolling over the notional at the prevailing zero rates until maturity. Thus, the value of all the floating payments on a payment date for the swap is equal to the notional. We conclude that V_{float} is the present value of the next floating payment (which was determined at the prior swap payment date) plus the notional.

To illustrate swap valuation with an example, consider a 19–month semiannual swap on a $10 million notional with 3% fixed rate and paying semiannually compounded LIBOR. The next floating payment that will be made in one month is $125,000 (and was determined five months ago, at the previous cash flow date).

The cash flow dates of the swap are 1 month, 7 months, 13 months, and 19 months. The value of the swap for the party receiving fixed payments is $V_{swap} = V_{fix} - V_{float}$. The value of the 3% semiannual coupon bond corresponding to the fixed leg of the swap is

$$
\begin{aligned}
V_{fix} \;=\; & 150,000 \cdot \mathrm{disc}\left(\frac{1}{12}\right) \\
& + 150,000 \cdot \mathrm{disc}\left(\frac{7}{12}\right) \\
& + 150,000 \cdot \mathrm{disc}\left(\frac{13}{12}\right) \\
& + 10,150,000 \cdot \mathrm{disc}\left(\frac{19}{12}\right), \quad (3.71)
\end{aligned}
$$

where $\mathrm{disc}(t)$ denotes the discount factor corresponding

to time t. For example, if the discount factors are given in terms of the semiannually compounded LIBOR rate LIBOR(t), then

$$\text{disc}(t) \ = \ \left(1 + \frac{\text{LIBOR}(t)}{2}\right)^{-2t}.$$

On the next cash flow date, i.e., in one month, the floating leg of the swap is equal to the value of the next floating payment, $125,000, plus the $10 million notional, i.e., $10,125,000. Thus, the value of the floating leg of the swap today is the present value of $10,125,000 in one month, i.e.,

$$V_{float} \ = \ 10,125,000 \cdot \text{disc}\left(\frac{1}{12}\right). \qquad (3.72)$$

The value of the swap for the party receiving fixed payments is $V_{swap} = V_{fix} - V_{float}$, where V_{fix} and V_{float} are given by (3.71) and (3.72), respectively. $\square$

Question 17. By how much will the price of a ten year zero coupon bond change if the yield increases by ten basis points?

Answer: The first order approximation of the change ΔB in the bond price in terms of the change Δy in the bond yield is

$$\Delta B \ \approx \ \frac{\partial B}{\partial y}\Delta y = -DB\Delta y, \qquad (3.73)$$

where $D = -\frac{1}{B}\frac{\partial B}{\partial y}$ is the duration of the bond. Thus,

$$\frac{\Delta B}{B} \ \approx \ -D\Delta y. \qquad (3.74)$$

Note that $D = 10$, since the duration of a zero coupon bond is equal to the maturity of the bond. Moreover, the change in the bond yield is $\Delta y = 0.001$, since $1\% = 100$ basis points (bps), and therefore 10 bps $= 0.1\% = 0.001$.

From (3.74), we find that the percentage change in the value of the bond can be estimated as follows:

$$\frac{\Delta B}{B} \approx -10 \times 0.001 = -0.01.$$

In other words, the price of the ten year zero-coupon bond decreases by 1% if the yield increases by ten basis points. □

Question 18. A five year bond with 3.5 years duration is worth 102. What is the value of the bond if the yield decreases by fifty basis points?

Answer: The value of the bond will increase, since the yield of the bond decreases. More precisely, recall from (3.73) that

$$\Delta B \approx -DB\Delta y. \qquad (3.75)$$

For $B = 102$, $D = 3.5$ and $\Delta y = -0.005$ (1% = 100 bps, and therefore 50 bps = 0.5% = 0.005), we find from (3.75) that $\Delta B \approx 1.785$. Thus, the new value of the bond is $B + \Delta B \approx 103.785$. □

Question 19. What is a forward contract? What is the forward price?

Answer: A forward contract is an agreement between two parties in which one party (the long position) agrees to buy a specified quantity of the underlying asset from the other party (the short position) at a given time in the future and for a price, called the forward price, that is agreed upon at the inception of the forward contract. The forward price is chosen such that the forward contract has value 0 at inception.[9]

If the underlying asset has spot price S_0 and pays dividends continuously at rate q, the forward price for a for-

---

[9]Note that the forward price is not the price of the forward.

ward contract maturing at T is

$$F = S_0 e^{(r-q)T},$$

where r is the risk–free rate between 0 and T.   □

Question 20. What is the forward price for treasury futures contracts? What is the forward price for commodities futures contracts?

Answer: A short position in a forward contract (i.e., selling a forward contract) is perfectly hedged by buying one unit of the underlying asset and holding that position until the maturity of the forward contract. The forward price is the future value at the maturity of the forward contract of the cost of buying one unit of the underlying asset.

For a treasury futures contract, buying the underlying treasury bond generates a positive cash flow from receiving all the bond coupon payments until the maturity of the forward contract. If S_0 is the spot price of the underlying treasury bond and if C is the present value of all the coupon payments received during the life of the futures contract, then the forward price F of a treasury futures contract with maturity T is the future value at time T of $S_0 - C$, i.e., $F = (S_0 - C)e^{rT}$.

For a commodities futures contract, buying the underlying commodity incurs storage costs; for example, for a gold futures contract, buying the underlying gold would require storing the gold safely. If S_0 is the spot price of the underlying commodity and if C is the present value of all the storage costs, then the forward price F of a commodities futures contract with maturity T is the future value at time T of $S_0 + C$, i.e., $F = (S_0 + C)e^{rT}$, where r denotes the risk–free rate between 0 and T.   □

Question 21. What is a Eurodollar futures contract?

Answer: A Eurodollar is a dollar deposit in a bank outside the United States. The Eurodollar rate is the interest rate earned on Eurodollars deposited by one bank with another bank (and is very close to LIBOR for short term maturities).

A Eurodollar futures contract is a futures contract on a Eurodollar rate, and deliveries are for up to ten years in the future.

For example, a three-month Eurodollar futures contract is a futures contract on the three month (90-day) Eurodollar rate. The start date is the third Wednesday of the delivery month (March, June, September, December). □

Question 22. What are the most important differences between forward contracts and futures contracts?

Answer: The main differences between the ways forward and futures contracts are structured, settled, and traded are:

• Futures contracts trade on exchanges and have standard features, while forward contracts are over-the-counter instruments.

• Futures are marked to market and settled in a margin account on a daily basis, while forward contracts are not settled before maturity.

• Futures carry almost no credit and counterparty risk, since they are settled daily, while entering into a forward contract carries some credit risk.

• Futures have a range of delivery dates, while forward contracts have a specified delivery date.

• Futures contracts require the delivery of the underlying asset for the futures price, while forward contracts can be settled in cash at maturity, without the delivery of a physical asset. □

Question 23. What is the ten–day 99% VaR of a portfolio with a five–day 98% VaR of $10 million?

Answer: If we assume normally distributed short term portfolio returns, the VaR (Value at Risk) of a portfolio is proportional to the square root of the time horizon. More precisely, if the N day $C\%$ VaR of the portfolio is denoted by $\text{VaR}(N, C)$, where N is the number of days in the time horizon and C is the confidence level, then

$$\text{VaR}(N, C) \approx \sigma_V z_C \sqrt{\frac{N}{252}} \, V(0), \qquad (3.76)$$

where, σ_V is the (annualized) standard deviation of the rate of return of the portfolio, z_C is the z–score of the standard normal distribution corresponding to C, i.e., $P(Z \leq z_C) = C$, and $V(0)$ is the current value of the portfolio.

From (3.76), it follows that

$$\text{VaR}(10 \text{ days}, 99\%) \approx \frac{z_{99}\sqrt{10}}{z_{98}\sqrt{5}} \, \text{VaR}(5 \text{ days}, 98\%).$$

For $\text{VaR}(5 \text{ days}, 98\%) = \$10,000,000$, and since $z_{99} \approx 2.326348$ and $z_{98} \approx 2.053749$, we obtain that

$$\text{VaR}(10 \text{ days}, 99\%) \approx \$16,019,255. \quad \square$$

3.4   C++. Data structures.

Question 1. How do you declare an array?

Answer: An array can be declared either on stack, or on heap.

```
//created on stack, uninitialized
T identifier[size];

//created on stack, initialized
T identifier[] = initializer_list;

//created on heap, uninitialized
T* identifier = new T[size];
```

Example:

```
int foo[3];
int bar[] = {1,2,3};
int* baz = new int[3];
```

Question 2. How do you get the address of a variable?

Answer: Use the ampersand before the name of the variable, e.g.

```
T var;
T* ptr = &var
```

Example:

```
int foo = 1;
int* foo_ptr = &foo;
```

Question 3. How do you declare an array of pointers?

Answer: The same way as declaring an array, but making the type, T, a pointer:

```
T* identifier[size];
T* identifier[] = initializer_list;
T** identifier = new T*[size];
```

Example:

```
int a = 1; int b = 2; int c = 3;
int* foo[3];
int* bar[] = {&a, &b, &c};
int** baz = new int*[3];
```

Question 4. How do you declare a const pointer, a pointer to a const and a const pointer to a const?

Answer:

```
//pointer to a read only variable
const T* identifier;
T const* identifier;

//read only pointer to a variable
T *const identifier = rvalue;

//read only pointer to a read only variable
const T *const identifier = rvalue;
T const *const identifier = rvalue;
```

Example:

```
//read only variables
const int a = 2; const int b = 2;
int c  = 1;

//pointer to a read only b
int const* foo_two;
foo_two = &a; foo_two = &b;

//pointer to read only a
```

```
const int* foo;
foo = &a; foo = &b;

//read only pointer to c
//it needs to be initialized
int *const bar = &c;

//read only pointer to read only a
//it needs to be initialized
const int *const baz = &a;
```

Question 5. How do you declare a dynamic array?

Answer:

```
T* identifier = new T[size];
T* identifier = nullptr;
T* identifier;

delete[] identifier;
```

Example:

```
int *foo = new int[4];
int *bar = nullptr;
bar = new int[4];
```

Question 6. What is the general form for a function signature?

Answer:

```
return_type function_name(parameter_list);
```

Example:

```
int my_sum(int a, int b);
```

Question 7. How do you pass-by-reference?

Answer:

```
return_type function_name(T & identifier);
```

The identifier is now an alias for the argument.

Question 8. How do you pass a read only argument by reference?

Answer:

```
return_type function_name(const T & identifier);
```

Once you define a parameter as `const`, you will not be able to modify it in the function.

Question 9. What are the important differences between using a pointer and a reference?

Answer: Several differences between using a pointer and a reference are:

• A pointer can be re-assigned any number of times, while a reference cannot be reassigned after initialization.

• A pointer can point to NULL (`nullptr` in C++11), while a reference can never be referred to NULL.

• It is not possible to take the address of a reference as it is done with pointers.

• There is no reference arithmetic.

Question 10. How do you set a default value for a parameter?

Answer:

```
return_type function_name(T identifier = rvalue)
```

The parameters with default value must be placed at the end of the parameter list.

Question 11. How do you create a template function?

Answer:

```
template<class T>
return_type function_name(parameter_list);

template<typename T>
return_type function_name(parameter_list);
```

Note that the parameter type can be specified, when calling the function, explicitly or implicitly. Also note that there is no technical difference between using **class** or **typename** besides code readability (**typename** for primitive types and **class** for classes).

Example:

```
template<typename T>
T temp_sum(T a, T b) {return a+b;}

struct Processor{
    int a;
    int apply(int b) {return a+b;}
};

template<class T>
int temp_sum_2(int a, int b) {
    T processor;
    processor.a = a;
    return processor.apply(b);
}

int main(){
    //implicit, foo equals 3
    int foo = temp_sum(1,2);

    //explicit, bar equals 3
    int bar = temp_sum_2<Processor>(1,2);
}
```

Question 12. How do you declare a pointer to a function?

Answer:

```
return_type (*identifier) (list_parameter_types)
```

Example:

```
int my_sum(int a, int b) {return a+b;}
int main(){
    int(*p_func)(int,int);
    p_func = & my_sum;

    // foo equals 3
    int foo = p_func(1,2);
}
```

Question 13. How do you prevent the compiler from doing an implicit conversion with your class?

Answer: Use the keyword **explicit** to define the constructor:

```
explicit Classname(parameter_list)
```

Question 14. Describe all the uses of the keyword **static** in C++.

Answer: Inside a function, using the keyword **static** means that once the variable has been initialized it remains in memory up until the end of the program.

Inside a class definition, either for a variable or for a member function, using the keyword **static** means that the there is only one copy of them per class, and shared between instances.

As a global variable in a file of code, using the keyword **static** means that the variable is private within the scope of the file.

Question 15. Can a static member function be const?

Answer: When the `const` qualifier is applied to a non-static member function it implies that member function can not change the instance class when called (i.e. can not change any non `mutable` members from `*this`). Since static member function are defined at a class level, where there is no notion of `this` the `const` qualifier for member functions does not apply.

Question 16. C++ constructors support the initialization of member data via an initializer list. When is this preferable to initialization inside the body of the constructor?

Answer: The initialization list has to be used for const members, references and with members without default constructors, but for any type of members initialization through the initialization list is still preferable, since it is for efficient. Using the initialization list, the members are initialized calling directly their constructors. If the initialization is done in the body of the constructor for each member being initialized there is an instance of it created and then a copy assignment operation is called to assign that instance to its respective member.

Question 17. What is a copy constructor, and how can the default copy constructor cause problems when you have pointer data members?

Answer: A copy constructor allows you to create a new object as a copy of an existing instance. The default copy constructor creates the new object by copying the existing object, member by member, and thus when there are member pointer you end up with two objects pointing to the same object.

It is important to note that the copy constructor is

called every time a function receives an object via the
pass-by-value mechanism. This means that the copy con-
structor needs to be implemented using a pass by refer-
ence. Otherwise you will be recursively calling the copy
constructor. You should always set the parameter for a
copy constructor to be const.

```
ClassName( const ClassName& other );
ClassName( ClassName& other );
ClassName( volatile const ClassName& other);
ClassName( volatile ClassName& other );
```

Question 18. What is the output of the following code:

```
#include <iostream>
using namespace std;

class A
{
public:
    int * ptr;
    ~A()
    {
        delete(ptr);
    }
};

void foo(A object_input)
{
    ;
}

int main()
{
    A aa;
    aa.ptr = new int(2);
```

```
    foo(aa);
    cout<<(*aa.ptr)<<endl;
    return 0;
}
```

Answer: The output of the code is an uncertain number, depending on the compiler used; for some compilers it could generate an error. The reason for this is that we do not define our own copy constructor.

When we call the foo function, the compiler will generate a default copy constructor which will shallow copy every data members defined in class A. This will lead to the result that two pointers, one in temporary object and the other in the object aa, will point to the same area in memory. When we get out the foo function, the compiler will automatically call the destructor function of the temporary object in which the pointer will be deleting and the area it points to will be free. In this situation, the pointer in aa will still point to the same area which has been free. When we try to visit the data through the pointer in aa, we will get garbage information.

Question 19. How do you overload an operator?

Answer:

```
type operator symbol (parameter_list);
```

If you define the operator outside of the class, then it will be a global operator function.

Example:

```
struct FooClass{int a;};
int operator + (FooClass lhs, FooClass rhs) {
    return lhs.a + rhs.a;
}
```

Question 20. What are smart pointers?

Answer: A smart pointer is a class built to mimic a pointer (offering dereferencing, indirection, arithmetic) that also offers extra features to simplify the usage, sharing and management of resources.

C++11 comes with three implementations of smart pointers: `shared_ptr`, `unique_ptr`, and `weak_ptr`.

Example:

```
//shared_pointer maintains
//a reference count
//when the count is zero the object
//pointed to is destroyed
std::shared_ptr<int> foo(new int(3));
std::shared_ptr<int> bar = foo;

//memory not released
//bar is still in scope
foo.reset();

//releases the memory,
//since no one is using it
bar.reset();
}
```

Question 21. What is encapsulation?

Answer: Encapsulation is the ability to expose an interface while hiding implementation. This is usually achieved through access modifiers (public, private, protected, etc.).

Question 22. What is a polymorphism?

Answer: Polymorphism is the ability for a set of classes to all be referenced through a common interface.

Question 23. What is inheritance?

Answer: Inheritance is the ability for one class to extend another through sub-classing. This is also referred to as "white-box" (the opposite of "black-box") re-use. A library can provide base classes that may be extended by the application developer.

Question 24. What is a virtual function? What is a pure virtual function and when do you use it?

Answer: Virtual functions are functions that are resolved by the compiler, at runtime, to the most derived version with the same signature. This means that if a function that was defined using a base class `Foo`, with a virtual member function `f`, is called using an instance of a sub class `FooChild`, that function is going to be dynamically binded to the implementation of the sub class (regardless that the actual code only refers to the base class).

A pure virtual function is a virtual function with no implementation in the base class, making the base class abstract (and thus can't be instantiated). Derived classes are forced to override the pure virtual function if they want to be instantiated. You use the same syntax as the virtual function but add `=0` to its declaration within the class.

Question 25. Why are virtual functions used for destructors? Can they be used for constructors?

Answer: Destructors are recommended to be defined as virtual, so the proper destructor (in the class hierarchy) is called at running time.

When calling a constructor, the caller needs to know the exact type of the object to be created, and thus they cannot be virtual.

Question 26. Write a function that computes the factorial of a positive integer.

Answer:

```
//for implementation
int factorial(int n){
    int output =1;
    for (int i =2 ; i <= n ; ++i)
        output *= i;
    return output;
}

//recursive implementation
int factorial(int n){
    if (n == 0) return 1;
    return n*factorial(n-1);
}

//tail recursive implementation
int factorial(int n, int last = 1){
    if (n == 0) return last;
    return factorial(n-1, last * n);
}
```

Question 27. Write a function that takes an array and returns the subarray with the largest sum.

Answer:

```
#include <vector>
#include <algorithm> // std::max

using namespace std;

template <typename T>
T max_sub_array(vector<T> const & numbers){
    T max_ending = 0, max_so_far = 0;
    for(auto & number: numbers){
        max_ending = max(0, max_ending + number);
```

```
        max_so_far = max(max_so_far, max_ending);
    }
    return max_so_far;
}
```

Question 28. Write a function that returns the prime factors of a positive integer.

Answer:

```cpp
#include <vector>
using namepsace std;

vector<int> prime_factors(int n){
    vector<int> factors;
    for (int i = 2; i <= n/i ; ++i)
        while (n % i == 0) {
            factors.push_back(i);
            n /= i;
        }
    if (n > 1)
        factors.push_back(n);
    return factors;
}
```

Question 29. Write a function that takes a 64-bit integer and swaps the bits at indices i and j.

Answer:

```cpp
long swap_bits (long x, const int &i, const int &j){
    if ( ((x >>i) & 1L) != ((x>>j) & 1L) )
        x ^= (1L <<i ) | (1L <<j);
    return x;
}
```

Question 30. Write a function that reverses a single linked list.

Answer:

```cpp
#include <memory> // shared_ptr
using namespace std;

template<typename T>
struct node_t {
    T data;
    shared_ptr<node_T<T>> next;
};

//recursive implementation
template<typename T> shared_ptr<node_t<T>>
    reverse_linked_list(
        const shared_ptr<node_t <T>> &head){
    if (!head || !head->next) {
        return head;
    }

    shared_ptr<node_t<T>>
      new_head = reverse_linked_list(head->next);
    head->next->next = head;
    head->next = nullptr;
    return new_head;
}

//while implementation
template<typename T> shared_ptr<node_t<T>>
    reverse_linked_list(
        const shared_ptr<node_t <T>> &head){
    shared_ptr<node_t<T>>
        prev = nullptr, curr = head;
    while(curr) {
  shared_ptr<node_t<T>> temp = curr-> next;
  curr->next = prev;
  prev = curr;
  curr = temp;
```

```
    }
    return prev;
}
```

Question 31. Write a function that takes a string and returns true if its parenthesis are balanced.

Answer:

```
#include<string>
#include<stack>
using namespace std;

bool is_par_balanced(const string input)
{
    //"())())"=> false
    //"(a(dd)()(()))"=>true
    stack<char> par_stack;
    for(auto &c: input)
    {
        if(c==')')
        {
            if(par_stack.empty())
            return false;
            else if(par_stack.top()=='(')
            par_stack.pop();
        }
        else if(c=='(')
        par_stack.push(c);
    }
    return par_stack.empty();
}
```

Question 32. Write a function that returns the height of an arbitrary binary tree.

Answer:

```cpp
#include<memory> //std::shared_ptr
#include<algorithm> //std::max
using namespace std;

template <typename T>
struct BinaryTree {
    T data;
    shared_ptr<BinaryTree<T>> left, right;
};


template <typename T>
int height(
    const shared_ptr<BinaryTree<T>> &tree,
    int count = -1){
        if (!tree) return count;
        return max(
            height(tree->left, count + 1),
            height(tree->right, count +1));
}
```

3.5 Monte Carlo simulations. Numerical methods.

Question 1. How would you compute π using Monte Carlo simulations? What is the standard deviation of this method?

Answer: An Acceptance–Rejection type method can be used to approximate π as follows: generate N points uniformly in the $[-1, 1] \times [-1, 1]$ square and accept a point (x, y) if the point is in the unit disk $D(0, 1)$, i.e., if $x^2 + y^2 \leq 1$. If the number of accepted points is A, then the ratio $\frac{A}{N}$ converges in the limit as $N \to \infty$ to the ratio of the area of the unit disk to the area of the square, which is equal to $\frac{\pi}{4}$. Therefore,

$$\pi \approx \frac{4A}{N}.$$

The standard deviation of the method is $O\left(\frac{1}{\sqrt{N}}\right)$. We make this more precise by computing below the coefficient of $\frac{1}{\sqrt{N}}$ in $O\left(\frac{1}{\sqrt{N}}\right)$; see (3.79).

Let $U_1, U_2, \cdots$ be a sequence of independent identically distributed bivariate random variables uniformly distributed in $[-1, 1] \times [-1, 1]$. Denote by $\mathbf{1}_{D(0,1)}$ the indicator function of the unit disk $D(0, 1)$, i.e.,

$$\mathbf{1}_{D(0,1)}(x, y) = \begin{cases} 1, & \text{if } (x, y) \in D(0, 1); \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$X_i = \mathbf{1}_{D(0,1)}(U_i), \quad \forall\, i \geq 1.$$

The random variables X_i, $i \geq 1$, are independent and

identically distributed, with

$$E[X_i] = E\left[\mathbf{1}_{D(0,1)}(U_i)\right] = \iint\limits_{D(0,1)} \frac{1}{4} dx dy$$

$$= \frac{\pi}{4}. \tag{3.77}$$

Since X_i, $i \geq 1$, are integrable, it follows from the strong law of large numbers that

$$\lim_{n \to \infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\pi}{4} \quad \text{almost surely.}$$

Note that $X_1 + X_2 + \cdots + X_n$ counts how many points out of the randomly selected n points reside in the disk $D(0,1)$. Thus, for N large enough,

$$\frac{X_1 + X_2 + \cdots + X_N}{N} \approx \frac{\pi}{4}. \tag{3.78}$$

To calculate the variance of the estimation in (3.78), note that, for $1 \leq i \leq N$,

$$\left(\mathbf{1}_{D(0,1)}(U_i)\right)^2 = \mathbf{1}_{D(0,1)}(U_i), \quad \forall\, 1 \leq i \leq N,$$

and therefore, using (3.77), we find that

$$\begin{aligned}
\text{var}(X_i) &= E[X_i^2] - (E[X_i])^2 \\
&= E\left[\left(\mathbf{1}_{D(0,1)}(U_i)\right)^2\right] - \left(\frac{\pi}{4}\right)^2 \\
&= E\left[\mathbf{1}_{D(0,1)}(U_i)\right] - \frac{\pi^2}{16} \\
&= \frac{\pi}{4} - \frac{\pi^2}{16} \\
&= \frac{\pi(4-\pi)}{16}.
\end{aligned}$$

Then,

$$\text{var}\left(\frac{X_1 + X_2 + \cdots + X_N}{N}\right)$$

$$= \frac{1}{N^2}\left(\text{var}(X_1) + \cdots + \text{var}(X_N)\right)$$

$$= \frac{1}{N^2} \cdot N \,\frac{\pi(4 - \pi)}{16}$$

$$= \frac{\pi(4 - \pi)}{16N}.$$

By taking the square root of the variance, we conclude that the standard deviation of this Monte Carlo method for estimating π is

$$\frac{\sqrt{\pi(4 - \pi)}}{4}\,\frac{1}{\sqrt{N}} \approx 0.82\,\frac{1}{\sqrt{N}}. \qquad (3.79)$$

Question 2. What methods do you know for generating independent samples of the standard normal distribution?

Answer: The three most commonly used methods to generate independent standard normal samples are:

• Box–Muller (using the Marsaglia–Bray algorithm in order to avoid estimating trigonometric functions);

• Acceptance–Rejection;

• Inverse Transform.

For details on these methods, see Glasserman [2]. □

Question 3. How do you generate a geometric Brownian motion stock path using random numbers from a normal distribution?

Answer: Consider a stock whose price follows the geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \qquad (3.80)$$

where μ and σ are the drift and the volatility of the stock price, and W_t is a Weiner process. To generate a price path for the stock between time 0 and time T, discretize the time interval into m equal time steps of size $\delta t = \frac{T}{m}$, and let $t_j = j\delta t$, for $j = 0 : m$.

By integrating (3.80) between t_j and t_{j+1}, it follows that

$$S_{t_{j+1}} - S_{t_j} = \mu \int_{t_j}^{t_{j+1}} S_t dt + \sigma \int_{t_j}^{t_{j+1}} S_t dW_t. \quad (3.81)$$

We use the following approximations:

$$\int_{t_j}^{t_{j+1}} S_t dt \approx S_{t_j}(t_{j+1} - t_j)$$
$$= S_{t_j} \delt; \quad (3.82)$$
$$\int_{t_j}^{t_{j+1}} S_t dW_t \approx S_{t_j}(W_{t_{j+1}} - W_{t_j})$$
$$= S_{t_j} \sqrt{\delta t}\, Z_{j+1}, \quad (3.83)$$

where Z_{j+1} is a standard normal variable, since W_t is a Wiener process and therefore $W_{t_{j+1}} - W_{t_j}$ is a normal variable with mean 0 and variance $t_{j+1} - t_j = \delta t$, i.e.,

$$W_{t_{j+1}} - W_{t_j} = \sqrt{\delta t} Z_{j+1}. \quad (3.84)$$

Note that Z_j, for $j = 1 : m$, are independent standard normals.

If $z_1, z_2, \ldots, z_m$ are independent samples of the standard normal distribution, we obtain from (3.81–3.83) that (3.80) can be discretized as follows:

$$S_{t_{j+1}} - S_{t_j} = \mu S_{t_j} \delta t + \sigma S_{t_j} \sqrt{\delta t} z_{j+1},$$

for $j = 0 : (m - 1)$, which can be written as

$$S_{t_{j+1}} = S_{t_j} \left(1 + \mu \delta t + \sigma \sqrt{\delta t} z_{j+1}\right),$$

for $j = 0 : (m - 1)$.

Note that there is a very small probability that the price path above becomes negative, which is a drawback of using this discretization.

A price path which is always positive can be generated using Itô's formula to express (3.80) as

$$d(\ln(S_t)) = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW_t. \quad (3.85)$$

By integrating (3.85) between t_j and t_{j+1}, it follows that

$$\ln(S_{t_{j+1}}) - \ln(S_{t_j}) = \ln\left(\frac{S_{t_{j+1}}}{S_{t_j}}\right)$$

$$= \left(\mu - \frac{\sigma^2}{2}\right)(t_{j+1} - t_j) + \sigma(W_{t_{j+1}} - W_{t_j})$$

$$= \left(\mu - \frac{\sigma^2}{2}\right)\delta t + \sigma\sqrt{\delta t}Z_{j+1}, \quad (3.86)$$

for $j = 0 : (m - 1)$, where (3.84) was used for (3.86).

Then, (3.86) can be discretized as follows:

$$\ln\left(\frac{S_{t_{j+1}}}{S_{t_j}}\right) = \left(\mu - \frac{\sigma^2}{2}\right)\delta t + \sigma\sqrt{\delta t}z_{j+1},$$

for $j = 0 : (m - 1)$, and therefore

$$S_{t_{j+1}} = S_{t_j} \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)\delta t + \sigma\sqrt{\delta t}z_{j+1}\right),$$

for all $j = 0 : (m - 1)$. $\quad\square$

Question 4. How do you generate a sample of the standard normal distribution from 12 independent samples of the uniform distribution on $[0, 1]$?

Answer: If $u_1, u_2, \ldots, u_{12}$ are 12 independent samples of the uniform distribution on $[0, 1]$, then

$$\sum_{i=1}^{12} u_i - 6 \quad (3.87)$$

can be used as a sample of the standard normal distribution.

To see this, recall from the Central Limit Theorem that, if X_i, $i \geq 1$, are independent identically distributed random variables with finite expected value $E[X]$ and standard deviation $\sigma(X)$, then

$$\lim_{n \to \infty} \frac{\frac{1}{n} \left(\sum_{i=1}^{n} X_i \right) - E[X]}{\frac{\sigma(X)}{\sqrt{n}}} = Z, \qquad (3.88)$$

where Z is the standard normal distribution and the convergence in (3.88) is in distribution.

Let $U_1, U_2, \ldots, U_{12}$ be 12 independent uniform distributions on $[0, 1]$. Using (3.88) for $n = 12$ and $X_i = U_i$, $i = 1 : 12$, we infer that

$$Z \approx \frac{\frac{1}{12} \left(\sum_{i=1}^{12} U_i \right) - E[U]}{\frac{\sigma(U)}{\sqrt{12}}}, \qquad (3.89)$$

where

$$E[U] = \int_0^1 u \, du = \frac{1}{2}; \qquad (3.90)$$

$$\sigma^2(U) = E[U^2] - (E[U])^2 = \int_0^1 u^2 du - \left(\frac{1}{2} \right)^2$$

$$= \frac{1}{3} - \frac{1}{4} = \frac{1}{12},$$

and thus

$$\sigma(U) = \frac{1}{\sqrt{12}}. \qquad (3.91)$$

From (3.89–3.91), we obtain that

$$Z \approx \frac{\frac{1}{12} \left(\sum_{i=1}^{12} U_i \right) - \frac{1}{2}}{\frac{1}{12}}$$

$$= \sum_{i=1}^{12} U_i - 6,$$

and therefore (3.87) can be used as an approximate sample of the standard normal distribution.

Note that this is an inefficient method, since it uses 12 samples from the uniform distribution to generate one approximate sample of the standard normal distribution, and all the samples that it generates are in the interval $[-6, 6]$. More efficient methods for generating independent standard normal samples are Box–Muller, which uses two uniform distribution samples to generate two samples of the standard normal distribution, and Acceptance–Rejection. □

Question 5. What is the rate of convergence for Monte Carlo methods?

Answer: If n is the number of paths in the Monte Carlo simulation and m is the number of time steps between 0 and T used in the discretization of each path, then the convergence rate of the Monte Carlo simulation is

$$O\left(\max\left(\frac{T}{m}, \frac{1}{\sqrt{n}}\right)\right).$$

The estimate above holds for Monte Carlo simulations on multi asset derivative securities, i.e., is independent of the number of underlying assets of the derivative security, unlike finite differences and numerical integration methods, where the convergence slows down as the number of underlying assets increases. □

Question 6. What variance reduction techniques do you know?

Answer: The variance reduction techniques are used to reduce the constant factor corresponding to the Monte Carlo approximation error $O\left(\frac{1}{\sqrt{n}}\right)$. Some of the most commonly used variance reduction techniques are:

• Control Variates;

- Antithetic Variables;

- Moment Matching.

For details on these methods and their implementation, see Glasserman [2]. □

Question 7. How do you generate samples of normal random variables with correlation ρ?

Answer: Assume that you can generate two samples z_1 and z_2 from two independent standard normal variables Z_1 and Z_2. Let $X_1 = Z_1$ and $X_2 = \rho Z_1 + \sqrt{1 - \rho^2} Z_2$. Note that X_2 is a linear combination of independent normal variables, and therefore X_2 is a normal variable as well. Also,

$$
\begin{aligned}
\text{corr}(X_1, X_2) &= \text{corr}(Z_1, \rho Z_1 + \sqrt{1 - \rho^2} Z_2) \\
&= \rho \, \text{corr}(Z_1, Z_1) \\
&\quad + \sqrt{1 - \rho^2} \, \text{corr}(Z_1, Z_2) \\
&= \rho \, \text{var}(Z_1) \\
&= \rho,
\end{aligned}
$$

since $\text{var}(Z_1) = 1$ and $\text{corr}(Z_1, Z_2) = 0$, since Z_1 and Z_2 are independent.

We conclude that X_1 and X_2 are normal random variables with correlation ρ. Thus, starting with two independent standard normal samples z_1 and z_2,

$$
x_1 = z_1 \quad \text{and} \quad x_2 = \rho z_1 + \sqrt{1 - \rho^2} z_2
$$

are samples of normal random variables with correlation ρ. □

Question 8. What is the order of convergence of the Newton's method?

Answer: If it is convergent, Newton's method is quadratically (second order) convergent.

Recall that, given an initial guess x_0, the Newton's method recursion for solving $f(x) = 0$, where $f : \mathbb{R} \to \mathbb{R}$, is

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \ \forall \ k \geq 0. \quad (3.92)$$

The quadratic convergence of Newton's method can be stated formally as follows: Let x^* be a solution of $f(x) = 0$. If $f(x)$ is a twice differentiable function with $f''(x)$ continuous, if $f'(x^*) \neq 0$, and if x_0 is close enough to x^*, then there exists $M > 0$ and n_M a positive integer such that

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} < M, \ \forall \ k \geq n_M. \quad (3.93)$$

To provide the intuition behind (3.93), note that, since $f(x^*) = 0$, the recursion (3.92) can be written as

$$
\begin{aligned}
&x_{k+1} - x^* \\
&= \ x_k - x^* - \frac{f(x_k) - f(x^*)}{f'(x_k)} \\
&= \ \frac{f(x^*) - f(x_k) + (x_k - x^*)f'(x_k)}{f'(x_k)}. \quad (3.94)
\end{aligned}
$$

Recall from the linear Taylor expansion of the function $f(x)$ around the point x_k that, if $f''(x)$ is continuous, then there exists a point c_k between x^* and x_k such that

$$
\begin{aligned}
f(x^*) \ = \ & f(x_k) + (x^* - x_k)f'(x_k) \\
& + \frac{(x^* - x_k)^2}{2} f''(c_k). \quad (3.95)
\end{aligned}
$$

From (3.94) and (3.95), we find that

$$x_{k+1} - x^* = (x^* - x_k)^2 \frac{f''(c_k)}{2f'(x_k)},$$

and conclude that

$$\frac{|x_{k+1} - x^*|}{|x_k - x^*|^2} = \left| \frac{f''(c_k)}{2f'(x_k)} \right|. \quad (3.96)$$

If $f''(x)$ and $f'(x)$ are continuous functions such that $f'(x^*) \neq 0$, it follows that, if x_k is close to x^*, then $\left| \frac{f''(c_k)}{2f'(x_k)} \right|$ is close to $\left| \frac{f''(x^*)}{2f'(x^*)} \right| < \infty$. Therefore, the term $\left| \frac{f''(c_k)}{2f'(x_k)} \right|$ is uniformly bounded if x_k is close enough to x^*, and (3.96) can be written formally as (3.93). $\square$

Question 9. Which finite difference method corresponds to trinomial trees?

Answer: Forward Euler. As an explicit finite difference method, the Forward Euler discretization of the Black–Scholes PDE gives the finite difference value of the option at a node as a linear combination of the option values at three nodes on the prior time step, which is similar to the risk neutral formula for trinomial trees.

If the calibration of the trinomial trees is done in the log space, i.e., if the up and down factors are calibrated to the normal distribution of $\ln(S)$, and if the Forward Euler discretization is done for the constant coefficients PDE obtained by the change of variables $x = \ln(S)$, the classical trinomial tree recursion and the Forward Euler recursion are almost identical. $\square$

3.6 Probability. Stochastic calculus.

Question 1. What is the exponential distribution? What are the mean and the variance of the exponential distribution?

Answer: The density function of the exponential random variable X with parameter $\alpha > 0$ is

$$f(x) = \begin{cases} \alpha\, e^{-\alpha x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

The expected value and the variance of the exponential random variable X are

$$E[X] = \frac{1}{\alpha} \quad \text{and} \quad \text{var}(X) = \frac{1}{\alpha^2}.$$

To see this, use integration by parts to find that

$$\begin{aligned} \int_0^\infty x e^{-\alpha x} dx &= -\frac{x e^{-\alpha x}}{\alpha}\Big|_0^\infty + \frac{1}{\alpha}\int_0^\infty e^{-\alpha x} dx \\ &= 0 - \frac{e^{-\alpha x}}{\alpha^2}\Big|_0^\infty = \frac{1}{\alpha^2}; \\ \int_0^\infty x^2 e^{-\alpha x} dx &= -\frac{x^2 e^{-\alpha x}}{\alpha}\Big|_0^\infty + \frac{2}{\alpha}\int_0^\infty x e^{-\alpha x} dx \\ &= 0 + \frac{2}{\alpha}\times\frac{1}{\alpha^2} = \frac{2}{\alpha^3}. \end{aligned}$$

Then,

$$E[X] = \int_{-\infty}^\infty x f(x)\, dx = \alpha \int_0^\infty x e^{-\alpha x}\, dx = \frac{1}{\alpha};$$

$$E[X^2] = \int_{-\infty}^\infty x^2 f(x)\, dx = \alpha \int_0^\infty x^2 e^{-\alpha x}\, dx = \frac{2}{\alpha^2}.$$

Therefore,

$$\text{var}(X) = E[X^2] - (E[X])^2 = \frac{2}{\alpha^2} - \left(\frac{1}{\alpha}\right)^2 = \frac{1}{\alpha^2}. \quad \square$$

Question 2. If X and Y are independent exponential random variables with mean 6 and 8, respectively, what is the probability that Y is greater than X?

Answer: The probability density functions of X and Y are, respectively:

$$f_X(x) = \begin{cases} \frac{1}{6}e^{-\frac{x}{6}}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0; \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1}{8}e^{-\frac{y}{8}}, & \text{if } y \geq 0; \\ 0, & \text{if } y < 0. \end{cases}$$

Since X and Y are independent, the joint probability density function $f_{XY}(x,y)$ of (X,Y) is the product of the marginal probability density functions, i.e.,

$$\begin{aligned} f_{XY}(x,y) &= f_X(x)\, f_Y(y) \\ &= \begin{cases} \frac{1}{48}e^{-\frac{x}{6}-\frac{y}{8}}, & \text{if } x \geq 0,\, y \geq 0; \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Let

$$A = \{(x,y) \in \mathbb{R}^2 : y \geq x\}.$$

The probability that Y is greater than X can be found by evaluating the double integral

$$\begin{aligned} P(Y \geq X) &= \iint_A f_{XY}(x,y)\,dxdy \\ &= \frac{1}{48}\int_0^\infty \int_0^y e^{-\frac{x}{6}-\frac{y}{8}}\,dxdy \end{aligned}$$

as follows:

$$
\begin{aligned}
P(Y \geq X) &= \frac{1}{48} \int_0^\infty e^{-\frac{y}{8}} \left(\int_0^y e^{-\frac{x}{6}} dx \right) dy \\
&= \frac{1}{48} \int_0^\infty e^{-\frac{y}{8}} \left(-6e^{-\frac{x}{6}} \Big|_0^y \right) dy \\
&= \frac{1}{8} \int_0^\infty e^{-\frac{y}{8}} \left(1 - e^{-\frac{y}{6}} \right) dy \\
&= \frac{1}{8} \int_0^\infty \left(e^{-\frac{y}{8}} - e^{-\frac{7}{24}y} \right) dy \\
&= \frac{1}{8} \left(8 - \frac{24}{7} \right) \\
&= \frac{4}{7}. \quad \square
\end{aligned}
$$

Question 3. What are the expected value and the variance of the Poisson distribution?

Answer: A Poisson distribution is a random variable X taking nonnegative integer values with probabilities

$$
P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad \forall \, k \geq 0,
$$

where $\lambda > 0$ is a fixed positive number.

We show that the expected value and the variance of the Poisson distribution X are

$$
E[X] = \lambda \quad \text{and} \quad \text{var}(X) = \lambda.
$$

By definition,

$$
\begin{aligned}
E[X] &= \sum_{k=0}^\infty P(X = k) \cdot k = \sum_{k=1}^\infty \frac{e^{-\lambda} \lambda^k}{k!} k \\
&= e^{-\lambda} \lambda \sum_{k=1}^\infty \frac{\lambda^{k-1}}{(k-1)!}. \quad (3.97)
\end{aligned}
$$

Since the Taylor series expansion for e^t is

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!},$$

it follows that

$$\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{\lambda}; \tag{3.98}$$

$$\sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = e^{\lambda}. \tag{3.99}$$

From (3.97) and (3.98), we find that $E[X] = \lambda$.

To calculate $\text{var}(X)$, note that

$$\begin{aligned}
E[X^2] &= \sum_{k=0}^{\infty} P(X = k) \cdot k^2 \\
&= \sum_{k=1}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} k^2 \\
&= e^{-\lambda} \sum_{k=1}^{\infty} \frac{k \lambda^k}{(k-1)!},
\end{aligned}$$

which can be written as

$$\begin{aligned}
E[X^2] &= e^{-\lambda} \sum_{k=1}^{\infty} \frac{(k-1)\lambda^k}{(k-1)!} + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\
&= e^{-\lambda} \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda^2 + \lambda,
\end{aligned}$$

where (3.98) and (3.99) were used for the last equality.

We conclude that

$$\text{var}(X) = E[X^2] - (E[X])^2 = \lambda. \quad \square$$

Question 4. A point is chosen uniformly from the unit disk. What is the expected value of the distance between the point and the center of the disk?

Answer: The expected value of the distance between a uniformly chosen point in the unit disk D and the center of the disk can be computed as $E\left[\sqrt{X^2 + Y^2}\right]$, where (X, Y) is uniformly distributed in the unit disk D. The probability density function of (X, Y) is

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & \text{if } x \in D; \\ \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$E\left[\sqrt{X^2 + Y^2}\right] = \frac{1}{\pi} \iint\limits_D \sqrt{x^2 + y^2} \, dx dy. \qquad (3.100)$$

Using the polar coordinates substitution $x = r \cos\theta$ and $y = r \sin\theta$, with $0 \leq r \leq 1$ and $0 \leq \theta < 2\pi$, and recalling that $dx dy = r dr d\theta$, we obtain from (3.100) that

$$\begin{aligned} & E\left[\sqrt{X^2 + Y^2}\right] \\ &= \frac{1}{\pi} \int_0^{2\pi} \int_0^1 \sqrt{r^2 \left(\cos^2\theta + \sin^2\theta\right)} \, r dr d\theta \\ &= \frac{1}{\pi} \int_0^1 \int_0^{2\pi} r^2 d\theta \, dr \qquad (3.101) \\ &= \frac{1}{\pi} \int_0^1 2\pi r^2 \, dr = 2 \int_0^1 r^2 dr \\ &= \frac{2}{3}, \end{aligned}$$

where for (3.101) we used the fact that $\cos^2\theta + \sin^2\theta = 1$ for all θ. $\qquad \square$

Question 5. Consider two random variables X and Y with mean 0 and variance 1, and with joint normal dis-

tribution. If $\text{cov}(X, Y) = \frac{1}{\sqrt{2}}$, what is the conditional probability $P(X > 0 | Y < 0)$?

Answer: From the definition of conditional probability, it follows that

$$P(X > 0 | Y < 0) = \frac{P(X > 0, Y < 0)}{P(Y < 0)}. \qquad (3.102)$$

Note that

$$P(Y < 0) = \frac{1}{2}, \qquad (3.103)$$

since Y is a standard normal random variable.

In order to compute $P(X > 0, Y < 0)$, let

$$W = \sqrt{2}\, X - Y. \qquad (3.104)$$

Since $E[X] = E[Y] = 0$, it follows that $E[W] = 0$. Moreover, since

$$\text{var}(X) = \text{var}(Y) = 1 \ \ \text{and} \ \ \text{cov}(X, Y) = \frac{1}{\sqrt{2}},$$

we obtain that

$$\begin{aligned}
\text{var}(W) &= \text{var}\left(\sqrt{2}\, X - Y\right) \\
&= \text{var}\left(\sqrt{2}\, X\right) - 2\,\text{cov}\left(\sqrt{2}X, Y\right) + \text{var}(Y) \\
&= 2\text{var}(X) - 2\sqrt{2}\,\text{cov}(X, Y) + \text{var}(Y) \\
&= 1,
\end{aligned}$$

and

$$\begin{aligned}
\text{cov}(W, Y) &= \text{cov}\left(\sqrt{2}\, X - Y, Y\right) \\
&= \sqrt{2}\,\text{cov}(X, Y) - \text{var}(Y) \\
&= 0.
\end{aligned}$$

Note that $W = \sqrt{2}\, X - Y$ is a normal random variable since X and Y have joint normal distribution. Moreover,

since $E[W] = 0$, $\text{var}(W) = 1$, and $\text{cov}(W, Y) = 0$, it follows that W and Y are independent standard normal variables.

From (3.104), we find that

$$X = \frac{1}{\sqrt{2}}(W + Y).$$

Then, the probability of the event $\{X > 0, Y < 0\}$ can be written as

$$
\begin{aligned}
& P(X > 0, Y < 0) \\
= \ & P\left(\frac{1}{\sqrt{2}}(W + Y) > 0, Y < 0\right) \\
= \ & P(W + Y > 0, Y < 0).
\end{aligned} \tag{3.105}
$$

The two straight lines $w + y = 0$ and $y = 0$ cut the (w, y) plane into four wedges:

$$
\begin{aligned}
R_1 &= \{w + y > 0, y < 0\}; \\
R_2 &= \{w + y > 0, y > 0\}; \\
R_3 &= \{w + y < 0, y < 0\}; \\
R_4 &= \{w + y < 0, y > 0\}.
\end{aligned}
$$

Note that

$$P(W + Y > 0, Y < 0) = P((W, Y) \in R_1). \tag{3.106}$$

Since W and Y are independent normal random variables, their joint probability density function is rotationally symmetric, and therefore

$$
\begin{aligned}
P((W, Y) \in R_1) &= P((W, Y) \in R_4); & (3.107) \\
P((W, Y) \in R_2) &= P((W, Y) \in R_3); & (3.108) \\
P((W, Y) \in R_2) &= 3P((W, Y) \in R_1);
\end{aligned}
$$
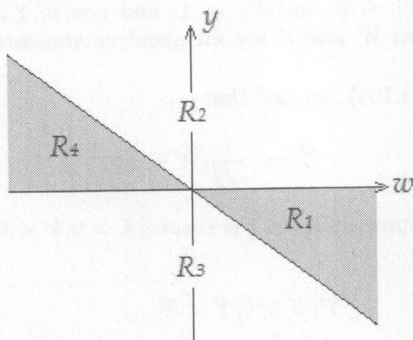
see Figure 3.1.

Figure 3.1: The regions R_1 to R_4 in the (w, y) plane.

Also, note that

$$\sum_{i=1}^{4} P((W, Y) \in R_i) = 1, \qquad (3.109)$$

since $P(W + Y = 0 \text{ or } Y = 0) = 0$.

From (3.107–3.109), we find that

$$
\begin{aligned}
1 &= \sum_{i=1}^{4} P((W, Y) \in R_i) \\
  &= 2P((W, Y) \in R_1) + 2P((W, Y) \in R_2) \\
  &= 8P((W, Y) \in R_1),
\end{aligned}
$$

and therefore

$$P((W, Y) \in R_1) = \frac{1}{8}.$$

Thus,

$$P(W + Y > 0, Y < 0) = \frac{1}{8}; \qquad (3.110)$$

see (3.106).

Then, from (3.105) and (3.110), it follows that

$$P(X > 0, Y < 0) = \frac{1}{8}. \qquad (3.111)$$

From (3.102), (3.103), and (3.111), we conclude that

$$P(X > 0 | Y < 0) = \frac{P(X > 0, Y < 0)}{P(Y < 0)}$$

$$= \frac{1}{4}. \quad \square$$

Question 6. If X and Y are lognormal random variables, is their product XY lognormally distributed?

Answer: First, note that, if X and Y are independent lognormal random variables, then XY is lognormally distributed, since $\ln(XY) = \ln(X) + \ln(Y)$ is the sum of two independent normal random variables, and therefore it is normally distributed.

In a more general setting, if $\ln(X)$ and $\ln(Y)$ have joint normal distribution, then $\ln(X) + \ln(Y)$ is normally distributed and therefore XY is a lognormal random variable.

Otherwise, $\ln(X) + \ln(Y)$ may not be normally distributed even if $\ln(X)$ and $\ln(Y)$ are normally distributed, in which case XY is not lognormally distributed. $\quad \square$

Question 7. Let X be a normal random variable with mean μ and variance σ^2, and let Φ be the cumulative distribution function of the standard normal distribution. Find the expected value of $Y = \Phi(X)$.

Answer: Let Z be a standard normal random variable independent of X. Then,

$$Y = \Phi(X) = P(Z \le X | X) = E[\mathbf{1}_{Z \le X} | X],$$

and therefore

$$E[Y] = E[E[\mathbf{1}_{Z \leq X} | X]]. \tag{3.112}$$

Recall from the Tower Property for conditional expectation[10] that, for any two random variables T and W,

$$E[T] = E[E[T|W]]. \tag{3.113}$$

Using (3.113) for $T = \mathbf{1}_{Z \leq X}$ and $W = X$, we obtain that

$$\begin{aligned} E[E[\mathbf{1}_{Z \leq X} | X]] &= E[\mathbf{1}_{Z \leq X}] \\ &= P(Z \leq X). \end{aligned} \tag{3.114}$$

From (3.112) and (3.114), we obtain that

$$E[Y] = P(Z \leq X). \tag{3.115}$$

Recall that X is a normal random variable with mean μ and variance σ^2, and that X and Z are independent. Then, $Z - X$ is a normal random variable with the following mean and variance:

$$\begin{aligned} E[Z - X] &= -\mu; \\ \mathrm{var}(Z - X) &= \mathrm{var}(Z) + \mathrm{var}(X) = 1 + \sigma^2. \end{aligned}$$

and $Z - X$ is a normal random variable with mean $-\mu$ and variance $1 + \sigma^2$. Thus, $\frac{Z - X + \mu}{\sqrt{1 + \sigma^2}}$ is a standard normal random variable and therefore

$$\begin{aligned} P(Z \leq X) &= P(Z - X \leq 0) \\ &= P\left(\frac{Z - X + \mu}{\sqrt{1 + \sigma^2}} \leq \frac{\mu}{\sqrt{1 + \sigma^2}} \right) \\ &= \Phi\left(\frac{\mu}{\sqrt{1 + \sigma^2}} \right). \end{aligned} \tag{3.116}$$

---

[10]In other words, to calculate the expected value of T, one can first calculate the conditional expected value of T knowing the extra information from W, then average out the resulting conditional expected value over W.

From (3.115) and (3.116) we conclude that

$$E[Y] = \Phi\left(\frac{\mu}{\sqrt{1+\sigma^2}}\right). \tag{3.117}$$

We note that, if X is the standard normal variable, then $Y = \Phi(X)$ is uniformly distributed in the interval $[0,1]$, and therefore $E[Y] = \frac{1}{2}$. Note that, for $\mu = 0$ and $\sigma = 1$ in (3.117), i.e., if X is a standard normal variable, then $E[Y] = \Phi(0) = \frac{1}{2}$, which is consistent to the comment above. $\square$

Question 8. What is the law of large numbers?

Answer: There is a strong law of large numbers and there is a weak law of large numbers. The strong law of large numbers states that the average of a large number of independent identically distributed integrable random variables converges almost surely to their common mean; in the case of the weak law of large numbers, the convergence is only in probability.

More precisely, let X_1, X_2, ... be a sequence of independent identically distributed random variables with finite expected value $\mu = E[X_i]$, and let $S_n = X_1 + \cdots + X_n$.

The strong law of large numbers states that $\frac{S_n}{n} \rightarrow \mu$ almost surely, i.e.,

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu\right) = 1. \tag{3.118}$$

The weak law of large numbers states that $\frac{S_n}{n} \rightarrow \mu$ in probability, i.e.,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) = 0, \quad \forall\, \epsilon > 0. \tag{3.119}$$

Note that, if a sequence of random variables convergences almost surely, than it also converges in probability, and therefore, if (3.118) holds true, then (3.119) also holds

true. This is the sense in which the strong law of large numbers is "stronger" than the weak law of large numbers.
□

Question 9. What is the central limit theorem?

Answer: The central limit theorem states that the limiting distribution of the centered and scaled sum of an independent identically distributed sequence of random variables is a normal distribution if the common distribution of the random variables has finite variance.

More precisely, let X_1, X_2, ...be a sequence of independent identically distributed random variables with finite expected value $\mu = E[X_i]$ and finite variance $\sigma^2 = \text{var}[X_i]$. Let $S_n = X_1 + \cdots + X_n$. Then,

$$\lim_{n \to \infty} \frac{S_n - n\mu}{\sigma\sqrt{n}} = Z,$$

where Z is the standard normal distribution, and the convergence is in distribution, i.e.,

$$\lim_{n \to \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) = P(Z \leq t).$$

Putting together the Law of Large Numbers and the Central Limit Theorem, the following approximation as n goes to infinity holds:

$$\frac{S_n}{n} \approx \mu + \frac{\sigma}{\sqrt{n}} Z. \quad \square$$

Question 10. What is a martingale? How is it related to option pricing?

Answer: Let $(\Omega, \mathcal{F}_t, P)$ be a filtered probability space, where Ω is the sample space, $\mathcal{F}_t$ is a filtration, and P is a probability measure on Ω. A stochastic process X_t

is called a *martingale* with respect to the filtration $\{\mathcal{F}_t\}$ if and only if

(i) X_t is adapted, i.e., X_t is $\mathcal{F}_t$-measurable for all t;

(ii) X_t is integrable for all t, i.e., $E[|X_t|] < \infty$ for all t;

(iii) $E[X_t|\mathcal{F}_s] = X_s$ almost surely for all $s < t$.

In other words, a martingale is a stochastic process in which, given the available information $\mathcal{F}_s$ up to the current time s, the optimal estimation (in the least square sense) of the process in the future time t, i.e., $E[X_t|\mathcal{F}_s]$, is the current value X_s almost surely.

The martingale concept is one of the cornerstones of option pricing theory. The fundamental theorem of asset pricing states that, if a market model is arbitrage–free, then there exists a risk neutral probability so that each discounted asset price process under the risk neutral probability is a martingale. Thus, a way to price a derivative security is to figure out a partial differential equation, called the pricing equation, usually deduced by applying Itô's formula, so that the discounted price process of the derivative is a martingale under risk neutral probability.

Question 11. Explain the assumption $(dW_t)^2 = dt$ used in the informal derivation of Itô's Lemma.

Answer: The notation in differential form $(dW_t)^2 = dt$ is a shorthand for the conventional integral notation used in Riemann integral

$$\int_0^T (dW_t)^2 = \int_0^T dt. \qquad (3.120)$$

The intuition behind (3.120) is related to the *quadratic variation* of a Brownian path. By definition, the quadratic variation $QV_W[0, T]$ of the Brownian path W in the interval $[0, T]$ is

$$QV_W[0, T] = \lim_{n \to \infty} \sum_{i=1}^n \left| W_{t_i} - W_{t_{i-1}} \right|^2,$$

where $t_i = \frac{i}{n}T$, for $i = 0 : n$. Therefore, in expectation, we have that

$$E\left[\lim_{n\to\infty}\sum_{i=1}^{n}\left|W_{t_i}-W_{t_{i-1}}\right|^2\right]$$

$$=\lim_{n\to\infty}\sum_{i=1}^{n}E\left[\left|W_{t_i}-W_{t_{i-1}}\right|^2\right]$$

$$=\lim_{n\to\infty}\sum_{i=1}^{n}(t_i-t_{i-1})$$

$$=T,$$

since $W_{t_i} - W_{t_{i-1}} \sim N(0, t_i - t_{i-1})$ for $i = 1 : n$. In fact, it can be shown that the convergence

$$\sum_{i=1}^{n}\left|W_{t_i}-W_{t_{i-1}}\right|^2 \longrightarrow T \text{ as } n \to \infty$$

is in L^2 sense, i.e.,

$$\lim_{n\to\infty}E\left[\left|\sum_{i=1}^{n}\left|W_{t_i}-W_{t_{i-1}}\right|^2-T\right|^2\right]=0.$$

By mimicking the notations used in the conventional Riemann integral, we write that

$$\int_0^T (dW_t)^2 = \lim_{n\to\infty}\sum_{i=1}^{n}\left|W_{t_i}-W_{t_{i-1}}\right|^2 = T$$

$$= \int_0^T dt,$$

whose shorthand notation in differential form is

$$(dW_t)^2 = dt. \quad \square$$

Question 12. If W_t is a Wiener process, find $E[W_t W_s]$.

Answer: Assume that $s \leq t$. Write W_t as $W_t = (W_t - W_s) + W_s$, and note that

$$W_t W_s = (W_t - W_s)W_s + W_s^2. \qquad (3.121)$$

Since W_τ, $\tau \geq 0$, is a Wiener process, it follows that $W_t - W_s$ and W_s are independent normal random variables of mean 0, and therefore

$$\begin{aligned} E[(W_t - W_s)W_s] &= E[W_t - W_s]E[W_s] \\ &= 0. \end{aligned} \qquad (3.122)$$

Also, W_s is normal of mean $E[W_s] = 0$ and variance $\text{var}(W_s) = s$. Thus,

$$\begin{aligned} \text{var}(W_s) &= E[W_s^2] - (E[W_s])^2 \\ &= E[W_s^2], \end{aligned}$$

and therefore

$$E[W_s^2] = \text{var}(W_s) = s. \qquad (3.123)$$

From (3.121–3.123), we obtain that

$$\begin{aligned} E[W_t W_s] &= E[(W_t - W_s)W_s] + E[W_s^2] \\ &= s. \end{aligned} \qquad (3.124)$$

Since (3.124) was derived under the assumption $s \leq t$, we conclude that

$$E[W_t W_s] = \min(s, t). \quad \square \qquad (3.125)$$

Question 13. If W_t is a Wiener process, what is $\text{var}(W_t + W_s)$?

Answer: Assume that $s \leq t$. Write W_t as $W_t = (W_t - W_s) + W_s$, and note that $W_t + W_s = (W_t - W_s) + 2W_s$.

Then,

$$
\begin{aligned}
\text{var}(W_t + W_s) &= \text{var}(W_t - W_s) \\
&\quad + 4\text{cov}(W_t - W_s, W_s) \\
&\quad + 4\text{var}(W_s).
\end{aligned}
\tag{3.126}
$$

Since W_τ, $\tau \geq 0$, is a Wiener process, it follows that

$$
\begin{aligned}
\text{cov}(W_t - W_s, W_s) &= 0; &&(3.127) \\
\text{var}(W_t - W_s) &= t - s; &&(3.128) \\
\text{var}(W_s) &= s, &&(3.129)
\end{aligned}
$$

since $W_t - W_s$ and W_s are independent normal variables of variance $t - s$ and s, respectively.

From (3.126–3.129), we obtain that

$$
\text{var}(W_t + W_s) = (t - s) + 4s = t + 3s.
\tag{3.130}
$$

Since (3.130) was derived under the assumption $s \leq t$, we conclude that

$$
\text{var}(W_t + W_s) = \max(s, t) + 3\min(s, t). \quad \square
$$

Question 14. Let W_t be a Wiener process. Find Find

$$
\int_0^t W_s \, dW_s \quad \text{and} \quad E\left[\int_0^t W_s \, dW_s\right].
$$

Answer: Since $\int x \, dx = \frac{x^2}{2} + C$, we begin by computing $d\left(\frac{W_t^2}{2}\right)$. Recall from Itô's lemma that, if $f(x, t)$ is a continuously differentiable function, then

$$
df = \frac{\partial f}{\partial t} \, dt + \frac{\partial f}{\partial x} \, dW_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} \, dt.
$$

For $f(x, t) = \frac{x^2}{2}$, we find that

$$
d\left(\frac{W_t^2}{2}\right) = W_t dW_t + \frac{1}{2}dt,
$$

and therefore

$$d\left(\frac{W_t^2 - t}{2}\right) = W_t dW_t. \qquad (3.131)$$

By integrating (3.131) between 0 and t, and since $W_0 = 0$, we obtain that

$$\int_0^t W_s \, dW_s = \frac{W_t^2 - t}{2}, \quad \forall \, t \geq 0. \qquad (3.132)$$

From (3.132), it follows that

$$E\left[\int_0^t W_s \, dW_s\right] = \frac{E[W_t^2] - t}{2} = 0,$$

since W_t is normally distributed with mean 0 and variance t and therefore

$$\begin{aligned}
E[W_t^2] &= \text{var}(W_t) + (E[W_t])^2 \\
&= t. \quad \square
\end{aligned}$$

Question 15. Find the distribution of the random variable

$$X = \int_0^1 W_t dW_t.$$

Answer: Recall from Itô's formula that, if W_t is a Wiener process and $f(x)$ is a function with continuous second order derivative, then

$$df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt. \qquad (3.133)$$

For $f(x) = x^2$, we obtain from (3.133) that

$$dW_t^2 = 2W_t dW_t + dt. \qquad (3.134)$$

By integrating (3.134) from 0 to 1, we obtain that

$$W_1^2 = 2 \int_0^1 W_t dW_t + 1 = 2X + 1. \qquad (3.135)$$

Note that W_1 is a standard normal random variable, and let $W_1 = Z$. By solving (3.135) for X, we find that

$$X = \frac{W_1^2 - 1}{2} = \frac{Z^2 - 1}{2}.$$

Let $f_X(x)$ and $F_X(x)$ be the probability density function of X and the cumulative distribution function of X, respectively.

Note that

$$P\left(X \le -\frac{1}{2}\right) = P(Z^2 \le 0) = 0.$$

Thus, if $x < -\frac{1}{2}$, then $F_X(x) = P(X \le x) = 0$ and therefore

$$f_X(x) = 0,$$

since $f_X(x) = F_X'(x)$.

If $x \ge -\frac{1}{2}$, then

$$
\begin{aligned}
F_X(x) &= P(X \le x) = P\left(\frac{Z^2 - 1}{2} \le x\right) \\
&= P\left(Z^2 \le 2x + 1\right) \\
&= P\left(-\sqrt{2x+1} \le Z \le \sqrt{2x+1}\right) \\
&= 2 P\left(0 \le Z \le \sqrt{2x+1}\right) \\
&= \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{2x+1}} e^{-\frac{y^2}{2}} dy.
\end{aligned}
$$

By differentiating $F_X(x)$, it follows that, for $x < -\frac{1}{2}$,

$$\begin{aligned}
f_X(x) &= F_X'(x) \\
&= \sqrt{\frac{2}{\pi}} \frac{d}{dx} \left(\int_0^{\sqrt{2x+1}} e^{-\frac{y^2}{2}} dy \right) \\
&= \sqrt{\frac{2}{\pi}} e^{-\frac{(\sqrt{2x+1})^2}{2}} \left(\sqrt{2x+1} \right)' \\
&= \sqrt{\frac{2}{\pi}} \frac{e^{-\frac{2x+1}{2}}}{\sqrt{2x+1}}.
\end{aligned}$$

We conclude that

$$f_X(x) = \begin{cases} \sqrt{\frac{2}{\pi}} \dfrac{e^{-\frac{2x+1}{2}}}{\sqrt{2x+1}}, & \text{if } x > -\frac{1}{2}; \\ 0 & \text{if } x \le -\frac{1}{2}. \end{cases} \qquad \Box$$

Question 16. Let W_t be a Wiener process. Find the mean and the variance of

$$\int_0^t W_s^2 dW_s.$$

Answer: Let $\mathcal{B}_t$ be the Borel σ-algebra over the time interval $[0, t]$ and let $\mathcal{F}_t$ be the filtration for the probability space in which the Wiener process W_t resides. We will use the following results:[11]

Theorem 3.1. *(Martingality)*
Let f_s be progressively measurable and square integrable in $[0, T]$, i.e., f_t is $\mathcal{B}_t \otimes \mathcal{F}_t$-measurable for every $t \in [0, T]$ and $E\left[\int_0^T |f_t|^2 dt \right] < \infty$. Then, the stochastic integral

---

[11]For the proofs of Theorem 3.1 and Theorem 3.2, see Theorem 2.8 on p.65 and Theorem 3.1 on p.67, respectively, from Friedman [1].

$\int_0^t f_s dW_s$ defines a zero mean, square integrable martingale for $t \in [0, T]$. In particular,

$$E\left[\int_0^t f_s dW_s\right] = 0, \quad \forall t \in [0, T]. \qquad (3.136)$$

Theorem 3.2. *(Itô's isometry)*
Let f_t and g_t be progressively measurable and square integrable processes. Then,

$$E\left[\int_0^t f_s dW_s \int_0^t g_s dW_s\right] = \int_0^t E[f_s\, g_s]ds.$$

In particular, if $f_s = g_s$, it follows that

$$E\left[\left(\int_0^t f_s dW_s\right)^2\right] = \int_0^t E[f_s^2]ds. \qquad (3.137)$$

For our problem, we need to compute $E[X]$ and $\text{var}(X)$, where

$$X = \int_0^t W_s^2 dW_s.$$

We first check that the integrand W_s^2 is progressively measurable and square integrable in $[0, t]$.

Note that W_s^2 is progressively measurable because it is adapted and continuous.

Furthermore, since $W_s \sim N(0, s)$, it follows that $W_s = \sqrt{s}Z$, where Z is a standard normal random variable, and therefore

$$E[W_s^4] = E\left[\left(\sqrt{s}Z\right)^4\right] = s^2\, E[Z^4] = 3s^2, \qquad (3.138)$$

where we used the fact that the fourth moment of the standard normal distribution is 3, i.e., $E[Z^4] = 3$. From (3.138), it follows that

$$\int_0^t E[W_s^4]ds = \int_0^t 3s^2 ds = t^3 < \infty. \qquad (3.139)$$

In other words, W_s^2 is square integrable in $[0, t]$.

We can therefore apply both Theorems 3.1 and 3.2 with $f_s = W_s^2$ and $g_s = W_s^2$.

From (3.136) for $f_s = W_s^2$, we find that

$$E[X] = E\left[\int_0^t W_s^2 dW_s\right] = 0. \qquad (3.140)$$

From (3.137) for $f_s = g_s = W_s^2$, and using (3.139), it follows that

$$\begin{aligned} E[X^2] &= E\left[\left(\int_0^t W_s^2 dW_s\right)^2\right] = \int_0^t E[W_s^4] ds \\ &= t^3. \end{aligned}$$

Since $E[X] = 0$, see (3.140), we find that

$$\text{var}(X) = E[X^2] - (E[X])^2 = t^3. \quad \square$$

Question 17. If W_t is a Wiener process, find the variance of

$$X = \int_0^1 \sqrt{t}e^{\frac{W_t^2}{8}} dW_t.$$

Answer: We will use Theorems 3.1 and 3.2 to solve this problem. To do so, we first check that the integrand $\sqrt{t}e^{\frac{W_t^2}{8}}$ is progressively measurable and square integrable in $[0, 1]$.

The process $\sqrt{t}e^{\frac{W_t^2}{8}}$ is progressively measurable because it is adapted and continuous.

Furthermore, since $W_t \sim N(0,t)$, the probability density function of W_t is $\frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}}$, and therefore

$$
\begin{aligned}
E\left[e^{\frac{W_t^2}{4}}\right] &= \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty}e^{\frac{x^2}{4}}e^{-\frac{x^2}{2t}}\,dx \\
&= \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty}e^{\left(\frac{1}{4}-\frac{1}{2t}\right)x^2}\,dx \\
&= \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\cdot\frac{2-t}{2t}x^2}\,dx \\
&= \frac{1}{\sqrt{2\pi t}}\sqrt{2\pi\cdot\frac{2t}{2-t}} \qquad (3.141) \\
&= \sqrt{\frac{2}{2-t}}, \qquad (3.142)
\end{aligned}
$$

where, for (3.141), we used the identity

$$
\int_{-\infty}^{\infty}e^{-\frac{Ax^2}{2}}\,dx = \sqrt{\frac{2\pi}{A}},
$$

for any positive constant $A > 0$. From (3.142), it follows that

$$
\begin{aligned}
\int_0^1 E\left[te^{\frac{W_t^2}{4}}\right]dt &= \int_0^1 tE\left[e^{\frac{W_t^2}{4}}\right]dt \\
&= \int_0^1 t\cdot\sqrt{\frac{2}{2-t}}\,dt \\
&= -\frac{2\sqrt{2}}{3}\left((t+4)\sqrt{2-t}\right)\Big|_0^1 \\
&= \frac{2}{3}\left(8-5\sqrt{2}\right). \qquad (3.143)
\end{aligned}
$$

Thus, $\int_0^1 E\left[te^{\frac{W_t^2}{4}}\right]dt < \infty$, and we conclude that both Theorems 3.1 and 3.2 can be applied here.

From (3.136), it follows that

$$
E[X] = E\left[\int_0^1 \sqrt{t}e^{\frac{W_t^2}{8}}\,dW_t\right] = 0. \qquad (3.144)
$$

From (3.137), and using (3.143), we obtain that

$$
\begin{aligned}
E[X^2] &= E\left[\left(\int_0^1 \sqrt{t}e^{\frac{W_t^2}{8}}\,dW_t\right)^2\right] \\
&= \int_0^1 E\left[\left(\sqrt{t}e^{\frac{W_t^2}{8}}\right)^2\right]dt \\
&= \int_0^1 E\left[te^{\frac{W_t^2}{4}}\right]dt \\
&= \frac{2}{3}(8 - 5\sqrt{2}).
\end{aligned}
\tag{3.145}
$$

From (3.144) and (3.145), we conclude that

$$
\operatorname{var}(X) = E[X^2] - (E[X])^2 = \frac{2}{3}\left(8 - 5\sqrt{2}\right). \quad \square
$$

Question 18. If W_t is a Wiener process, what is $E\left[e^{W_t}\right]$?

Answer:

Solution 1: If W_t is a Wiener process and $Y = e^{W_t}$, then

$$
\ln(Y) = W_t \cong \sqrt{t}Z, \tag{3.146}
$$

where Z is the standard normal random variable, and therefore Y is a lognormal random variable. Recall that the expected value of a lognormal random variable given by $\ln(V) = \mu + \sigma Z$ is

$$
E[V] = e^{\mu + \frac{\sigma^2}{2}}. \tag{3.147}
$$

From (3.146), and using (3.147) with $\mu = 0$ and $\sigma = \sqrt{t}$, we conclude that

$$
E\left[e^{W_t}\right] = E[Y] = e^{\frac{t}{2}}.
$$

Solution 2: Since $W_t \cong \sqrt{t}Z$, it follows that

$$
\begin{aligned}
E\left[e^{W_t}\right] &= E\left[e^{\sqrt{t}Z}\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\sqrt{t}x}e^{-\frac{x^2}{2}}\,dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\sqrt{t}x-\frac{x^2}{2}}\,dx. \qquad (3.148)
\end{aligned}
$$

By completing the square, we find that

$$
\sqrt{t}x - \frac{x^2}{2} = -\frac{1}{2}\left(x-\sqrt{t}\right)^2 + \frac{t}{2},
$$

and therefore

$$
e^{\sqrt{t}x-\frac{x^2}{2}} = e^{\frac{t}{2}}e^{-\frac{1}{2}\left(x-\sqrt{t}\right)^2}. \qquad (3.149)
$$

From (3.148) and (3.149), it follows that

$$
\begin{aligned}
E\left[e^{W_t}\right] &= e^{\frac{t}{2}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(x-\sqrt{t}\right)^2}\,dx \\
&= e^{\frac{t}{2}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy \qquad (3.150) \\
&= e^{\frac{t}{2}}, \qquad\qquad\qquad\qquad (3.151)
\end{aligned}
$$

where we used the substitution $y = x - \sqrt{t}$ for (3.150), and, for (3.151), we used the fact that

$$
\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy = 1,
$$

since $\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$ is the probability density function of the standard normal distribution. $\square$

Question 19. If W_t is a Wiener process, find the variance of

$$
\int_0^t s\,dW_s.
$$

Answer: Recall that, for any deterministic square integrable function $f : [a, b] \to \mathbb{R}$, the stochastic integral $\int_a^b f(s) dW_s$ is normally distributed with mean 0 and variance equal to the square of the L^2 norm of f, i.e.,

$$\int_a^b f(s) dW_s \sim N\left(0, \int_a^b f^2(s) ds\right).$$

Then,

$$\int_0^t s \, dW_s \sim N\left(0, \int_0^t s^2 ds\right) = N\left(0, \frac{t^3}{3}\right),$$

and we conclude that

$$\text{var}\left(\int_0^t s dW_s\right) = \frac{t^3}{3}. \quad \square$$

Question 20. Let W_t be a Wiener process, and let

$$X_t = \int_0^t W_\tau d\tau. \tag{3.152}$$

What is the distribution of X_t? Is X_t a martingale?

Answer: A solution to this question was given in Chapter 1 using integration by parts; we include a different solution here.

Note that X_t is not a martingale because, if we rewrite (3.152) in differential form as

$$dX_t = W_t dt = W_t dt + 0 \, dW_t,$$

we can think of X_t as being a diffusion process with only the drift part W_t.

Recall that the integral of a one-parameter family of Gaussian random variables remains Gaussian. Since W_t is a Gaussian family, X_t is normally distributed. Furthermore,

$$E[X_t] = \int_0^t E[W_\tau] d\tau = 0.$$

Therefore,

$$\text{var}(X_t) \;=\; E[X_t^2] - (E[X_t])^2 \;=\; E[X_t^2]. \quad (3.153)$$

Note that

$$X_t^2 \;=\; \left(\int_0^t W_s ds \right)^2 = \int_0^t \int_0^t W_s W_u ds du, \quad (3.154)$$

and recall that

$$E[W_s W_u] \;=\; \min\{s, u\}, \quad \forall \, s, u > 0; \quad (3.155)$$

see (3.125).

From (3.153–3.155), we obtain that

$$
\begin{aligned}
\text{var}(X_t) \;=\; E[X_t^2] \;&=\; \int_0^t \int_0^t E[W_s W_u] ds du \\
&=\; \int_0^t \int_0^t \min\{s, u\} ds du \\
&=\; \int_0^t \left(\int_0^u s ds + \int_u^t u ds \right) du \\
&=\; \int_0^t \left(\frac{u^2}{2} + u(t-u) \right) du \\
&=\; \int_0^t \left(ut - \frac{u^2}{2} \right) du \\
&=\; \frac{t^3}{2} - \frac{t^3}{6} \\
&=\; \frac{t^3}{3}.
\end{aligned}
$$

We conclude that X_t is normally distributed with mean 0 and variance $\frac{t^3}{3}$, i.e.,

$$X_t \;\sim\; N\left(0, \frac{t^3}{3} \right). \quad \square$$

Question 21. What is an Itô process?

Answer: An Itô process is a generic term referring to a stochastic process X_t determined by the solution of a stochastic differential equation (SDE) of the form

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t, \qquad (3.156)$$

where W_t is a Wiener process. The coefficient $a(x, t)$ of the dt term is the drift of X_t; the coefficient $b(x, t)$ of the dW_t term is the diffusion of X_t. The SDE (3.156) is, by definition, the shorthand notation for the stochastic integral equation

$$X_t = X_0 + \int_0^t a(X_s, s) \, ds + \int_0^t b(X_s, s) \, dW_s.$$

We note that a sufficient condition for the existence and uniqueness of the (strong) solution to an SDE is for the drift and the diffusion coefficients $a(x, t)$ and $b(x, t)$ to be locally Lipschitz functions of at most linear growth in x. □

Question 22. What is Itô's lemma?

Answer: Itô's lemma, also known as Itô's formula, states that, if X_t is an Itô process satisfying the SDE

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t,$$

then for any function $f(x, t)$ with continuous second order partial derivative in x and continuous first order partial derivative in t, the process $f(X_t, t)$ is also an Itô process, driven by the same Wiener process W_t, and the drift and diffusion parts of $f(X_t, t)$ are determined according to the Taylor expansion of $f(x, t)$ to first order for the t part and

up to second order for the x part:

$$
\begin{aligned}
df(X_t, t) &= \frac{\partial f}{\partial t}\, dt \;+\; \frac{\partial f}{\partial x}\, dX_t \;+\; \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\,(dX_t)^2 \\
&= \frac{\partial f}{\partial t}\, dt \;+\; \frac{\partial f}{\partial x}\,[a(X_t,t)dt \;+\; b(X_t,t)dW_t] \\
&\quad + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\,[a(X_t,t)dt + b(X_t,t)dW_t]^2 \\
&= \left[\frac{\partial f}{\partial t} + a(X_t,t)\frac{\partial f}{\partial x} + \frac{b^2(X_t,t)}{2}\frac{\partial^2 f}{\partial x^2}\right]\, dt \\
&\quad + b(X_t,t)\frac{\partial f}{\partial x}\, dW_t. \tag{3.157}
\end{aligned}
$$

Note that for (3.157) we used the fact that

$$
\begin{aligned}
(dX_t)^2 &= [a(X_t,t)dt + b(X_t,t)dW_t]^2 \\
&= b^2(X_t,t)dt
\end{aligned}
$$

since $(dW_t)^2 = dt$ (see Problem 3.6 in this section), $(dt)^2 = 0$, and $dW_t\, dt = 0$. $\square$

Question 23. If W_t is a Wiener process, is the process $X_t = W_t^2$ a martingale?

Answer: Recall that a stochastic process M_t defined in a filtered probability space $(\Omega, \mathcal{F}_t, P)$ is a martingale if and only if

(i) M_t is adapted, i.e., M_t is $\mathcal{F}_t$-measurable for all t;
(ii) M_t is integrable for all t, i.e., $E[M_t] < \infty$ for all t;
(iii) $E[|M_t|\,|\mathcal{F}_s] = M_s$ almost surely for all $s < t$.

We check whether the process X_t satisfies the conditions (i), (ii), and (iii) above.

Since any continuous function of the Wiener process W_t is adapted, the process X_t is adapted, and therefore condition (i) is satisfied.

Also, $E[X_t] = E[W_t^2] = t < \infty$, since W_t is a normal random variable of mean 0 and variance t, and therefore

$E[W_t^2] = \text{var}[W_t] = t$. Thus, the random variable X_t is integrable for every t, and X_t satisfies condition (ii).

To check whether X_t satisfies condition (iii), note that, for every $s < t$,

$$
\begin{aligned}
E\left[X_t|\mathcal{F}_s\right] &= E\left[W_t^2|\mathcal{F}_s\right] \\
&= E\left[(W_t - W_s + W_s)^2|\mathcal{F}_s\right] \\
&= E\left[(W_t - W_s)^2|\mathcal{F}_s\right] + 2E\left[(W_t - W_s)W_s|\mathcal{F}_s\right] \\
&\quad + E\left[W_s^2|\mathcal{F}_s\right].
\end{aligned} \tag{3.158}
$$

Since the Wiener process W_t has independent increments, i.e, $W_t - W_s$ is independent of $\mathcal{F}_s$, and stationary increment, i.e., $W_t - W_s \sim W_{t-s}$. Moreover, $W_t - W_s$ is a normal random variable of mean 0 and variance $t - s$, i.e., $E[W_{t-s}] = 0$ and $\text{var}(W_{t-s}) = E[W_{t-s}^2] = t - s$. Then,

$$
E\left[(W_t - W_s)^2|\mathcal{F}_s\right] = E[W_{t-s}^2] = t - s, \tag{3.159}
$$

and

$$
\begin{aligned}
E\left[(W_t - W_s)W_s|\mathcal{F}_s\right] &= W_s E\left[W_t - W_s|\mathcal{F}_s\right] \\
&= W_s E[W_{t-s}] \\
&= 0.
\end{aligned} \tag{3.160}
$$

Since W_s is $\mathcal{F}_s$-measurable, it follows that

$$
E[W_s^2|\mathcal{F}_s] = W_s^2. \tag{3.161}
$$

From (3.158–3.161), we find that

$$
\begin{aligned}
&E\left[X_t|\mathcal{F}_s\right] \\
&= E\left[(W_t - W_s)^2|\mathcal{F}_s\right] + 2E\left[(W_t - W_s)W_s|\mathcal{F}_s\right] \\
&\quad + E\left[W_s^2|\mathcal{F}_s\right] \\
&= t - s + W_s^2.
\end{aligned}
$$

Thus,

$$
E\left[X_t|\mathcal{F}_s\right] = t - s + W_s^2 \neq W_s^2 = X_s,
$$

and therefore the process X_t does not satisfy condition (iii).

We conclude that X_t is not a martingale. $\square$

Question 24. If W_t is a Wiener process, is the process

$$N_t = W_t^3 - 3tW_t$$

a martingale?

Answer:

Solution 1: A stochastic process M_t defined in a filtered probability space $(\Omega, \mathcal{F}_t, P)$ is called a martingale if and only if

(i) M_t is adapted, i.e., M_t is $\mathcal{F}_t$-measurable for all t;

(ii) M_t is integrable for all t, i.e., $E[|M_t|] < \infty$ for all t;

(iii) $E[M_t|\mathcal{F}_s] = M_s$ almost surely for all $s < t$.

We check whether the process N_t satisfies the conditions (i), (ii), and (iii) above.

Since any continuous function of the Wiener process W_t is adapted, the process N_t is adapted, and therefore condition (i) is satisfied.

Also,

$$E[N_t] = E[W_t^3] - 3tE[W_t] = 0 < \infty,$$

since $W_t \sim N(0,t)$, therefore $E[W_t] = E[W_t^3] = 0$. Thus, the random variable N_t is integrable for every t, and N_t satisfies condition (ii).

To check whether N_t satisfies condition (iii), note that, for every $s < t$,

$$
\begin{aligned}
&E\left[N_t|\mathcal{F}_s\right] \\
=\ &E\left[W_t^3 - 3tW_t|\mathcal{F}_s\right] \\
=\ &E\left[(W_t - W_s + W_s)^3|\mathcal{F}_s\right] - 3tE\left[W_t|\mathcal{F}_s\right] \\
=\ &E\left[(W_t - W_s)^3|\mathcal{F}_s\right] + 3E\left[(W_t - W_s)^2 W_s|\mathcal{F}_s\right] \\
&+\ 3E\left[(W_t - W_s)W_s^2|\mathcal{F}_s\right] + E\left[W_s^3|\mathcal{F}_s\right] \\
&-\ 3tE\left[W_t|\mathcal{F}_s\right].
\end{aligned}
$$

Recall that the Wiener process W_t has independent increments, i.e, $W_t - W_s$ is independent of $\mathcal{F}_s$, and stationary increment, i.e., $W_t - W_s \sim W_{t-s}$, and the fact that W_s is $\mathcal{F}_s$-measurable. Then,

$$
\begin{aligned}
E\left[(W_t - W_s)^3 | \mathcal{F}_s\right] &= E[W_{t-s}^3] \\
&= 0; \qquad\qquad (3.162) \\
E\left[(W_t - W_s)^2 W_s | \mathcal{F}_s\right] &= W_s\, E[W_{t-s}^2] \\
&= (t-s)W_s; \quad (3.163) \\
E\left[(W_t - W_s) W_s^2 | \mathcal{F}_s\right] &= W_s^2\, E[W_{t-s}] \\
&= 0; \qquad\qquad (3.164) \\
E\left[W_s^3 | \mathcal{F}_s\right] &= W_s^3, \qquad\quad (3.165)
\end{aligned}
$$

since $W_{t-s} \sim N(0, t-s)$, and therefore

$$E[W_{t-s}] = E[W_{t-s}^3] = 0;$$

$$E[W_{t-s}^2] = \mathrm{var}[W_{t-s}] = t - s.$$

Moreover, since W_t is a martingale, it follows that

$$E[W_t | \mathcal{F}_s] = W_s. \qquad (3.166)$$

From (3.162–3.166), we find that

$$
\begin{aligned}
&E\left[N_t | \mathcal{F}_s\right] \\
={}& E\left[(W_t - W_s)^3 | \mathcal{F}_s\right] + 3E\left[(W_t - W_s)^2 W_s | \mathcal{F}_s\right] \\
&+ 3E\left[(W_t - W_s) W_s^2 | \mathcal{F}_s\right] + E\left[W_s^3 | \mathcal{F}_s\right] \\
&- 3t E\left[W_t | \mathcal{F}_s\right] \\
={}& 0 + 3(t-s)W_s + 0 + W_s^3 - 3t W_s \\
={}& W_s^3 - 3s W_s \\
={}& N_s.
\end{aligned}
$$

Thus, the process N_t satisfies condition (iii), and we conclude that N_t is a martingale.

Solution 2: By applying Itô's formula to N_t, we obtain that

$$
\begin{aligned}
dN_t &= d(W_t^3 - 3tW_t) \\
&= 3W_t^2 dW_t + \frac{1}{2} \cdot 6W_t dt - 3(W_t dt + t dW_t) \\
&= 3(W_t^2 - t) dW_t.
\end{aligned}
$$

The process N_t has zero drift and can be written as a stochastic integral as

$$
N_t = \int_0^t 3(W_s^2 - s) dW_s.
$$

Moreover,

$$
\begin{aligned}
&\int_0^t E\left[\left(3(W_s^2 - s)\right)^2\right] ds \\
&= 9\left[\int_0^t \left(E[W_s^4] - 2sE[W_s^2] + s^2\right) ds\right] \\
&= 9\left[\int_0^t \left(3s^2 - 2s^2 + s^2\right) ds\right] \qquad (3.167) \\
&= 6t^3 < \infty;
\end{aligned}
$$

for (3.167), we used the fact that $W_s = \sqrt{s}Z$, where Z is a standard normal random variable, and therefore

$$
\begin{aligned}
E[W_s^2] &= E\left[\left(\sqrt{s}Z\right)^2\right] = sE[Z^2] = s; \\
E[W_s^4] &= E\left[\left(\sqrt{s}Z\right)^4\right] = s^2 E[Z^4] = 3s^2.
\end{aligned}
$$

Therefore, the process $3(W_s^2 - s)$ is square integrable for $s \in [0, t]$, and, from Theorem 3.1, we conclude that N_t is a martingale.

Question 25. What is Girsanov's theorem?

Answer: Girsanov's theorem is providing a way to change the drift of a Wiener process by defining a new probability measure via a Radon-Nikodym derivative. More precisely, let $(\Omega, \mathcal{F}_t, P)$, for $0 \leq t \leq T$, be a filtered probability space, and let W_t be a Wiener process in the probability measure P. Let h_t be a progressively measurable stochastic process such that the stochastic exponential

$$\mathcal{E}_t(h) = \exp\left(\int_0^t h_s dW_s - \frac{1}{2}\int_0^t h_s^2 ds\right)$$

is a martingale in the probability measure P. Define a new probability measure $\widetilde{P}$ over Ω given by the Radon-Nikodym derivative as follows:

$$\frac{d\widetilde{P}}{dP} = \exp\left(\int_0^T h_t dW_t - \frac{1}{2}\int_0^T h_t^2 dt\right). \quad (3.168)$$

Then, W_t is a Wiener process with drift h under the new probability measure $\widetilde{P}$. Equivalently, if we define $\widetilde{W}_t$ by $\widetilde{W}_t = W_t - h_t$, then $\widetilde{W}_t$ is a Wiener process in the $\widetilde{P}$-measure. More generally, if X_t is the diffusion process satisfying the SDE

$$dX_t = a(X_t, t)dW_t + b(X_t, t)dt$$

under the probability measure P, then under the probability measure $\widetilde{P}$ defined in (3.168), X_t satisfies the following SDE

$$\begin{aligned}
dX_t &= a(X_t, t)dW_t + b(X_t, t)dt \\
&= a(X_t, t)\left(d\widetilde{W}_t + h_t dt\right) + b(X_t, t)dt \\
&= a(X_t, t)d\widetilde{W}_t + \left(b(X_t, t) + h_t a(X_t, t)\right)dt.
\end{aligned}$$

In other words, in the P-measure, X_t has drift part b and diffusion part a; whereas in the $\widetilde{P}$-measure, the diffusion part stays the same while the drift part becomes $b + h\,a$.

In particular, if we choose h to be $-\frac{b}{a}$, then X_t becomes driftless in the $\widetilde{P}$-measure provided that the stochastic exponential $\mathcal{E}_t(h)$ is a martingale.

We note that the martingality[12] of the stochastic exponential $\mathcal{E}_t(h)$ guarantees that the new measure $\widetilde{P}$ defined in (3.168) is a probability measure, i.e., $\int_\Omega d\widetilde{P} = 1$, since

$$
\begin{aligned}
\int_\Omega d\widetilde{P} &= \int_\Omega \frac{d\widetilde{P}}{dP} dP \\
&= E\left[\frac{d\widetilde{P}}{dP}\right] \\
&= E\left[\mathcal{E}_T(h)\right] \\
&= \mathcal{E}_0(h) \\
&= 1. \quad \square
\end{aligned}
$$

Question 26. What is the martingale representation theorem, and how is it related to option pricing and hedging?

Answer: Let W_t be a Wiener process defined on the filtered probability space $(\Omega, \mathcal{F}_t, P)$, where the filtration $\{\mathcal{F}_t\}$ is generated by the Wiener process W_t. Let M_t be a martingale with respect to $\{\mathcal{F}_t\}$ such that M_t is square integrable for every t, i.e., $E[M_t^2] < \infty$, for all t. The martingale representation theorem asserts that, for any such a martingale M_t, there exists an $\mathcal{F}_t$-adapted square integrable process θ_t such that M_t has the stochastic integral representation

$$
M_t = E[M_0] + \int_0^t \theta_s dW_s
$$

almost surely. Note that such martingales have to be continuous.

---

[12]A sufficient condition which ensures the martingality of $\mathcal{E}_t(h)$ is the Novikov's condition $E\left[e^{\frac{1}{2}\int_0^T h_t^2 dt}\right] < \infty$.

The relationship between the martingale representation theorem and option pricing and hedging is as follows. For simplicity, assume that the risk free rate is zero. Assume that the price process S_t of the underlying asset follows the diffusion process determined by the SDE

$$dS_t = \sigma_t S_t dW_t$$

under risk neutral probability P, where the driving Wiener process W_t is defined on the filtered probability space $(\Omega, \mathcal{F}_t, P)$ with the filtration $\{\mathcal{F}_t\}$ is generated by W_t. Consider a derivative whose payoff function (possibly path dependent) is φ_T at maturity T. The Fundamental Theorem of Asset Pricing asserts that, because the risk free rate is assumed zero, the value process V_t of the derivative security is a martingale under risk neutral probability. In fact,

$$V_t = E[\varphi_T | \mathcal{F}_t].$$

Moreover, if the payoff function φ_T is square integrable, i.e., $E[\varphi_T^2] < \infty$, then V_t is a square integrable martingale. Therefore, by applying the martingale representation theorem, we find that there exists a $\mathcal{F}_t$-adapted process θ_t such that

$$
\begin{aligned}
\varphi_T &= V_T = E[V_0] + \int_0^T \theta_t dW_s \\
&= E[V_0] + \int_0^T \frac{\theta_t}{\sigma_t S_t} dS_t.
\end{aligned}
$$

Note that the quantity $\frac{\theta_t}{\sigma_t S_t}$ indicates the amount of shares to hold in order to dynamically hedge the position of the derivative with payoff function φ_T at maturity T. In particular, if φ_T is the payoff of a call option, then $\frac{\theta_t}{\sigma_t S_t}$ corresponds to the delta of the call. $\quad\square$

Question 27. Solve

$$dY_t = Y_t dW_t, \qquad (3.169)$$

where W_t is a Wiener process.

Answer:

Solution 1: Note that (3.169) is a particular case for $\mu = 0$ and $\sigma = 1$ of the stochastic differential equation

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \qquad (3.170)$$

which is the model for the evolution of an asset in the Black–Scholes framework. The solution of (3.170) has the distribution

$$S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma\sqrt{t}Z\right), \qquad (3.171)$$

with $t > 0$, where Z is the standard normal variable. For $\mu = 0$ and $\sigma = 1$ in (3.171), we obtain that the solution to (3.169) is

$$Y_t = Y_0 \exp\left(-\frac{t}{2} + \sqrt{t}Z\right). \qquad (3.172)$$

Solution 2: From Itô's lemma it follows that, if Y_t is a stochastic process satisfying $dY_t = Y_t dW_t$ and $f(y)$ is a function with continuous second order derivative, then

$$d(f(Y_t)) = \frac{1}{2}f''(Y_t)Y_t^2 dt + f'(Y_t)Y_t dW_t. \qquad (3.173)$$

For $f(y) = \ln(y)$, we obtain from (3.173) that

$$d(\ln(Y_t)) = -\frac{1}{2}dt + dW_t. \qquad (3.174)$$

By integrating (3.174) between 0 and t, we find that

$$\ln(Y_t) - \ln(Y_0) = -\frac{t}{2} + W_t - W_0 = -\frac{t}{2} + W_t,$$

since $W_0 = 0$, and therefore

$$Y_t = Y_0 \exp\left(-\frac{t}{2} + W_t\right).$$

Since W_t is a Wiener process, W_t is a normal random variable of mean 0 and variance t, i.e., $W_t = \sqrt{t}Z$, where Z is the standard normal variable. Thus, Y_t has the distribution

$$Y_t = Y_0 \exp\left(-\frac{t}{2} + \sqrt{t}Z\right),$$

which is the same as (3.172). $\square$

Question 28. Solve the following SDEs:

(i) $dY_t = \mu Y_t dt + \sigma Y_t dW_t$;

(ii) $dX_t = \mu dt + (aX_t + b)dW_t$.

Answer:
(i) Recall from Itô's formula that, if Y_t satisfies the SDE

$$dY_t = \mu Y_t dt + \sigma Y_t dW_t,$$

and if $f(y)$ is a function with continuous second order derivative, then

$$
\begin{aligned}
df(Y_t) &= f'(Y_t)dY_t + \frac{1}{2}f''(Y_t)\sigma^2 Y_t^2 dt \\
&= \left[\mu Y_t f'(Y_t) + \frac{\sigma^2}{2}Y_t^2 f''(Y_t)\right]dt \\
&\quad + \sigma Y_t f'(Y_t)dW_t.
\end{aligned}
\tag{3.175}
$$

For $f(y) = \ln(y)$, we obtain from (3.175) that

$$d\ln(Y_t) = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma dW_t. \tag{3.176}$$

By integrating (3.176) from 0 to t, we obtain that

$$\ln(Y_t) - \ln(Y_0) = \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t, \tag{3.177}$$

and by solving (3.177) for Y_t, we conclude that

$$Y_t = Y_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}.$$

(ii) We look for a solution of

$$dX_t = \mu dt + (aX_t + b)dW_t \tag{3.178}$$

of the form

$$X_t = U_t V_t, \tag{3.179}$$

where the process U_t is defined by the solution to the SDE

$$dU_t = aU_t dW_t, \quad \text{with } U_0 = 1, \tag{3.180}$$

and the process V_t is the solution to the SDE

$$dV_t = \alpha_t dt + \beta_t dW_t, \quad \text{with } V_0 = X_0, \tag{3.181}$$

where the coefficients α_t and β_t are to be determined.

Recall from (i) that the solution to

$$dY_t = \mu Y_t dt + \sigma Y_t dW_t$$

is

$$Y_t = Y_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}. \tag{3.182}$$

By letting $\mu = 0$ and $\sigma = a$ in (3.182), we find that the solution to (3.180) is

$$U_t = e^{aW_t - \frac{a^2}{2}t}. \tag{3.183}$$

Note that

$$\begin{aligned}
dU_t \, dV_t &= (aU_t dW_t)(\alpha_t dt + \beta_t dW_t) \\
&= a\beta_t U_t dt, \tag{3.184}
\end{aligned}$$

since $(dW_t)^2 = dt$ and $dW_t dt = 0$ (because this term has order $(dt)^{3/2}$). By applying Itô's product rule to $X_t = U_t V_t$ and using (3.184), we obtain that

$$
\begin{aligned}
dX_t &= d(U_t V_t) = U_t dV_t + V_t dU_t + dU_t \, dV_t \\
&= U_t(\alpha_t dt + \beta_t dW_t) + V_t(aU_t dW_t) \\
&\quad + a\beta_t U_t dt \\
&= (a\beta_t + \alpha_t)U_t dt + (aU_t V_t + \beta_t U_t)dW_t \\
&= (a\beta_t + \alpha_t)U_t dt + (aX_t + \beta_t U_t)dW_t \quad (3.185)
\end{aligned}
$$

Then, $X_t = U_t V_t$ is a solution to (3.178) if and only if the coefficients of the dW_t terms and of the dt terms in (3.178) and (3.185) are equal.

From the dW_t terms, we obtain that

$$
\begin{aligned}
& aX_t + b = aX_t + \beta_t U_t \\
\iff & \beta_t U_t = b \\
\iff & \beta_t = bU_t^{-1} = b\, e^{-aW_t + \frac{a^2}{2}t}. \quad (3.186)
\end{aligned}
$$

From the dt terms, we obtain that

$$
\begin{aligned}
& (a\beta_t + \alpha_t)U_t = \mu \\
\iff & a\beta_t + \alpha_t = \mu U_t^{-1} \\
\iff & \alpha_t = \mu U_t^{-1} - a\beta_t \\
\iff & \alpha_t = \mu\, e^{-aW_t + \frac{a^2}{2}t} - abe^{-aW_t + \frac{a^2}{2}t} \\
\iff & \alpha_t = (\mu - ab)\, e^{-aW_t + \frac{a^2}{2}t}. \quad (3.187)
\end{aligned}
$$

Note that the solution V_t to (3.181) is given by

$$
V_t = X_0 + \int_0^t \alpha_s ds + \int_0^t \beta_s dW_s. \quad (3.188)
$$

From (3.179), (3.183), and (3.188), we find that the solution X_t to (3.178) is

$$
X_t = e^{aW_t - \frac{a^2}{2}t} \left(X_0 + \int_0^t \alpha_s ds + \int_0^t \beta_s dW_s \right),
$$

and, using (3.186) and (3.187), we obtain that

$$
\begin{aligned}
X_t &= e^{aW_t - \frac{a^2}{2}t}\left[X_0 + \int_0^t (\mu - ab)\, e^{-aW_s + \frac{a^2}{2}s}\, ds \right. \\
&\qquad \left. + \int_0^t b\, e^{-W_s + \frac{a^2}{2}s}\, dW_s\right] \\
&= X_0\, e^{aW_t - \frac{a^2}{2}t} \\
&\qquad + (\mu - ab)\int_0^t e^{a(W_t - W_s) - \frac{a^2}{2}(t-s)}\, ds \\
&\qquad + b\int_0^t e^{a(W_t - W_s) - \frac{a^2}{2}(t-s)}\, dW_s. \quad \square
\end{aligned}
$$

Question 29. What is the Heston model?

Answer: In Heston's stochastic volatility model, it is assumed that the price of the underlying asset satisfies the same SDE as in the lognormal model, i.e.,

$$
dS_t = \mu S_t dt + \sqrt{v_t}\, S_t dW_t,
$$

whereas the instantaneous variance v_t itself follows a mean reverting CIR (Cox-Ingersoll-Ross) process

$$
dv_t = \lambda(v_t - m)dt + \eta\sqrt{v_t}\, dZ_t,
$$

where $\lambda > 0$ and $m > 0$ are positive constants. The driving Wiener processes W_t and Z_t are correlated with constant correlation ρ, i.e.,

$$
\mathrm{corr}(dW_t, dZ_t) = \rho dt.
$$

The Heston model is a benchmark, and is commonly used in derivative pricing because it has following features:

- it takes into account the leverage effect, namely, the driving Wiener processes W_t and Z_t are correlated; empirically, the correlation ρ is negative and that is why the word "leverage";

- it has a quasi closed form (up to an inverse Fourier transform) solution for the prices of European options, which make the calibration more tractable;

- the variance process is mean reverting with rate of reversion λ and long term mean m.

Note that, if the parameters of the volatility process are in the regime $2\lambda m < \eta^2$, then zero is an attainable boundary for the volatility process. Practitioners usually assume that the boundary behavior at zero is either absorption, i.e., the process is stuck at zero once it hits zero, or reflection, i.e, the process bounces back right after it hits zero. On the other hand, the other boundary infinity is an unattainable boundary. □

3.7 Brainteasers.

Question 1. A flea is going between two points which are 100 inches apart by jumping (always in the same direction) either one inch or two inches at a time. How many different paths can the flea travel by?

Answer: Let a_n denote the number of different paths of the flea that covers the distance of n inches, jumping either one inch or two inches at a time. We want to find a_{100}.

Since the flea can jump either one inch or two inches at a time, it could have made the last jump either from the end of the $(n-1)$st inch or from the end of the $(n-2)$nd inch. Hence, the total number of ways the flea can cover the distance of n inches, jumping either one inch or two inches at a time, is the sum of the number of ways the flea can cover the distance of $n-1$ inches, jumping either one inch or two inches at a time, and the number of ways the flea can cover the distance of $n-2$ inches, jumping either one inch or two inches at a time.

In other words, for $n > 2$, we have

$$a_n = a_{n-1} + a_{n-2}. \tag{3.189}$$

Note that $a_1 = 1$, since the flea can cover 1 inch in only one way, jumping one inch once; while $a_2 = 2$, since the flea can cover 2 inches in two ways, either jumping one inch twice, or jumping two inches once.

Note that (3.189) is the recurrence relation for the Fibonacci sequence. Let

$$\phi_1 = \frac{1 + \sqrt{5}}{2} \quad \text{and} \quad \phi_2 = \frac{1 - \sqrt{5}}{2}$$

be the roots of the characteristic equation $x^2 - x - 1 = 0$ corresponding to (3.189).

Then,

$$a_n = C_1 \phi_1^n + C_2 \phi_2^n, \quad \forall\, n \geq 1,$$

where the constants C_1 and C_2 are such that $a_1 = 1$ and $a_2 = 2$.

By solving the linear system

$$\begin{cases} C_1\phi_1 + C_2\phi_2 & = & 1 \\ C_1\phi_1^2 + C_2\phi_2^2 & = & 2, \end{cases}$$

we obtain that

$$C_1 = \frac{1+\sqrt{5}}{2\sqrt{5}} = \frac{\phi_1}{\sqrt{5}};$$

$$C_2 = -\frac{\sqrt{5}-1}{2\sqrt{5}} = -\frac{\phi_2}{\sqrt{5}},$$

and therefore

$$a_n = \frac{1}{\sqrt{5}}\left(\phi_1^{n+1} - \phi_2^{n+1}\right).$$

Plugging $n = 100$ into the last expression gives the answer. $\square$

Question 2. I have a bag containing three pancakes: one golden on both sides, one burnt on both sides, and one golden on one side and burnt on the other. You shake the bag, draw a pancake at random, look at one side, and notice that it is golden. What is the probability that the other side is golden?

Answer:

Solution 1: Label the pancakes 0, 1, and 2, according to the number of burnt sides. Let E_i, $i = 0, 1, 2$, denote the event that the pancake with i burnt sides is drawn from the bag. Let A denote the event that the side (of the randomly drawn pancake) we look at is golden.

The probability of the pancake drawn from the bag having no burnt sides given that one side is golden is $P(E_0|A)$. Then, from the Bayes' formula, we obtain that

$$P(E_0|A) = \frac{P(E_0 \cap A)}{P(A)} =$$

$$\frac{P(E_0)P(A|E_0)}{P(A|E_0)P(E_0) + P(A|E_1)P(E_1) + P(A|E_2)P(E_2)}$$

$$= \frac{\frac{1}{3} \cdot 1}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}}$$

$$= \frac{2}{3}.$$

Solution 2: Out of the six possible sides that we could have seen, three are golden. Out of these, two belong to a pancake that is golden on both sides. Therefore, the probability of the other side being golden is $\frac{2}{3}$. □

Question 3. Alice and Bob are playing heads and tails, Alice tosses $n + 1$ coins, Bob tosses n coins. The coins are fair. What is the probability that Alice will have strictly more heads than Bob?

Answer:

Solution 1: Alice flips the coin more often than Bob, so either she must end up with more heads or with more tails than Bob. She cannot, however, end up with more heads and more tails, because she only flips one more coin than Bob. We deduce that either Alice gets more heads or she gets more tails.

Since these events are equally likely, they both have probability $\frac{1}{2}$, and therefore the probability that Alice will have strictly more heads than Bob is $\frac{1}{2}$.

Solution 2: Suppose that Alice and Bob begin by flipping n coins each. Let p be the probability that Alice gets more heads than Bob, and let q be the probability that both Alice and Bob get an equal number of heads. Note that $2p + q = 1$ since the probability that Alice gets more

heads than Bob is, by symmetry, equal to the probability that Bob gets more heads than Alice.

Alice then flips the $(n+1)$st coin. For Alice to have more heads, she either had more heads than Bob before flipping the last coin; or the same number of heads as Bob before flipping the last coin and her $(n+1)$st flip must come up heads. Hence, the probability of Alice winning is $p + \frac{1}{2}q$.

Since $2p + q = 1$, then $p + \frac{1}{2}q = \frac{1}{2}$. $\square$

Question 4. Alice is in a restaurant trying to decide between three desserts. How can she choose one of three desserts with equal probability with the help of a fair coin? What if the coin is biased and the bias is unknown?

Answer:

Solution 1: Denote the desserts by A, B, and C. First, suppose the coin is fair. Denote heads by H and tails by T. The procedure to choose one of three desserts with equal probability is as follows: toss the coin twice; let the outcomes TH, HT, and TT correspond to choosing desserts A, B, and C, respectively; if the outcome is HH, repeat the procedure.

Note that the probability of our procedure not being repeated is $p = \frac{3}{4}$; hence, the number of times our procedure is repeated is a geometric random variable with p as its parameter. The expected number of times our procedure is repeated is $\frac{1}{p} = \frac{4}{3}$. Since each procedure involves two coin tosses, the expected number of coin tosses before Alice chooses one of three desserts with equal probability is $\frac{8}{3}$.

Now, suppose the coin is biased. One procedure to choose one of three desserts with equal probability would be as follows: toss the coin four times; denote by $THHT$, $HTTH$, and $THTH$ the outcomes corresponding to choosing desserts A, B, and C, respectively; all the other 4-toss outcomes result in repeating the procedure.

Solution 2: An alternative procedure is as follows: toss the coin three times; denote by HTT, THT, and HTT the outcomes corresponding to choosing desserts A, B, and C, respectively; all the other 3-toss outcomes result in repeating the procedure.

Using an argument similar to the case with a fair coin, one finds that the expected number of tosses of a coin with an unknown bias, before Alice chooses one of three desserts with equal probability, is $\frac{64}{3}$ and 8, respectively, for the two procedures described above. □

Question 5. What is the expected number of times you must flip a fair coin until it lands on head? What if the coin is biased and lands on head with probability p?

Answer: Denote by X be the number of times you must flip a fair coin until it lands on head. If the first coin toss is a head (which happens with probability $\frac{1}{2}$), then $X = 1$. If the first coin toss is a tail (which also happens with probability $\frac{1}{2}$), then the coin tossing process resets and the number of steps before the coin lands on head will be 1 plus the expected number of coin tosses until the coin lands heads. In other words, the expected number of coin toss $E[X]$ satisfies the equation

$$E[X] = \frac{1}{2} + \frac{1}{2}\left(1 + E[X]\right). \qquad (3.190)$$

Solving (3.190) for $E[X]$, we conclude that $E[X] = 2$, i.e., the expected number of times you must flip a fair coin until it lands heads is 2.

On the other hand, if the coin is biased with the probability p of landing on head, the same argument still applies. However, in this case, (3.190) reads

$$E[X] = p + (1 - p)\left(1 + E[X]\right).$$

Again, solve for $E[X]$, we obtain $E[X] = \frac{1}{p}$. □

Question 6. What is the expected number of coin tosses of a fair coin in order to get two heads in a row? What if the coin is biased with 25% probability of getting heads?

Answer: We solve the general case of a biased coin with probability p of the coin toss resulting in heads. The outcomes of the first two tosses are as follows:

• If the first toss is tails, which happens with probability $1 - p$, then the process resets and the expected number of tosses increases by 1.

• If the first toss is heads, and if the second toss is also heads, which happens with probability p^2, then two consecutive heads are obtained after two tosses.

• If the first toss is heads, and if the second toss is tails, which happens with probability $p(1 - p)$, then the process resets and the expected number of tosses increases by 2.

If $E[X]$ denotes the expected number of tosses in order to get two heads in a row, we conclude that

$$E[X] = (1 - p)(1 + E[X]) + 2p^2 + p(1 - p)(2 + E[X]).$$

By solving for $E[X]$, we obtain that

$$E[X] = \frac{1 + p}{p^2}. \tag{3.191}$$

For an unbiased coin, i.e., for $p = \frac{1}{2}$, we find from (3.191) that $E[X] = 6$, i.e., the expected number of coin tosses to obtain two heads in a row is 6.

For a biased coin with 25% probability of getting heads, i.e., for $p = \frac{1}{4}$, we find from (3.191) that $E[X] = 20$, i.e., the expected number of coin tosses to obtain two heads in a row in this case is 20. $\square$

Question 7. A fair coin is tossed n times. What is the probability that no two consecutive heads appear?

Answer: The total number of sequences of heads and tails of length n is 2^n. Let a_n be the number of sequences of

heads and tails of length n, such that no two consecutive
heads appear. Then, the probability that no two consec-
utive heads appear is $\frac{a_n}{2^n}$.

Note that $a_1 = 2$ (H and T do not contain two consec-
utive heads) and $a_2 = 3$ (out of HH, HT, TH, and TT,
only HH contains two consecutive heads). We find the
closed formula for a_n by deriving a recurrence relation as
follows:

A sequence of $n \geq 3$ coin tosses does not contain two
consecutive heads if and only if: (i) either it begins with a
tail, followed by a sequence of $n-1$ coin tosses with no two
consecutive heads; (ii) or it begins with a head, followed
by a tail, and followed by a sequence of $n - 2$ coin tosses
with no two consecutive heads. Since these two scenarios
are mutually exclusive, it follows that

$$a_n = a_{n-1} + a_{n-2}, \quad \forall \, n \geq 3. \qquad (3.192)$$

Note that (3.192) is the recurrence relation for the Fi-
bonacci sequence. Let $\phi_1 = \frac{1+\sqrt{5}}{2}$ and $\phi_2 = \frac{1-\sqrt{5}}{2}$ be
the roots of the characteristic equation $x^2 - x - 1 = 0$
corresponding to (3.192). Then,

$$a_n = C_1 \phi_1^n + C_2 \phi_2^n, \quad \forall \, n \geq 1,$$

where the constants C_1 and C_2 are such that $a_1 = 2$ and
$a_2 = 3$.

By solving the linear system

$$\begin{cases} C_1 \phi_1 + C_2 \phi_2 = 2 \\ C_1 \phi_1^2 + C_2 \phi_2^2 = 3, \end{cases}$$

we obtain that

$$C_1 = \frac{3 + \sqrt{5}}{2\sqrt{5}} = \frac{\phi_1^2}{\sqrt{5}};$$

$$C_2 = \frac{3 - \sqrt{5}}{2\sqrt{5}} = -\frac{\phi_2^2}{\sqrt{5}}.$$

We conclude that

$$a_n = \frac{1}{\sqrt{5}} \left(\phi_1^{n+2} - \phi_2^{n+2} \right),$$

and therefore the probability that no two consecutive heads appear in n tosses of a fair coin, which is equal to $\frac{a_n}{2^n}$, is

$$\frac{1}{2^n \sqrt{5}} \left(\phi_1^{n+2} - \phi_2^{n+2} \right). \quad \square$$

Question 8. You have two identical Fabergé eggs, either of which would break if dropped from the top of a building with 100 floors. Your task is to determine the highest floor from which the an egg could be dropped without breaking. What is the minimum number of drops required to achieve this? You are allowed to break both eggs in the process.

Answer: Consider the following more general problem:

Find the largest number of floors $h_e(n)$ a building could have in order to be able to determine the highest floor from which the an egg could be dropped without breaking using e eggs and n drops.

Since one drop can only determine one floor, it follows that

$$h_e(1) = 1. \tag{3.193}$$

If we have only one egg at our disposal, the only possible strategy is to try the floors one by one from bottom to top; hence,

$$h_1(n) = n.$$

When $e \geq 2$ and $n \geq 2$, the first drop cannot be from the floor higher than $h_{e-1}(n-1) + 1$, since if the egg breaks, there are only $e - 1$ eggs and $n - 1$ drops left, and the highest floor we can still handle is $h_{e-1}(n-1)$. If the first drop does not break an egg, we can treat floor

$h_{e-1}(n-1) + 2$ as the new floor 1, and reduce it to a problem with e eggs and $n-1$ drops, and therefore

$$h_e(n) = 1 + h_{e-1}(n-1) + h_e(n-1).$$

Iterating this argument, we obtain that

$$
\begin{aligned}
& h_e(n) \\
= \; & 1 + h_{e-1}(n-1) + h_e(n-1) \\
= \; & 2 + h_{e-1}(n-1) + h_{e-1}(n-2) + h_e(n-2) \\
= \; & \ldots \\
= \; & (n-1) + \sum_{j=1}^{n-1} h_{e-1}(j) + h_e(1) \\
= \; & n + \sum_{j=1}^{n-1} h_{e-1}(j),
\end{aligned}
$$

since $h_e(1) = 1$; see (3.193).

For $e = 2$, the formula above becomes

$$
\begin{aligned}
& h_2(n) \\
= \; & n + \sum_{j=1}^{n-1} h_1(j) = n + \sum_{j=1}^{n-1} j = n + \frac{(n-1)n}{2} \\
= \; & \frac{n(n+1)}{2},
\end{aligned}
\tag{3.194}
$$

where the next-to-last step follows from the summation formula $\sum_{j=1}^{k} = \frac{k(k+1)}{2}$.

Since $h_2(13) = 91 < 100 \le 105 = h_2(14)$, the required number of drops is 14.

Note that our iterative argument also provides an al-

gorithm: you drop the first egg from the floors

$$14 = 1 + h_1(13),$$

$$27 = 2 + h_1(13) + h_1(12),$$

$$39 = 3 + \sum_{j=11}^{13} h_1(j),$$

$$50 = 4 + \sum_{j=10}^{13} h_1(j),$$

$$60 = 5 + \sum_{j=9}^{13} h_1(j),$$

$$69 = 6 + \sum_{j=8}^{13} h_1(j),$$

$$77 = 7 + \sum_{j=7}^{13} h_1(j),$$

$$84 = 8 + \sum_{j=6}^{13} h_1(j),$$

$$90 = 9 + \sum_{j=5}^{13} h_1(j),$$

$$95 = 10 + \sum_{j=4}^{13} h_1(j),$$

$$99 = 11 + \sum_{j=3}^{13} h_1(j),$$

and 100; that is, move up by $14 = 1 + h_1(13)$ floors, then by $13 = 1 + h_1(12)$ floors, then by $12 = 1 + h_1(11)$ floors, and so on, until the first egg breaks (or does not) from the 100th floor. Calling the floor from which the first egg breaks f and the previously tested floor f', you drop the

second egg from the intervening floors $f'+1, f'+2, \ldots, f-$ 1 in that order.   □

Question 9. An ant is in the corner of a $10 \times 10 \times 10$ room and wants to go to the opposite corner. What is the length of the shortest path the ant can take?

Answer: For clarity, assume that the ant is in a corner by the ceiling. Denote that corner by A, and denote by B the opposite corner to A, which is by the floor. The shortest path from A to B would require the ant to go on a straight line across a wall of the room to a side of the floor and from there on a straight line along the floor to the B. If you imagine laying down to the floor the vertical wall the ant went down from A to the floor, you have a 10×20 rectangle with A and B opposite corners in the rectangle. The shortest path for the ant to go from A to B is by following the diagonal of the rectangle, which has length $10\sqrt{5} \approx 22.36$.   □

Question 10. A $10 \times 10 \times 10$ cube is made of $1,000$ unit cubes. How many unit cubes can you see on the outside?

Answer: If all the outside unit cubes are removed, what remains is an $8 \times 8 \times 8$ cube, which is made of $8^3 = 512$ unit cubes. Thus, there are $1000 - 512 = 488$ outside unit cubes.   □

Question 11. Fox Mulder is imprisoned by aliens in a large circular field surrounded by a fence. Outside the fence is a vicious alien that can run four times as fast as Mulder, but is constrained to stay near the fence. If Mulder can contrive to get to an unguarded point on the fence, he can quickly scale the fence and escape. Can he get to a point on the fence ahead of the alien?

Answer: Let R denote the radius of the circular field, whose center we denote by C. Denote Mulder's speed by

v. The alien's speed is then $4v$. Denote Mulder's and alien's positions by M and A, respectively.

Mulder cannot just run for the fence along the straight line connecting C with the point on the fence diametrically opposite to A. Indeed, while it takes Mulder $\frac{R}{v}$ time to cover the distance R, the alien would cover the distance πR in $\frac{\pi R}{4v}$ time and the alien would catch up with Mulder, since $\frac{R}{v} > \frac{\pi R}{4v}$.

To optimize this strategy, Mulder needs to start running for the fence from a point that is closer to the fence than C. Assume that Mulder somehow managed to be at a point M that is xR away from C (where $0 < x < 1$), with M, C, and A collinear, and C between M and A. Denote by P the point on the circle diametrically opposite to A; see Figure 3.2. Then, $MC = xR$ and $MP = (1-x)R$. It takes Mulder $\frac{(1-x)R}{v}$ time to reach the fence running from M to P, while the alien needs $\frac{\pi R}{4v}$ time to reach the point P going from A to P on a semicircle. Note that

$$\frac{(1-x)R}{v} < \frac{\pi R}{4v}, \quad \text{if } x > 1 - \frac{\pi}{4}.$$

Thus, if $x > 1 - \frac{\pi}{4}$, Mulder would be able to escape the alien.

We are now ready to describe Mulder's escape strategy. He sets $x = 1 - \frac{\pi}{4} + 0.01$. Note that $x < \frac{1}{4}$, since $0.01 < \frac{\pi-3}{4}$. Regardless of the alien's movement, Mulder first runs from C to any point on the circle of radius xR, centered at C. Then he runs around that circle, until his position M is such that M, C, and A are collinear, with C between M and A. He is able to do so, since $x < \frac{1}{4}$ and the alien's speed is only 4 times his speed.[13] Finally,

---

[13]Mulder is able to do so since, if $x < \frac{1}{4}$, his angular speed, $\frac{\theta v}{\pi R x}$, is larger than the angular speed $\frac{4\theta v}{\pi R}$ of the alien:

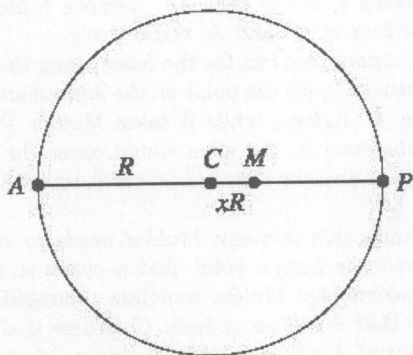$$\frac{\theta v}{\pi R x} > \frac{4\theta v}{\pi R} \iff \frac{1}{4} > x.$$

Figure 3.2: Mulder can reach P from M before the alien can do so from A.

Mulder runs from M to P and will reach P before alien does, as shown above, since $x > 1 - \frac{\pi}{4}$. □

Question 12. At your subway station, you notice that of the two trains running in opposite directions which are supposed to arrive with the same frequency, the train going in one direction comes first 80% of the time, while the train going in the opposite direction comes first only 20% of the time. What do you think could be happening?

Answer: One thing that could be happening is that the train that comes first 80% of the time comes in fact more frequently than the other one. However, even if both trains run with the same frequency, one train might come first 80% of the time. For example, assuming that your arrival in the station is uniformly distributed, if both trains run every ten minutes, and train A comes into the station at 1:00, 1:10, 1:20, ..., while train B comes at 1:12, 1:22, 1:32, ..., then train A will come first 80% of the time.

□

Question 13. You start off with one amoeba. Every minute, this amoeba can either die, do nothing, split into two amoebas, or split into three amoebas; all these scenarios being equally likely to happen. All further amoebas behave the same way. What is the probability that the amoebas eventually die off?

Answer:

Solution 1: (Due to Yu Gan, Baruch MFE'14.) Denote by A_k the event that no amoebas are alive after k minutes. Let $p_k = P(A_k)$. Note that $p_1 = \frac{1}{4}$ and $A_k \subseteq A_{k+1}$, for all $k \geq 1$.

The probability p that the amoebas eventually die off is the probability that at some point in time, i.e., after n minutes for some n, no amoebas are alive. In other words,

$$p = P\left(\bigcup_{k=1}^{\infty} A_k\right).$$

Note that $\bigcup_{k=1}^{n} A_k = A_n$, since $A_k \subseteq A_{k+1}$ for all $k \geq 1$. Then,

$$
\begin{aligned}
p &= P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\lim_{n \to \infty} \bigcup_{k=1}^{n} A_k\right) \\
&= \lim_{n \to \infty} P\left(\bigcup_{k=1}^{n} A_k\right) = \lim_{n \to \infty} P(A_n) \\
&= \lim_{n \to \infty} p_n.
\end{aligned}
\tag{3.195}
$$

Given the four equally probable outcomes after the first minute, i.e, the amoeba can die, remain one amoeba, split into two amoebas, or split into three amoebas, it follows that

$$p_n = \frac{1}{4} + \frac{1}{4}p_{n-1} + \frac{1}{4}p_{n-1}^2 + \frac{1}{4}p_{n-1}^3, \tag{3.196}$$

for all $n \geq 2$.

The sequence $(p_n)_{n \geq 0}$ is increasing. Recall that $A_n \subseteq A_{n+1}$, and therefore

$$p_n = P(A_n) \leq P(A_{n+1}) = p_{n+1}, \ \forall n \geq 1.$$

Also, we can see by induction that the sequence $(p_n)_{n \geq 0}$ is bounded from above by $\sqrt{2} - 1$, since $p_1 = \frac{1}{4} < \sqrt{2} - 1$, and, if we assume that $p_{n-1} < \sqrt{2} - 1$ for some $n \geq 2$, then we obtain from (3.196) that

$$p_n \ < \ \frac{1}{4} + \frac{1}{4}(\sqrt{2} - 1) + \frac{1}{4}(\sqrt{2} - 1)^2 + \frac{1}{4}(\sqrt{2} - 1)^3$$
$$= \ \sqrt{2} - 1.$$

Thus, the sequence $(p_n)_{n \geq 0}$ is bounded from above and increasing, and therefore convergent.

Recall from (3.195) that $\lim_{n \to \infty} p_n = p$, where p is the probability that the amoebas eventually die off. Since $p_n < \sqrt{2} - 1$ for all $n \geq 1$, it follows that $p \leq \sqrt{2} - 1$. Moreover, from (3.196), we find that

$$p = \frac{1}{4} + \frac{1}{4}p + \frac{1}{4}p^2 + \frac{1}{4}p^3,$$

which can be written as

$$0 = p^3 + p^2 - 3p + 1 = (p - 1)(p^2 + 2p - 1),$$

and has solutions 1, $\sqrt{2} - 1$, and $-\sqrt{2} - 1$. Since $0 < p \leq \sqrt{2} - 1$, we obtain that $p = \sqrt{2} - 1$.

In other words, the probability that the amoebas eventually die off is $\sqrt{2} - 1$.

Solution 2: Let p be the probability that the descendants of a single amoeba will die out eventually. Then, the probability that the descendants of n amoebas will all die out eventually is p^n, since each amoeba is independent of all other amoebas.

Furthermore, the probability that the descendants of an amoeba will die out eventually is independent of time when averaged over all the possibilities.

At the beginning, the probability that the descendants of an amoeba will die out eventually is, by definition, p. After one minute, the initial amoeba turns into 0, 1, 2, or 3 amoebas, with probability of $\frac{1}{4}$ for each case. Thus, the probability that the descendants of the amoeba die out is now

$$\frac{1}{4}\left(p^0 + p^1 + p^2 + p^3\right) = \frac{1}{4}\left(1 + p + p^2 + p^3\right).$$

These two probabilities must be equal, and therefore

$$p = \frac{1}{4}\left(1 + p + p^2 + p^3\right), \qquad (3.197)$$

which is the same as

$$p^3 + p^2 - 3p + 1 = (p-1)(p^2 + 2p - 1) = 0. \quad (3.198)$$

The roots of (3.198) are $p = 1$, $p = -\sqrt{2} - 1$, and $p = \sqrt{2} - 1$. The only root in the interval $(0,1)$ is $p = \sqrt{2} - 1$, and this is the probability that the amoebas eventually die off.

A more subtle question is why $p = 1$ is not one of the possible answers to the problem in hand? The right hand side of (3.197) is the generating function $h(p)$ of amoeba's branching process. It is a well-known theorem on branching processes that if the mean number of offspring produced by a single amoeba is bigger than 1, then the smallest positive root of the equation $p = h(p)$ is the probability that amoeba's descendants will die out eventually. In our case, the mean number of offspring produced by a single amoeba is $\frac{1}{4}(0 + 1 + 2 + 3) = \frac{3}{2} > 1$, so the theorem applies. $\square$

Question 14. Given a set X with n elements, choose two subsets A and B at random. What is the probability of A being a subset of B?

Answer: When two subsets A and B of X are chosen at random, each element of X is equally likely to end up in any of the following four sets:

$$A \setminus B, \quad B \setminus A, \quad A \cap B, \quad \text{and} \quad X \setminus (A \cup B).$$

For A to be a subset of B, $A \setminus B$ would have to be empty; in other words, none of the n elements of X would end up in $A \setminus B$. The probability of any element of X not ending up in $A \setminus B$ is $\frac{3}{4}$.

We conclude that the probability of A being a subset of B is

$$\left(\frac{3}{4}\right)^n. \quad \square$$

Question 15. Alice writes two distinct real numbers between 0 and 1 on two sheets of paper. Bob selects one of the sheets randomly to inspect it. He then has to declare whether the number he sees is the bigger or smaller of the two.

Is there any way Bob can expect to be correct more than half the times Alice plays this game with him?

Answer: Denote the numbers Alice writes on two sheets of paper by a_1 and a_2, $0 < a_1 < a_2 < 1$. Denote the number Bob selects by A. Bob's task is to guess (with probability of being correct bigger than $1/2$) whether $A = a_1$ or $A = a_2$.

Bob's strategy is as follows: after seeing A, Bob draws a number B uniformly at random from $(0, 1)$; if B is smaller than A, Bob declares that $A = a_2$; otherwise, he declares that $A = a_1$.

Denote by E the event that Bob is correct using this strategy.

Then,

$$
\begin{aligned}
P(E) &= P(E|A = a_1) \cdot P(A = a_1) \\
&\quad + P(E|A = a_2) \cdot P(A = a_2) \\
&= P(B > a_1) \cdot P(A = a_1) \\
&\quad + P(B < a_2) \cdot P(A = a_2) \\
&= (1 - a_1) \cdot 0.5 + a_2 \cdot 0.5 \\
&= 0.5 + 0.5(a_2 - a_1) \\
&> 0.5.
\end{aligned}
$$

The probability that Bob is correct using this strategy is therefore greater than $\frac{1}{2}$. $\quad\square$

Question 16. How many digits does the number 125^{100} have? You are not allowed to use values of $\log_{10} 2$ or $\log_{10} 5$.

Answer: Note that

$$
125^{100} = \left(\frac{1000}{8} \right)^{100} = \frac{1000^{100}}{2^{300}}. \tag{3.199}
$$

Since $2^{10} = 1024$, we obtain that

$$
125^{100} = \frac{1000^{100}}{1024^{30}} = \frac{1000^{70}}{1.024^{30}}. \tag{3.200}
$$

We first show that

$$
1 < 1.024^{30} < 10. \tag{3.201}
$$

From the binomial expansion, it follows that

$$
(1 + 0.024)^{30} = \sum_{j=0}^{30} \binom{30}{j} 0.024^j. \tag{3.202}
$$

Note that the ratio of every two consecutive terms in (3.202) is less than 0.72, since

$$\frac{\binom{30}{j+1}0.024^{j+1}}{\binom{30}{j}0.024^{j}} = \frac{\binom{30}{j+1}}{\binom{30}{j}} \cdot 0.024$$

$$= \frac{30!}{(j+1)!\,(30-j-1)!} \cdot \frac{j!\,(30-j)!}{30!} \cdot 0.024$$

$$= \frac{30-j}{j+1} \cdot 0.024$$

$$< 30 \cdot 0.024$$

$$= 0.72.$$

Then,

$$\binom{30}{j}0.024^{j} \le 0.72^{j} \quad \forall\, 0 \le j \le 30, \qquad (3.203)$$

and, from (3.202) and (3.203), we obtain that

$$(1+0.024)^{30} < \sum_{j=0}^{30} 0.72^{j} < \sum_{j=0}^{\infty} 0.72^{j}$$

$$= \frac{1}{1-0.72}$$

$$< 10;$$

note that, for the equality on the second line above, we used the geometric series identity

$$\sum_{j=1}^{\infty} z^{j} = \frac{1}{1-z} \quad \text{for} \quad z = 0.72.$$

Thus, the inequality (3.201) is proved.
From (3.200) and (3.201), we obtain that

$$10^{209} < 125^{100} = \frac{10^{210}}{1.024^{30}} < 10^{210}, \qquad (3.204)$$

and conclude that 125^{100} has 210 digits. □

Question 17. For every subset of $\{1, 2, 3, \ldots, 2013\}$, arrange the numbers in the increasing order and take the sum with alternating signs. The resulting integer is called the weight of the subset.[14] Find the sum of the weights of all the subsets of $\{1, 2, 3, \ldots, 2013\}$.

Answer: Let $w(S)$ denote the weight of subset S. Every subset S of $\{1, 2, 3, \ldots, 2013\}$ that does not contain element 1 can be uniquely paired with the subset $\{1\} \cup S$ that contains element 1. Since there are 2^{2013} subsets of $\{1, 2, 3, \ldots, 2013\}$, there are 2^{2012} such pairs. Note that $w(S) + w(\{1\} \cup S) = 1$; that is, the combined weight of each pair is 1. For example,

$$w(\{2, 5, 8\}) + w(\{1, 2, 5, 8\})$$
$$= (2 - 5 + 8) + (1 - 2 + 5 - 8)$$
$$= 1.$$

Hence, the sum of the weights of all the subsets of $\{1, 2, 3, \ldots, 2013\}$ is 2^{2012}. □

Question 18. Alice and Bob alternately choose one number from one of the following nine numbers: 1/16, 1/8, 1/4, 1/2, 1, 2, 4, 8, 16, without replacement. Whoever gets three numbers that multiply to one wins the game. Alice starts first. What should her strategy be? Can she always win?

Answer: First, notice that the numbers Alice and Bob play with are powers of two, namely, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, and 2^4. Next, imagine that Alice and Bob are playing on the 3×3 square, whose entries are as follows: the first row (from left to right) is 2^3, 2^{-4}, 2^1; the

---

[14]For example, the weight of the subset $\{3\}$ is 3. The weight of the subset $\{2, 5, 8\}$ is $2 - 5 + 8 = 5$.

second row is 2^{-2}, 2^0, 2^2; and the third row is 2^{-1}, 2^4, 2^{-3}.

The product of the entries in every row, every column, and every diagonal is 1, and these possibilities cover all the ways to choose three of the given numbers multiplying to 1. Thus, Alice and Bob are essentially playing Tic-Tac-Toe! It is well-known that best play by both players in Tic-Tac-Toe leads to a draw. We conclude that Alice does not have a winning strategy, although she cannot lose either. □

Question 19. Mr. and Mrs. Jones invite four other couples over for a party. At the end of the party, Mr. Jones asks everyone else how many people they shook hands with, and finds that everyone gives a different answer. Of course, no one shook hands with his or her spouse and no one shook the same person's hand twice. How many people did Mrs. Jones shake hands with?

Answer: Since each person shook hands with at most eight others, the nine different answers received by Mr. Jones are exactly the numbers 0 through 8. Denote by P_i the person with i handshakes, $i = 0, 1, \ldots, 8$. Mr. Jones is not assigned any additional notation.

P_8 shook hands with 8 people of the total of 9 other people. Thus, P_8 did not shake the hand of only one other person, so that person must be his or her spouse. On the other hand, P_8 did not shake the hand of P_0 since nobody did that. Therefore, P_8 and P_0 are married, and P_8 shook everyone's hand except for P_0. P_7 did not shake the hands of two people, one of whom was his/her spouse. One of these two people had to be P_0 as he or she did not shake anyone's hand, and the other one had to be P_1 as he or she had only one handshake, namely with P_8. Since spouses do not shake hands, the spouse of P_7 is either P_1 or P_0. However, P_0 is married to P_8, so P_1 must be married to P_7.

Proceeding similarly, we find that P_6 and P_2 must be married, and that P_5 and P_3 must be married. Then, P_4 must be Mrs. Jones, since this is the only person whose spouse was not identified, and Mr. Jones was not any one of $P_0, \ldots, P_8$.

We conclude that Mrs. Jones shook hands with four people. □

Question 20. The New York Yankees and the San Francisco Giants are playing in the World Series (best of seven format). You would like to bet $100 on the Yankees winning the World Series, but you can only place bets on individual games, and every time at even odds. How much should you bet on the first game?

Answer: Let P be a 5×5 matrix containing the net payoffs for all the states in this dynamic programming problem. More precisely, $P(i, j)$ denotes the net payoff in the state (i, j) when Yankees have won i and lost j games ($0 \le i, j \le 4$). Clearly, $P(4, j) = 100$ ($0 \le j \le 3$), and $P(i, 4) = -100$ ($0 \le i \le 3$). Moreover, $P(4, 4)$ is left blank, as $(4, 4)$ is not in our state space (4 wins and 4 losses cannot be achieved in a best of seven series).

Let B be a 4×4 matrix containing the bets we need to place at each state (i, j), given that we would like to bet $100 on the Yankees winning the World Series, and given that Yankees have won i and lost j games so far ($0 \le i, j \le 3$). Clearly, $B(3, 3) = 100$.

Given that Yankees have won i and lost j games so far, if we bet $B(i, j)$ on the Yankees for the next game, our payoff will be $P(i, j) + B(i, j)$ if the Yankees win, or $P(i, j) - B(i, j)$ if the Yankees lose. Therefore,

$$P(i + 1, j) = P(i, j) + B(i, j); \qquad (3.205)$$
$$P(i, j + 1) = P(i, j) - B(i, j). \qquad (3.206)$$

By adding and subtracting the equations (3.205) and

(3.206) we obtain that

$$P(i,j) = \frac{1}{2}(P(i+1,j) + P(i,j+1)) \quad (3.207)$$

$$B(i,j) = \frac{1}{2}(P(i+1,j) - P(i,j+1)) \quad (3.208)$$

Now, it is easy to compute all the entries of P, using (3.207) and working backwards from $P(4,j)$ and $P(i,4)$. For example,

$$P(3,3) = \frac{1}{2}(P(4,3) + P(3,4)) = \frac{1}{2}(100 - 100) = 0.$$

Once matrix P is computed, we use (3.208) to compute the matrix B. For example,

$$B(2,1) = \frac{1}{2}(P(3,1) - P(2,2)) = \frac{1}{2}(75 - 0) = 37.5.$$

In other words, given that Yankees have won 2 games and lost 1 game so far, and we would like to bet \$100 on the Yankees winning the World Series, we should bet \$37.5 on the Yankees for the next game. We include both matrices below:

$$P = \begin{pmatrix} 0 & -31.25 & -62.5 & -87.5 & -100 \\ 31.25 & 0 & -37.5 & -75 & -100 \\ 62.5 & 37.5 & 0 & -50 & -100 \\ 87.5 & 75 & 50 & 0 & -100 \\ 100 & 100 & 100 & 100 & 0 \end{pmatrix};$$

$$B = \begin{pmatrix} 31.25 & 31.25 & 25 & 12.5 \\ 31.25 & 37.5 & 37.5 & 25 \\ 25 & 37.5 & 50 & 50 \\ 12.5 & 25 & 50 & 100 \end{pmatrix}.$$

Therefore, we should bet $B(1,1) = 31.25$ dollars (on the Yankees) on the first game. Note that the probability of the Yankees winning or losing a single game does not affect your betting strategy.   □

Question 21. We have two red, two green and two yellow balls. For each color, one ball is heavy and the other is light. All heavy balls weigh the same. All light balls weigh the same. How many weighings on a scale are necessary to identify the three heavy balls?

Answer: It is clear that one weighing does not suffice. We show that the heavy balls can be identified in only two weighings.

Label the red balls R_1 and R_2, the green balls G_1 and G_2, and the yellow balls Y_1 and Y_2. Our first weighing is $\{R_1, G_1\}$ vs. $\{R_2, Y_1\}$.

If the scale is in balance, then either G_1 or Y_1 is heavy, but not both. Our second weighing is $\{G_1\}$ vs. $\{Y_1\}$. If G_1 is heavier, then the set of heavy balls is $\{R_2, G_1, Y_2\}$. If Y_1 is heavier, then the set of heavy balls is $\{R_1, G_2, Y_1\}$.

If $\{R_1, G_1\}$ is heavy, then either G_1 is heavy or Y_1 is light. Our second weighing is $\{G_1, Y_1\}$ vs. $\{G_2, Y_2\}$. If the scale is in balance, then G_1 is heavy; hence, the set of heavy balls is $\{R_1, G_1, Y_2\}$. If $\{G_1, Y_1\}$ is heavier, then G_1 and Y_1 are both heavy. The set of heavy balls is $\{R_1, G_1, Y_1\}$. If $\{G_2, Y_2\}$ is heavier, then G_2 and Y_2 are both heavy. The set of heavy balls is $\{R_1, G_2, Y_2\}$.

If $\{R_2, Y_1\}$ is heavy, then either Y_1 is heavy or G_1 is light. Our second weighing is $\{G_1, Y_1\}$ vs. $\{G_2, Y_2\}$. If the scale is in balance, then Y_1 is heavy; hence, the set of heavy balls is $\{R_2, G_2, Y_1\}$. If $\{G_1, Y_1\}$ is heavy, then G_1 and Y_1 are both heavy. The set of heavy balls is $\{R_2, G_1, Y_1\}$. If $\{G_2, Y_2\}$ is heavier, then G_2 and Y_2 are both heavy. The set of heavy balls is $\{R_2, G_2, Y_2\}$. $\square$

Question 22. There is a row of 10 rooms and a treasure in one of them. Each night, a ghost moves the treasure to an adjacent room. You are trying to find the treasure, but can only check one room per day. How do you find it?

Answer: The treasure can be found in at most 16 days. Label the rooms 1 through 10. Denote by T_k the room where the treasure is on day k, and let R_k denote the room you check on day k. Adopt the following strategy: for days $k = 1, 2, \ldots, 8$, let $R_k = k + 1$; for days $k = 9, 10, \ldots, 16$, let $R_k = 18 - k$.

If T_1 is even, then we will find the treasure in one of the first eight days. In other words, there exists k with $1 \leq k \leq 8$ such that $T_k = R_k$. Note that for $1 \leq k \leq 8$, T_k and R_k have the same parity, since T_1 is even, $R_1 = 2$, and both T_k and R_k change by at most 1 from day to day, according to ghost's moves and our strategy. Hence, $T_k - R_k$ is even. Furthermore, $T_1 \neq 1$ and $T_8 \neq 10$, which implies $T_1 - R_1 \geq 0$ and $T_8 - R_8 \leq 0$. Since $T_k - R_k$ can change by at most 2 from day to day, there must exist some k, $1 \leq k \leq 8$, such that $T_k - R_k = 0$.

If T_1 is odd, then we claim $T_k = R_k$ for some k, $9 \leq k \leq 16$. Note that T_9 is odd since T_1 is odd and the treasure is moved to an adjacent room each night. Furthermore, for $9 \leq k \leq 16$, T_k and R_k have the same parity since T_9 is odd, $R_9 = 9$, and both T_k and R_k change by at most 1 from day to day, according to ghost's moves and our strategy. Hence, $R_k - T_k$ is even. Furthermore, $T_9 \neq 10$ and $T_{16} \neq 1$, which implies $R_9 - T_9 \geq 0$ and $R_{16} - T_{16} \leq 0$. Since $R_k - T_k$ can change by at most 2 from day to day, there must exist some k, $9 \leq k \leq 16$, such that $R_k - T_k = 0$.

Finally, note that this strategy can be generalized to any number $n \geq 2$ of rooms, when the treasure will be found in $2n - 4$ days, by checking the rooms $2, 3, \ldots, n - 1, n - 1, n - 2, \ldots, 3, 2$, in that order. $\square$

Question 23. How many comparisons do you need to find the maximum in a set of n distinct numbers? How many comparisons do you need to find both the maximum and minimum in a set of n distinct numbers?

Answer: We can find the maximum in $n - 1$ comparisons as follows: let $\{x_1, x_2, \ldots, x_n\}$ denote the set of n distinct numbers. Scan the numbers from left to right, while maintaining the current maximum M. More precisely, set $x_1 = M$, and, in the ith comparison, compare M and x_{i+1}. If $M > x_{i+1}$, leave M as is; otherwise, set $M = x_{i+1}$. Do this for $i = 1, \ldots, n - 1$.

Similarly, one can find the minimum in a set of n distinct numbers in $n - 1$ comparisons.

Then, we can find the maximum M and the minimum m in $\{x_1, x_2, \ldots, x_n\}$ with $2n - 3$ comparisons: find the maximum M in $\{x_1, x_2, \ldots, x_n\}$ with $n - 1$ comparisons, and then find the minimum m in $\{x_1, x_2, \ldots, x_n\} \setminus \{M\}$ with $n - 2$ comparisons.

However, one can find m and M significantly faster.

If n is even, compare all $\frac{n}{2}$ consecutive pairs of numbers x_{2i-1} and x_{2i}, $i = 1 : \frac{n}{2}$, and put the smaller number into a set S and the larger number into a set L. This requires $\frac{n}{2}$ comparisons. Note that S and L have $\frac{n}{2}$ elements each. Then, find the minimum m in S using $\frac{n}{2} - 1$ comparisons, and find the maximum M in L using $\frac{n}{2} - 1$ comparisons. The total number of comparisons to find m and M is

$$\frac{n}{2} + \left(\frac{n}{2} - 1\right) + \left(\frac{n}{2} - 1\right) = \frac{3n}{2} - 2. \quad (3.209)$$

If n is odd, compare $\frac{n-1}{2}$ consecutive pairs of numbers x_{2i-1} and x_{2i}, $i = 1 : \frac{n-1}{2}$, and put the smaller number into a set S and the larger number into a set L. This requires $\frac{n-1}{2}$ comparisons. Place x_n into both S and L. Note that S and L have $\frac{n-1}{2} + 1 = \frac{n+1}{2}$ elements each. Then, find the minimum m in S using $\frac{n+1}{2} - 1 = \frac{n-1}{2}$ comparisons, and find the maximum M in L using $\frac{n+1}{2} - 1 = \frac{n-1}{2}$ comparisons. The total number of comparisons to find m and M is

$$\frac{n-1}{2} + \frac{n-1}{2} + \frac{n-1}{2} = \frac{3n+1}{2} - 2. \quad (3.210)$$

Note that the results of (3.209) and (3.210) can be written succinctly as

$$\left\lceil \frac{3n}{2} \right\rceil - 2 \text{ comparisons,}$$

where $\lceil x \rceil$ denotes the ceiling of x, the smallest integer greater than or equal to x. $\square$

Question 24. Given a cube, you can jump from one vertex to a neighboring vertex with equal probability. Assume you start from a certain vertex (does not matter which one). What is the expected number of jumps to reach the opposite vertex?

Answer: Label the vertices of the cube with 0, 1, 2, and 3, according to your distance from the opposite vertex. In other words, label the starting vertex with 3, the vertices adjacent to the starting vertex with 2, the vertices adjacent to the opposite vertex with 1, and the opposite vertex with 0. Call the opposite vertex your final destination.

Denote by E_i, $i = 0, 1, 2, 3$, the expected number of jumps yet to be made to reach the final destination, given that you are currently in one of the vertices labeled with i. Note that $E_0 = 0$, and we have to find E_3.

After the first jump, you are in one of the vertices labeled with 2, so

$$E_3 = 1 + E_2. \tag{3.211}$$

From a vertex labeled with 2, you can jump to three vertices: two of them are labeled with 1, and one of them is labeled with 3. Thus, you jump to a vertex labeled with 1 with probability $\frac{2}{3}$, or to a vertex labeled with 3 with probability $\frac{1}{3}$. Hence,

$$E_2 = 1 + \frac{2}{3}E_1 + \frac{1}{3}E_3. \tag{3.212}$$

Similarly, from one of the vertices labeled with 1, you jump to a vertex labeled with 2 with probability $\frac{2}{3}$, or to

a vertex labeled with 0 with probability $\frac{1}{3}$. Hence,

$$E_1 = 1 + \frac{2}{3}E_2 + \frac{1}{3}E_0 = 1 + \frac{2}{3}E_2, \qquad (3.213)$$

since $E_0 = 0$.

Solving (3.211–3.213) yields $E_1 = 7$, $E_2 = 9$, and $E_3 = 10$.

We conclude that it will take 10 jumps, on average, to reach the opposite vertex. $\square$

Question 25. Select numbers uniformly distributed between 0 and 1, one after the other, as long as they keep decreasing; i.e. stop selecting when you obtain a number that is greater than the previous one you selected.

(i) On average, how many numbers have you selected?

(ii) What is the average value of the smallest number you have selected?

Answer: We give three solutions for part (i); the third solution will be used to solve part (ii).

(i) *Solution 1:* Denote by $E(x)$ the expected number of numbers you have yet to select, given that you have just selected number x. For example, $E(0) = 1$, since the next number you select is greater than $x = 0$, upon which the game stops.

Assume that you have just selected number x. Denote by y the next number you select. We find $E(x)$ by conditioning on y. With probability $1 - x$, y is greater than x; the game stops, and, thus, you have selected only one number after selecting x. On the other hand, y could be smaller than x, in which case you expect to select $E(y)$ additional numbers after selecting y; in other words, you expect to select $1+E(y)$ additional numbers after selecting x. Since the probability density function of y is $f(y) = 1$,

the law of total probability gives

$$
\begin{aligned}
E(x) &= 1 \cdot (1 - x) + \int_0^x (1 + E(y)) f(y)\, dy \\
&= (1 - x) + \int_0^x (1 + E(y))\, dy \\
&= 1 + \int_0^x E(y)\, dy.
\end{aligned}
$$

Differentiating

$$
E(x) = 1 + \int_0^x E(y)\, dy
$$

with respect to x yields

$$
E'(x) = E(x).
$$

Thus, $E(x) = Ce^x$, where C is a constant. The condition $E(0) = 1$ gives $C = 1$. Hence, $E(x) = e^x$.

Note that the first number selected is automatically smaller than 1. Then, the number of numbers selected after starting with the number 1, which was denoted by $E(1)$, is equal to the number of numbers selected starting with a random number between 0 and 1. Therefore, the average number of numbers you have selected is $E(1) = e$.

(i) *Solution 2:* Denote by N the average number of numbers you have selected. Denote by x_i the ith number you selected, $i \geq 1$, and let p_i denote the probability that $x_i < x_{i-1} < \ldots < x_1$. Since there are $i!$ permutations of $x_1, \ldots, x_i$, then $p_i = \frac{1}{i!}$.

You select at least two numbers before stopping. The probability that you select exactly i numbers, $i \geq 2$, before stopping is equal to the probability that $x_{i-1} < \ldots < x_2 < x_1$ and $x_i > x_{i-1}$. The latter equals the probability that $x_{i-1} < \ldots < x_2 < x_1$ minus the probability that $x_i < x_{i-1} < \ldots < x_2 < x_1$; that is, $p_{i-1} - p_i$.

We conclude that the expected number of numbers you have selected is

$$
\begin{aligned}
N &= \sum_{i=2}^{\infty} i(p_{i-1} - p_i) \\
&= \sum_{i=2}^{\infty} i \left(\frac{1}{(i-1)!} - \frac{1}{i!} \right) \\
&= \sum_{i=2}^{\infty} i \cdot \frac{i-1}{i!} \\
&= \sum_{i=2}^{\infty} \frac{1}{(i-2)!} \\
&= \sum_{j=0}^{\infty} \frac{1}{j!} \\
&= e, \tag{3.214}
\end{aligned}
$$

where (3.214) follows from the Taylor series expansion of e^x around 0, i.e.

$$
e^x = \sum_{j=0}^{\infty} \frac{x^j}{j!}, \quad \forall \, x \in \mathbb{R},
$$

by letting $x = 0$.

(i) *Solution 3:* Denote by $p(x)\,dx$ the probability that a number between x and $x + dx$ is selected as part of the decreasing sequence. Let $p_i(x)\,dx$ denote the probability that a number between x and $x + dx$ is selected as the ith term of the decreasing sequence. Then,

$$
p(x)\,dx = \left(\sum_{i=1}^{\infty} p_i(x) \right) dx. \tag{3.215}
$$

The probability that a number between x and $x + dx$ is selected as the ith term of the decreasing sequence is

$$
p_i(x)\,dx = \frac{1}{(i-1)!} (1-x)^{i-1}\,dx, \tag{3.216}
$$

since $(1-x)^{i-1}$ is the probability that the first $i-1$ numbers selected are greater than x, and $\frac{1}{(i-1)!}$ is the probability that they are selected in decreasing order.

From (3.215) and (3.216), we obtain that

$$
\begin{aligned}
p(x)\,dx &= \sum_{i=1}^{\infty} \frac{(1-x)^{i-1}}{(i-1)!}\,dx \\
&= \sum_{j=0}^{\infty} \frac{(1-x)^j}{j!}\,dx \\
&= e^{1-x}\,dx;
\end{aligned}
\tag{3.217}
$$

where (3.217) follows from the Taylor series expansion of e^t around 0, i.e.

$$
e^t = \sum_{j=0}^{\infty} \frac{t^j}{j!}, \quad \forall\, t \in \mathbb{R},
$$

by letting $t = 1 - x$.

Therefore, the expected number of numbers selected in the decreasing sequence is

$$
\int_0^1 p(x)\,dx = \int_0^1 e^{1-x}\,dx = e - 1.
$$

Adding the last number selected (which is not in the decreasing sequence) gives an average of e numbers selected.

(ii) Denote by s the smallest number selected. It is the last number selected in the decreasing sequence. Since the probability that a number between x and $x + dx$ is selected as part of the decreasing sequence equals $e^{1-x}\,dx$, see (3.217), and since the probability that the next number selected is larger is $(1-x)$, then the probability that s is between x and $x + dx$ is $e^{1-x}(1-x)\,dx$.

Therefore, the expected value of the smallest number you have selected is

$$E(s) = \int_0^1 xe^{1-x}(1-x)\,dx$$

$$= \int_0^1 e^y y(1-y)\,dy, \qquad (3.218)$$

where (3.218) follows from the substitution $y = 1 - x$.

Using integration by parts to compute (3.218), we find that

$$E(s) = e^y(y-y^2)\Big|_0^1 - \int_0^1 e^y(1-2y)\,dy$$

$$= e^y(y-y^2)\Big|_0^1 - e^y\Big|_0^1 + 2\int_0^1 ye^y\,dy$$

$$= e^y(y-y^2-1)\Big|_0^1 + 2ye^y\Big|_0^1 - 2\int_0^1 e^y\,dy$$

$$= e^y(3y-y^2-1)\Big|_0^1 - 2e^y\Big|_0^1$$

$$= e^y(3y-y^2-3)\Big|_0^1$$

$$= 3 - e. \quad \square$$

Question 26. To organize a charity event that costs $100K, an organization raises funds. Independent of each other, one donor after another donates some amount of money that is exponentially distributed with a mean of $20K. The process is stopped as soon as $100K or more has been collected. Find the distribution, mean, and variance of the number of donors needed until at least $100K has been collected.

Answer: Denote by a_i the amount of money donated by donor i, that is exponentially distributed with mean $1/\lambda$. Let $s_n = \sum_{i=1}^n a_i$ be the total amount of money donated

by donors $1, \dots, n$, and let

$$N = \min_{n \geq 1}\{n \text{ such that } s_n \geq a\}$$

be the discrete random variable denoting the smallest index n such that s_n is at least a.

Denote by $P(n|a)$, $n \geq 1$, the probability mass function of N, that is, the probability that $N = n$ when a total of a needs to be raised. Note that

$$P(1|a) = P(a_1 \geq a) = e^{-\lambda a}. \tag{3.219}$$

We find $P(n|a)$, $n > 1$, by conditioning on a_1. Given that the first donor donated $a_1 = x < a$, N is equal to n if and only if the remaining amount $a - x$ is raised by the next $n - 1$ donors (and not by fewer than the next $n - 1$ donors), an event that by definition has probability $P(n - 1|a - x)$. Since the probability density function of a_1 is $f_{a_1}(x) = \lambda e^{-\lambda x}$, then, for $n > 1$, the law of total probability yields

$$\begin{aligned} P(n|a) &= \int_0^a P(n-1|a-x)f_{a_1}(x)\,dx \\ &= \int_0^a \lambda e^{-\lambda x} P(n-1|a-x)\,dx. \end{aligned} \tag{3.220}$$

We will prove that

$$P(n|a) = \frac{(\lambda a)^{n-1}}{(n-1)!}\,e^{-\lambda a}, \quad \forall\, a \geq 0, \tag{3.221}$$

by induction on n. The base case $n = 1$ was already established; see (3.219). Assume that (3.221) holds for $n > 1$; we will show that it also holds for $n + 1$.

From the induction hypothesis, we obtain that

$$P(n|a-x) = \frac{(\lambda(a-x))^{n-1}}{(n-1)!}\,e^{-\lambda(a-x)}, \tag{3.222}$$

for all $0 \leq x \leq a$. From (3.220), it follows that

$$P(n+1|a) = \int_0^a \lambda e^{-\lambda x} P(n|a-x) \, dx. \qquad (3.223)$$

From (3.222) and (3.223), we find that

$$
\begin{aligned}
P(n+1|a) \\
&= \int_0^a \lambda e^{-\lambda x} \cdot \frac{(\lambda(a-x))^{n-1}}{(n-1)!} \, e^{-\lambda(a-x)} \, dx \\
&= \frac{\lambda^n e^{-\lambda a}}{(n-1)!} \int_0^a (a-x)^{n-1} \, dx \\
&= \frac{(\lambda a)^n}{n!} \, e^{-\lambda a}.
\end{aligned}
$$

We conclude that (3.221) holds for $n+1$, and therefore (3.221) is proved by induction.

From (3.221), it follows that N has the same distribution as $1 + M$, where M has a Poisson distribution with mean λa. Then,

$$E[N] = 1 + \lambda a; \quad \text{Var}(N) = \lambda a.$$

For our problem, $1/\lambda = \$20\text{K}$ and $a = \$100\text{K}$. Thus, $E[N] = 1 + \lambda a = 6$ and $\text{Var}(N) = \lambda a = 5$.

We conclude that the number of donors needed until at least \$100K is collected has mean 6 and variance 5. $\quad \square$

Question 27. Consider a random walk starting at 1 and with equal probability of moving to the left or to the right by one unit, and stopping either at 0 or at 3.

(i) What is the expected number of steps to do so?

(ii) What is the probability of the random walk ending at 3 rather than at 0?

Answer: Denote by X_n^ℓ the position of the random walk at time n, where the superscript refers to the starting position of the walk; for example, $X_0^\ell = \ell$. In this question,

we are concerned with X_n^1. For a random walk starting at $\ell \in \{0, 1, 2, 3\}$, denote by T_ℓ the number of steps taken by the random walk in order to reach either 0 or 3 for the first time, and by $t_\ell = E[T_\ell]$ the expected number of steps.

(i) We are looking to find t_1. We derive a recurrence relation among the t_ℓ's as follows: for $\ell = 1, 2$,

$$
\begin{aligned}
t_\ell &= E[T_\ell] \\
&= E[T_\ell | X_1^\ell = \ell - 1] \cdot P(X_1^\ell = \ell - 1) \\
&\quad + E[T_\ell | X_1^\ell = \ell + 1] \cdot P(X_1^\ell = \ell + 1) \\
&= E[1 + T_{\ell-1}] \cdot P(X_1^\ell = \ell - 1) \quad (3.224) \\
&\quad + E[1 + T_{\ell+1}] \cdot P(X_1^\ell = \ell + 1) \quad (3.225) \\
&= (1 + t_{\ell-1}) \cdot \frac{1}{2} + (1 + t_{\ell+1}) \cdot \frac{1}{2}, \quad (3.226)
\end{aligned}
$$

where the following identities were used to derive (3.224) and (3.225):

$$
\begin{aligned}
E[T_\ell | X_1^\ell = \ell - 1] &= E[1 + T_{\ell-1}]; \quad (3.227) \\
E[T_\ell | X_1^\ell = \ell + 1] &= E[1 + T_{\ell+1}]. \quad (3.228)
\end{aligned}
$$

Note that (3.227) and (3.228) follow from the fact that, starting at ℓ, once the random walk took its first step to $\ell - 1$ (or $\ell + 1$), it becomes equivalent to a random walk starting afresh at $\ell - 1$ (or $\ell + 1$). The plus 1 term on the right hand side of (3.227) and (3.228) accounts for the first step.

Since $t_0 = t_3 = 0$, by letting $l = 1$ and then $l = 2$ in (3.226), we obtain the following linear system for t_1 and t_2:

$$
\begin{aligned}
t_1 &= \frac{1 + t_0}{2} + \frac{1 + t_2}{2} \\
&= \frac{1}{2} + \frac{1 + t_2}{2};
\end{aligned}
$$

$$t_2 = \frac{1 + t_1}{2} + \frac{1 + t_3}{2}$$

$$= \frac{1 + t_1}{2} + \frac{1}{2}.$$

Thus, $t_1 = 2$ and $t_2 = 2$.

We conclude that the expected number of steps before stopping is 2.

(ii) Denote by p_ℓ the probability that the random walk reaches 3 before it reaches 0 when its position is at ℓ, for $\ell \in \{0, 1, 2, 3\}$. We need to find p_1. Denote by τ_0^ℓ the first time the walk reaches 0 when starting at ℓ, and by τ_3^ℓ the first time the walk reaches 3 when starting at ℓ. We derive a recurrence relation for the p_ℓ's as follows: for $\ell = 1, 2$,

$$\begin{aligned}
p_\ell &= P\left(\tau_3^\ell < \tau_0^\ell\right) \\
&= P\left(\tau_3^\ell < \tau_0^\ell \,\middle|\, X_1^\ell = \ell + 1\right) P(X_1^\ell = \ell + 1) \\
&\quad + P\left(\tau_3^\ell < \tau_0^\ell \,\middle|\, X_1^\ell = \ell - 1\right) P(X_1^\ell = \ell - 1) \\
&= P\left(\tau_3^{\ell+1} < \tau_0^{\ell+1}\right) P(X_1^\ell = \ell + 1) \\
&\quad + P\left(\tau_3^{\ell-1} < \tau_0^{\ell-1}\right) P(X_1^\ell = \ell - 1) \\
&= \frac{1}{2} \cdot p_{\ell+1} + \frac{1}{2} \cdot p_{\ell-1}, \quad\quad\quad (3.229)
\end{aligned}$$

because the random walk is equally likely to move to the left or to the right and, once it moved, the random walk starts afresh.

Since $p_0 = 0$ and $p_3 = 1$, by letting $l = 1$ and then $l = 2$ in (3.229), we obtain the following linear system for p_1 and p_2:

$$p_1 = \frac{p_0}{2} + \frac{p_2}{2} = \frac{p_2}{2};$$

$$p_2 = \frac{p_1}{2} + \frac{p_3}{2} = \frac{p_1}{2} + \frac{1}{2}.$$

Thus, $p_1 = \frac{1}{3}$ and $p_2 = \frac{2}{3}$.

We conclude that the probability of the random walk ending at 3 rather than at 0 is $\frac{1}{3}$. $\square$

Question 28. A stick of length 1 drops and breaks at a random place uniformly distributed across the length. What is the expected length of the smaller part?

Answer:

Solution 1: Treating the stick as an interval $[0, 1]$, the breakpoint X becomes a random variable uniformly distributed on $(0, 1)$. Its probability density function $f_X(x)$ is 1, for $0 \leq x \leq 1$, and 0 otherwise. Denote by L the length of the smaller part. Then, $L = \min(x, 1 - x)$. We conclude that

$$
\begin{aligned}
E[L] &= \int_0^1 \min(x, 1 - x) \cdot f_X(x)\, dx \\
&= \int_0^1 \min(x, 1 - x)\, dx \\
&= \int_0^{1/2} x\, dx + \int_{1/2}^1 (1 - x)\, dx \\
&= \frac{1}{8} + \frac{1}{8} \\
&= \frac{1}{4}.
\end{aligned}
$$

Solution 2: Treating the stick as an interval $[0, 1]$, the breakpoint X becomes a random variable uniformly distributed on $(0, 1)$. Denote by L the length of the smaller part.

Let A be the event that the breakpoint X is in $\left(0, \frac{1}{2}\right)$. Then, $\overline{A}$ is the event that the breakpoint X is in $\left(\frac{1}{2}, 1\right)$. Clearly, $P(A) = P(\overline{A}) = \frac{1}{2}$. Given A, the length of the smaller part is X. Given $\overline{A}$, the length of the smaller part is $1 - X$.

The law of total probability yields

$$E[L] = E[L|A] \cdot P(A) + E[L|\overline{A}] \cdot P(\overline{A})$$

$$= E[X|A] \cdot \frac{1}{2} + E[1 - X|\overline{A}] \cdot \frac{1}{2}$$

$$= E[X|A] \cdot \frac{1}{2} + \frac{1}{2} - E[X|\overline{A}] \cdot \frac{1}{2}. \quad (3.230)$$

Note that, given A, X is uniformly distributed on $\left(0, \frac{1}{2}\right)$, and, Given $\overline{A}$, X is uniformly distributed on $\left(\frac{1}{2}, 1\right)$. Since the expectation of a random variable uniformly distributed on an interval (a, b) is equal to $\frac{a+b}{2}$, we obtain that

$$E[X|A] = \frac{1}{4} \quad \text{and} \quad E[X|\overline{A}] = \frac{3}{4}.$$

Then, (3.230) yields

$$E[L] = \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{2} - \frac{3}{4} \cdot \frac{1}{2}$$

$$= \frac{1}{4}. \quad \square$$

Question 29. You are given a stick of unit length.

(i) The stick drops and breaks at two places. What is the probability that the three pieces could form a triangle?

(ii) The stick drops and breaks at one place. Then the larger piece is taken and dropped again, breaking at one place. What is the probability that the three pieces could form a triangle?

Answer: We offer two solutions for part (i); the second solution will be used to solve part (ii).

(i) *Solution 1:* Denote by X and Y the two break points, and assume X and Y are independent random variables uniformly distributed on $(0, 1)$. To form a triangle, the sum of the lengths of any two pieces must be greater than

the length of the third piece. Equivalently, each piece must be of length less than $1/2$.

Assume that $Y > X$. Then, the length of the three pieces are X, $Y - X$, and $1 - Y$. Each of these pieces is of length less than $1/2$ if and only if the point (X, Y) belongs to the region $\{(x, y) : x < 1/2, y - x < 1/2, 1 - y < 1/2, x \in (0, 1), y \in (0, 1), x < y\}$ in the unit square. Since the area of this region is $1/8$, the probability of the three pieces forming a triangle, given $Y > X$, is $1/8$. Symmetrically, the probability of the three pieces forming a triangle, given $Y < X$, is $1/8$. Events $\{Y > X\}$ and $\{Y < X\}$ are disjoint; hence, the probability of the three pieces forming a triangle is $1/4$.

(i) *Solution 2:* Consider an equilateral triangle ABC with height of length 1. Given a point P in its interior, let h_a, h_b, and h_c be the lengths of the perpendiculars dropped from P to the sides BC, CA, and AB, respectively. Since the areas of triangles BPC, CPA, and APB sum up to the area of ABC, we conclude that $h_a + h_b + h_c = 1$, and, thus, is independent of the position of P. Breaking a stick of length 1 into three pieces of lengths h_a, h_b, and h_c, is clearly equivalent to (uniquely) specifying a point P in the interior of the triangle ABC.

Connect the midpoints A', B', C', of the sides of triangle ABC to split it into four congruent equilateral triangles, with the medial triangle $A'B'C'$ in the middle (see Figure 3.3). Each piece of the broken stick has length less than $1/2$ if and only if the corresponding point P belongs to the medial triangle. Since the area of the medial triangle is $1/4$ of the area of triangle ABC, the desired probability is $1/4$.

(ii) Assume that the pieces have lengths h and $(1 - h)$ after the first break, with $h < (1 - h)$, i.e., $h < 1/2$. With h fixed, the (larger) piece of length $1 - h$ is taken and dropped again, breaking at one place uniformly at random. The probability that the three pieces thus obtained
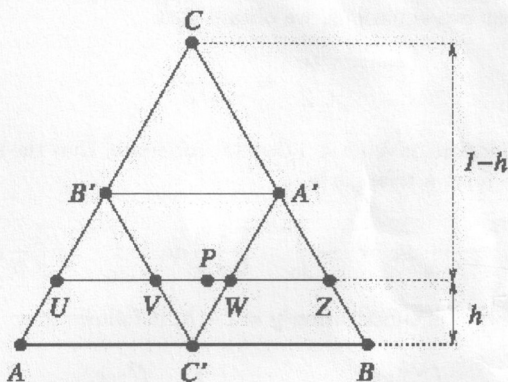
Figure 3.3: Point P inside the medial triangle $A'B'C'$ of the equilateral triangle ABC.

form a triangle clearly depends on h. It is close to 0 for h close to 0, and it is close to 1 for h close to 1/2.

More precisely, using the representation from the second solution of part (a) above, the probability that the three pieces form a triangle, given a fixed $h < 1/2$, is equal to the probability that the point P, that lies on the segment UZ parallel to side AB and at distance h from it, belongs to the segment VW, that is the intersection of UZ with the medial triangle $A'B'C'$ (see Figure 3.3). Since P is chosen (by the second break of the stick) from UZ uniformly at random, the probability that P belongs to VW is equal to the ratio of their lengths, namely $\frac{VW}{UZ}$.

Next, we express this ratio in terms of h. First, note that the medial triangle $A'B'C'$ has side length $AB/2$ and height 1/2. Since the triangles ABC and UZC are similar, we have $\frac{UZ}{AB} = \frac{1-h}{1}$. Since the triangles WVC'

and $A'B'C'$ are similar, we have $\frac{VW}{A'B'} = \frac{h}{1/2}$. Dividing the last two equations, we obtain that

$$\frac{VW}{UZ} = \frac{h}{1-h}.$$

Therefore, given $h < 1/2$, the probability that the three pieces form a triangle is

$$\int_0^{1/2} \frac{h}{1-h}\, dh = -\int_0^{1/2} \frac{1-h}{1-h}\, dh + \int_0^{1/2} \frac{1}{1-h}\, dh.$$

By using the substitution $y = 1 - h$, it follows that

$$\int_0^{1/2} \frac{h}{1-h}\, dh = -\frac{1}{2} + \int_{1/2}^1 \frac{1}{y}\, dy$$

$$= -\frac{1}{2} - \ln\left(\frac{1}{2}\right)$$

$$= \ln 2 - \frac{1}{2}.$$

Since the probability that $h < 1/2$, where h is chosen uniformly at random from $(0,1)$ (by the first break of the stick), is $1/2$, then the total probability that the three pieces form a triangle is

$$\frac{\ln 2 - 1/2}{1/2} = 2\ln 2 - 1. \quad \square$$

Question 30. Why is a manhole cover round?

Answer: A circle is the shape with minimal surface given a required minimal width in any direction. Moreover, the cover of a round manhole cannot fall through the hole. If the manhole were square, its cover turned on its edge could fall through the hole since the diagonal of a square is $\sqrt{2}$ times larger than its edge. $\square$

Question 31. When is the first time after 12 o'clock that the hour and minute hands of a clock meet again?

Answer: The minute hand moves at a speed of 360 degrees per hour, while the hour hand moves at a speed of 30 degrees per hour. They start together at 12 o'clock. The first time they meet, the minute hand made one full rotation more than the hour hand, which is the same as 360 degrees more than the hour hand. If t denotes the time (measured in hours) until the two hands meet again, this can be written as

$$360 \cdot t = 30 \cdot t + 360.$$

Thus, $t = \frac{12}{11}$ hours, which is approximately 1 hour, 5 minutes, and 27 seconds. $\square$

Question 32. Three light switches are in one room, and they turn three light bulbs in another. How do you figure out which switch turns on which bulb in one shot?

Answer: Turn on two switches for a couple of minutes, and then turn one of the switches off and go into the other room. The bulb that is lit corresponds to the switch that is still on; the bulb that is not lit but is hot corresponds to the switch that was turned on and then turned off; the bulb that is not lit and is cold corresponds to the switch that was never turned on. $\square$

POCKET BOOK GUIDES FOR
QUANT INTERVIEWS

FE PRESS

New York

Pocket Book Guides for Quant Interviews

1. 150 Most Frequently Asked Questions on Quant Interviews, by Dan Stefanica, Radoš Radoičić, and Tai-Ho Wang. FE Press, 2013

2. 50 Challenging Brainteasers from Quant Interviews, by Radoš Radoičić, Dan Stefanica, and Tai-Ho Wang. FE Press, 2014

3. Stochastic Calculus & Probability Quant Interview Questions, by Tai-Ho Wang, Radoš Radoičić, and Dan Stefanica. FE Press, 2014

Other Titles from FE Press

1. A Primer for the Mathematics of Financial Engineering, Second Edition, by Dan Stefanica. FE Press, 2011

2. Solutions Manual – A Primer for the Mathematics of Financial Engineering, Second Edition, by Dan Stefanica. FE Press, 2011

3. Numerical Linear Algebra Primer for Financial Engineering, by Dan Stefanica. FE Press, 2014

4. Solutions Manual – Numerical Linear Algebra Primer for Financial Engineering, by Dan Stefanica. FE Press, 2014

5. A Probability Primer for Mathematical Finance, by Elena Kosygina.

Bibliography

[1] Avner Friedman. *Stochastic Differential Equations and Applications*. Dover Publications, Mineola, New York, 2006.

[2] Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag New York, Inc., New York, 2004.

[3] Dan Stefanica. *A Mathematical Primer with Numerical Methods for Financial Engineering*. FE Press, New York, 2nd edition, 2011.

[4] Paul Wilmott. *Frequently Asked Questions in Quantitative Finance*. John Wiley & Sons Ltd, Chichester, West Sussex, 2nd edition, 2009.