Chapter 1

First Look: Fifteen Questions.

1. Nine months call options with strikes 20 and 25 on a non–dividend–paying underlying asset with spot price \$22 are trading for \$5.50 and \$1, respectively. Can you find an arbitrage?

2. (i) What is the sum of the eigenvalues of the correlation matrix of n random variables?

   (ii) Find a lower bound for the sum of the eigenvalues of the inverse of a nonsingular correlation matrix of n random variables.

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. An 8×8 matrix contains zeros and ones. You may repeatedly choose any 3×3 or 4×4 block and flip all bits in the block (that is, convert zeros to ones,

and ones to zeros). Can you always modify the original matrix into an all–zero matrix using these block flips?

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. Given a sample of size 1 from the normal distribution with mean μ and variance σ^2, with both μ and σ are unknown, give a *finite* confidence interval for σ^2 with confidence level at least 99%.

7. How would you generate uniformly distributed points on the surface of the 3-dimensional unit sphere?

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

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

with $\lambda > 0$, which is used, e.g., in the Vasicek model for interest rates.

10. Find all the integer solutions of the equation

$$x^3 + y^3 = 2013.$$

11. Let X and Y be standard normal variables with joint normal distribution with correlation ρ. Find the expectation

$$\mathbb{E}\left[\operatorname{sgn}(X)\operatorname{sgn}(Y)\right],$$

where $\operatorname{sgn}(\cdot)$ is the sign function given by $\operatorname{sgn}(x) = 1$, if $x > 0$, $\operatorname{sgn}(x) = -1$, if $x < 0$, and $\operatorname{sgn}(0) = 0$.

12. How do you create a long Gamma, short vega options trading strategy?

13. Let X_t and Y_t be geometric Brownian motions driven by

$$\begin{aligned} \frac{dX_t}{X_t} &= \mu_X dt + \sigma_X dW_t; \\ \frac{dY_t}{Y_t} &= \mu_Y dt + \sigma_Y dB_t, \end{aligned}$$

where W_t and B_t are correlated Brownian motions with constant correlation ρ. Show that

$$Z_t = \frac{X_t}{Y_t}$$

is also a geometric Brownian motion and determine its drift and volatility coefficients.

14. Find the k–th largest element in an unsorted array. Assume that k is always valid, i.e., $k \geq 1$ and k is less than or equal to the length of the array.

Note: You are looking for the k–th largest element in the sorted order, not the k–th distinct element of the array.

Example 1:

```
Input: [3,2,1,5,6,4] and k = 2
Output: 5
```

Example 2:

```
Input: [3,2,3,1,2,4,5,5,6] and k = 4
Output: 4
```

15. Given an array *nums*, there is a sliding window of size k which is moving from the very left of the array to the very right of the array. You can only see the k numbers in the window. Each time the sliding window moves right by one position. Assume that k is always valid, i.e., $k \geq 1$ and k is less than or equal to the size of the input array size for non-empty arrays.

    Write an algorithm that returns the maximum of the sliding window.

    Example:

    ```
    Input: nums = [1,3,-1,-3,5,3,6,7], and k = 3
    Output: [3,3,5,5,6,7]
    ```

    Explanation:

    ```
    Window Position                Max
    --------------------------    -----
    [1  3  -1] -3  5  3  6  7       3
    ```

```
1 [3  -1  -3] 5  3  6  7        3
1  3 [-1  -3  5] 3  6  7        5
1  3  -1 [-3  5  3] 6  7        5
1  3  -1  -3 [5  3  6] 7        6
1  3  -1  -3  5 [3  6  7]       7
```

Solutions

Question 1. Nine months call options with strikes 20 and 25 on a non–dividend–paying underlying asset with spot price \$22 are trading for \$5.50 and \$1, respectively. Can you find an arbitrage?

Answer: Note that a call option with strike 0 on a non–dividend–paying underlying asset is the same as one unit of the asset, since the call with strike 0 will always be exercised at maturity by paying \$0, i.e., the strike of the option, to receive one unit of the asset. Thus, we are implicitly given a third call option with strike $K = 0$ and price \$22 (i.e., the spot price of the asset), and we can proceed to identify whether there is convexity arbitrage for these three call options.

Let $K_1 = 0$, $K_2 = 20$, $K_3 = 25$ and $C_1 = 22$, $C_2 = 5.50$, $C_3 = 1$. Note that $20 = \frac{1}{5} \cdot 0 + \frac{4}{5} \cdot 25$, i.e.,

$$K_2 \;=\; \frac{1}{5}K_1 + \frac{4}{5}K_3.$$

Since

$$\frac{1}{5}C_1 + \frac{4}{5}C_3 \;=\; 5.20 \;<\; 5.50 \;=\; C_2, \qquad (1.1)$$

the convexity of option prices with respect to strike is violated.

The arbitrage strategy is to "buy low" $\frac{1}{5}C_1 + \frac{4}{5}C_3$ and "sell high" C_2. To normalize units, we multiply the positions by 500 to obtain the following arbitrage strategy: "buy low" $100C_1 + 400C_2$ and "sell high" $500C_2$. Note that buying $100C_1$, i.e., 100 calls with strike $K_1 = 0$, is equivalent to buying 100 units of the underlying asset since the asset does not pay dividends.

Arbitrage Strategy:
- buy 100 units of the underlying asset for \$2,200;
- buy 400 calls with strike $K_3 = 25$ for \$400;

• sell 500 calls with strike $K_2 = 20$ for \$2,750;
• realize a positive cash flow of \$150.

The positive cash flow \$150 represents risk–free profit since the arbitrage portfolio does not lose money at maturity:

The value of the arbitrage portfolio at the maturity T of the options is

$$\begin{aligned} V(T) &= 100S(T) \;-\; 500C_2(T) \;+\; 400C_3(T) \\ &= 100S(T) \;-\; 500\max(S(T)-20,0) \\ &\quad +\; 400\max(S(T)-25,0). \end{aligned}$$

If $S(T) \leq 20$,

$$V(T) \;=\; 100S(T) \;\geq\; 0.$$

If $20 < S(T) \leq 25$,

$$\begin{aligned} V(T) &= 100S(T) \;-\; 500(S(T)-20) \\ &= 10000 - 400S(T) \\ &\geq 0. \end{aligned}$$

If $25 < S(T)$,

$$\begin{aligned} V(T) &= 100S(T) \;-\; 500(S(T)-20) \\ &\quad +\; 400(S(T)-25) \\ &= 0. \end{aligned}$$

Note that $150 = 500 \cdot (5.50 - 5.20)$, i.e., the risk–free profit \$150 is equal to the size of the convexity disparity $\$5.50 - \$5.20$ times the amplifier factor 500. $\square$

Question 2. (i) What is the sum of the eigenvalues of the correlation matrix of n random variables?

(ii) Find a lower bound for the sum of the eigenvalues of the inverse of a nonsingular correlation matrix of n random variables.

Answer: (i) The sum of the eigenvalues of a matrix is equal to the trace of the matrix, i.e., to the sum of the main diagonal entries of the matrix.[1] Since the correlation matrix of n random variables is an $n \times n$ matrix with all main diagonal entries equal to 1, the trace of the correlation matrix is equal to n. We conclude that the sum of the eigenvalues of the correlation matrix of n random variables is n.

(ii) If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of the nonsingular $n \times n$ correlation matrix Ω, then $\lambda_i > 0$ for all $i = 1 : n$, since a nonsingular correlation matrix is symmetric positive definite. The eigenvalues of the inverse matrix Ω^{-1} are $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_n}$. Thus, the question asks us what could be said about the sum

$$\sum_{i=1}^{n} \frac{1}{\lambda_i}$$

of the eigenvalues of Ω^{-1}.

Recall from (i) that the sum of the eigenvalues of the correlation matrix Ω is n, i.e.,

$$\sum_{i=1}^{n} \lambda_i \; = \; n. \tag{1.2}$$

Also, recall from the Cauchy–Schwartz inequality that

$$\left(\sum_{i=1}^{n} a_i^2\right) \left(\sum_{i=1}^{n} b_i^2\right) \; \geq \; \left(\sum_{i=1}^{n} a_i b_i\right)^2. \tag{1.3}$$

Since $\lambda_i > 0$ for all $i = 1 : n$, we can use the Cauchy–Schwartz inequality (1.3) for $a_i = \sqrt{\lambda_i}$ and $b_i = \frac{1}{\sqrt{\lambda_i}}$,

[1] This property follows from the fact that the eigenvalues of a matrix A are the roots of the characteristic polynomial $P_A(t) = \det(tI - A)$ of the matrix A; see, e.g., Theorem 4.1 from Stefanica [5].

$i = 1 : n$, to obtain that

$$\left(\sum_{i=1}^{n} \lambda_i\right) \left(\sum_{i=1}^{n} \frac{1}{\lambda_i}\right) \;\geq\; \left(\sum_{i=1}^{n} 1\right)^2 \;=\; n^2, \qquad (1.4)$$

since $a_i^2 = \left(\sqrt{\lambda_i}\right)^2 = \lambda_i$, $b_i^2 = \left(\frac{1}{\sqrt{\lambda_i}}\right)^2 = \frac{1}{\lambda_i}$, and $a_i b_i = \sqrt{\lambda_i} \cdot \frac{1}{\sqrt{\lambda_i}} = 1$, for all $i = 1 : n$.

From (1.4) and using (1.2), we find that

$$n \left(\sum_{i=1}^{n} \frac{1}{\lambda_i}\right) \;\geq\; n^2$$

and therefore conclude that

$$\sum_{i=1}^{n} \frac{1}{\lambda_i} \;\geq n.$$

In other words, the sum of the eigenvalues of the inverse of a nonsingular correlation matrix of n random variables is bounded from below by n. $\square$

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

$$X_t \;=\; \int_0^t W_\tau \, d\tau. \qquad (1.5)$$

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

Answer: Note that we can rewrite (1.5) 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 3.7.

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. An 8×8 matrix contains zeros and ones. You may repeatedly choose any 3×3 or 4×4 block and flip all bits in the block (that is, convert zeros to ones, and ones to zeros). Can you always modify the original matrix into an all–zero matrix using these block flips?

Answer: No! Note that all the block flips are reversible, so it will suffice to show that there exist 8×8 matrices M

containing zeroes and ones that cannot be obtained using the block flips starting from an all–zero matrix.

Given a multiset of 3×3 and 4×4 blocks to be flipped in some order, the final matrix obtained is independent of the order in which the flips of the blocks in the multiset are applied. Moreover, we can remove all the block repetitions; in other words, we can reduce the multiset of the blocks flipped to a set with no repeated blocks by recognizing that flipping the same block twice does not affect the final matrix obtained at the end.

The total number of 3×3 blocks in an 8×8 matrix is 36: the upper left corner of the 3×3 block cannot be located in the 7–th or 8–th row or in the 7–th or 8–th column of the 8×8 matrix and therefore there are $6 \times 6 = 36$ possible positions for it. Similarly, the total number of 4×4 blocks in an 8×8 matrix is 25: the upper left corner of the 4×4 cannot be located in the 6–th, 7–th or 8–th row or in the 6–th, 7–th or 8–th column of the 8×8 matrix and therefore there are $5 \times 5 = 25$ possible positions for it.

Thus, there are $36+25 = 61$ blocks that can be flipped and the total number of different sets of blocks made with these 61 blocks (with no repeated blocks) is 2^{61}. Then, starting with an all–zero matrix, we can obtain at most 2^{61} distinct matrices. Since the total number of 8×8 matrices containing zeros and ones is 2^{64}, it follows that there exist matrices that cannot be obtained starting from an all–zero matrix by using block flips. $\square$

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

$$\Omega \ = \ \begin{pmatrix} 1 & 0.6 & -0.3 \\ 0.6 & 1 & \rho \\ -0.3 & \rho & 1 \end{pmatrix} \tag{1.6}$$

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 (1.6) has the following principal minors:

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

$$\begin{aligned} \det\begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix} &= 0.64; \\ \det\begin{pmatrix} 1 & -0.3 \\ -0.3 & 1 \end{pmatrix} &= 0.91; \\ \det\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} &= 1-\rho^2; \end{aligned}$$

$$\begin{aligned} \det(\Omega) &= 1 - 0.36\rho - 0.09 - 0.36 - \rho^2 \\ &= 0.55 - 0.36\rho - \rho^2. \end{aligned}$$

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

$$\begin{aligned} 1-\rho^2 &\geq 0; \\ 0.55-0.36\rho-\rho^2 &\geq 0, \end{aligned}$$

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

$$\rho^2 + 0.36\rho - 0.55 \leq 0. \tag{1.7}$$

Since the roots of the quadratic equation corresponding to (1.7) 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

$$-0.9432 \leq \rho \leq 0.5832. \quad \square \tag{1.8}$$

Question 6. Given a sample of size 1 from the normal distribution with mean μ and variance σ^2, with both μ and σ are unknown, give a *finite* confidence interval for σ^2 with confidence level at least 99%.

Answer: Denote by X the single observation from the normal distribution with mean μ and variance σ^2, where both μ and σ are unknown. We construct a confidence interval $[0, T(X)]$ for σ^2, where $T(\cdot)$ denotes some statistic. This interval will be a confidence interval with confidence level at least 99% if for every μ and $\sigma^2 > 0$:

$$\mathbb{P}_{\mu,\sigma^2}\left(\sigma^2 > T(X)\right) < 0.01.$$

Note that the probability density function $f_X(x)$ of X satisfies

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \leq \frac{1}{\sqrt{2\pi}\,\sigma}, \quad \forall\, x \in \mathbb{R}.$$

Then, for every $a \geq 0$, we have that

$$\begin{aligned} \mathbb{P}(|X| \leq a) &= \int_{-a}^{a} f_X(x)\,dx \leq \frac{2a}{\sqrt{2\pi}\,\sigma} \\ &\leq \frac{a}{\sigma}. \end{aligned} \tag{1.9}$$

By letting $a = 0.01\sigma$ in (1.9) we obtain that

$$\mathbb{P}\left(|X| \leq 0.01\sigma\right) \;\leq\; 0.01. \tag{1.10}$$

Note that

$$\begin{aligned} \mathbb{P}\left(|X| \leq 0.01\sigma\right) &= \mathbb{P}\left(X^2 \leq 0.0001\sigma^2\right) \\ &= \mathbb{P}\left(\sigma^2 \geq 10000X^2\right). \end{aligned} \tag{1.11}$$

From (1.10) and (1.11), we find that

$$\mathbb{P}\left(\sigma^2 \geq 10000X^2\right) \;\leq\; 0.01.$$

We conclude that $\left[0, 10000X^2\right]$ is a *finite* confidence interval for σ^2 with confidence level at least 99%. $\square$

Question 7. How would you generate uniformly distributed points on the surface of the 3-dimensional unit sphere?

Answer: We will describe two different methods to accomplish this task.

Method 1: Spherical coordinates provide a mapping from every point $P(x, y, z)$ on the surface of the 3-dimensional unit sphere to a pair of angles (θ, ϕ), where $\theta \in [0, 2\pi]$ is the azimuthal angle and $\phi \in [0, \pi]$ is the polar angle, via transformations: $x = \sin(\phi)\cos(\theta)$, $y = \sin(\phi)\sin(\theta)$, $z = \cos(\phi)$.

A tempting way to try generating uniformly distributed points on the surface of the 3-dimensional unit sphere would be to generate both θ and ϕ angles uniformly at random from their respective intervals, and then apply the transformation above. However, this algorithm is incorrect, as the points generated by it will be clustered around the poles ($\phi = 0$ and $\phi = \pi$) while sparse around the equator ($\phi = \pi/2$).

Why is that so? The reason is that the Jacobian of the transformation above is equal to $\sin(\phi)$. In other words, the differential surface element dA in spherical coordinates is not $d\phi\, d\theta$, but rather $\sin(\phi) d\phi\, d\theta$. So, close to the poles of the sphere (i.e., when $\phi = 0$ or $\phi = \pi$), the differential surface element gets smaller as $\sin(\phi) \to 0$.

Our task is, hence, a bit more delicate: we have to find and then draw samples from a probability distribution with joint density $f(\theta, \phi)$ that maps from the (θ, ϕ)–plane to a uniform distribution on the unit sphere. Since for every $P(x, y, z)$ on the surface of the 3-dimensional unit sphere $f(P)$ has to be constant for a uniform distribution, we obtain that $f(P) = \frac{1}{4\pi}$, since the surface area of the unit sphere is 4π. Therefore,

$$f(P)\, dA \;=\; \frac{1}{4\pi}\, dA \;=\; f(\theta, \phi)\, d\theta\, d\phi.$$

Since $dA = \sin(\phi) d\phi\, d\theta$, it follows that

$$f(\theta, \phi) \;=\; \frac{1}{4\pi} \sin(\phi).$$

Integrating the joint density $f(\theta, \phi)$ to get the marginal densities of θ and ϕ separately, we find that

$$\begin{aligned} f(\theta) = &\int_0^{\pi} f(\theta, \phi)\, d\phi \;=\; \frac{1}{2\pi}, \\ f(\phi) = &\int_0^{2\pi} f(\theta, \phi)\, d\theta \;=\; \frac{\sin(\phi)}{2}. \end{aligned}$$

Clearly, θ is uniformly distributed over $[0, 2\pi]$, and, hence, θ can be sampled as 2π times the output from a readily available uniform random generator in $[0, 1]$. How do we, however, use the same generator to sample ϕ from a probability distribution with density $f(\phi) = \frac{\sin(\phi)}{2}$? We use the inverse transform sampling method.

Note that the cumulative distribution function (cdf) for the distribution of ϕ is

$$F(\phi) \;=\; \int_0^{\phi} f(s)\, ds \;=\; \frac{1}{2}\left(1 - \cos(\phi)\right).$$

The function $F(\phi)$ is strictly increasing from $[0, \pi]$ to $[0, 1]$, and, as such, has an inverse function $F^{-1}(u)$ given by

$$F^{-1}(u) = \arccos(1 - 2u).$$

Let $U \sim U[0, 1]$ be uniformly distributed over $[0, 1]$. Then,

$$\mathbb{P}(U \leq F(\phi)) = F(\phi),$$

and

$$\mathbb{P}\left(F^{-1}(U) \leq \phi\right) = F(\phi).$$

Therefore, $F(\phi)$ is the cdf of the random variable $F^{-1}(U)$. In other words, $F^{-1}(U)$ has the same probability distribution as ϕ. Hence, to sample ϕ from a probability distribution with density $f(\phi) = \frac{\sin(\phi)}{2}$, we generate a uniform random number U from $[0, 1]$ using a readily available uniform random generator, then compute $\phi = \arccos(1 - 2U)$.

Finally, once we have sampled θ and ϕ, then x, y, and z are computed using the spherical transformation.

Method 2: Assume that we have available a random generator from the standard normal distribution, such as the Box–Muller method; see, e.g., Glasserman [2]. Generate three dependent standard normally distributed numbers X, Y, and Z to form a vector $\vec{v} = (X, Y, Z)$. Intuitively, this vector will point in a uniformly random direction in the 3-dimensional space. Next, we normalize the vector by dividing it by its norm, to obtain the point $P = \left(\frac{X}{\|\vec{v}\|}, \frac{Y}{\|\vec{v}\|}, \frac{Z}{\|\vec{v}\|}\right)$ on the unit sphere. In order to show that P is uniformly distributed on the surface of the unit sphere, it suffices to prove that $\vec{v}$ truly points in a uniformly random direction.

As X, Y, and Z are each sampled independently from the standard normal distribution, the probability density function of (X, Y, Z) is given by the product of their marginal densities:

$$\begin{aligned} & f(x,y,z) \;=\; f(\vec{v}) \\ = & \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\right) \\ = & \frac{1}{(2\pi)^{3/2}} e^{-\frac{1}{2}(x^2+y^2+z^2)} \\ = & \frac{1}{(2\pi)^{3/2}} e^{-\frac{1}{2}\|\vec{v}\|^2}. \end{aligned}$$

In other words, the probability density function of $\vec{v}$ depends only on its norm and not on any angles such as θ or ϕ. In conclusion, by finding where the ray $\vec{v}$ intersects the unit sphere, we obtain a sample from a uniform distribution on the surface of the unit sphere. □

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

$$dr_t \;=\; \lambda(\theta - r_t)dt \;+\; \sigma dW_t, \tag{1.12}$$

with $\lambda > 0$, which is used, e.g., in the Vasicek model for interest rates.

Answer: We can rewrite (1.12) as

$$dr_t + \lambda r_t dt \;=\; \lambda\theta dt + \sigma dW_t. \tag{1.13}$$

By multiplying (1.13) 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

$$d\left(e^{\lambda t} r_t\right) \;=\; \lambda\theta e^{\lambda t} dt + \sigma e^{\lambda t} dW_t. \tag{1.14}$$

By integrating (1.14) 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,

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

Question 10. Find all the integer solutions of the equation

$$x^3 + y^3 \; = \; 2013.$$

Answer: The equation has no integer solutions.

The challenge in this problem comes from the fact that we can write the equation as

$$(x+y)(x^2 - xy + y^2) \; = \; 3 \cdot 11 \cdot 61,$$

which means that $x + y$ has 16 possible values which are the positive and negative divisors of $2013 = 3 \cdot 11 \cdot 61$. This would lead to a long–winded solution.

However, there is a straightforward way to see that the equation has no integer solutions, by looking at residuals modulo 9.

Note that $2013 \equiv 6 \pmod 9$; here,

$$a \equiv b \pmod m \quad \Longleftrightarrow \quad m \mid (a-b),$$

where $m > 1$ is a positive integer and a and b are integers.

Furthermore:

- if $a \equiv 0 \pmod 3$, then $a^3 \equiv 0 \pmod 9$;

• if $a \equiv 1 \pmod 3$, then $a^3 \equiv 1 \pmod 9$;
• if $a \equiv 2 \pmod 3$, then $a^3 \equiv 8 \pmod 9$.

This means that $x^3 + y^3$ can only be equal to 0, 1, 2, 7 or 8 modulo 9, and cannot be equal to 2013 for any integers x and y since $2013 \equiv 6 \pmod 9$. $\square$

Question 11. Let X and Y be standard normal variables with joint normal distribution with correlation ρ. Find the expectation

$$\mathbb{E}\left[\operatorname{sgn}(X)\operatorname{sgn}(Y)\right],$$

where $\operatorname{sgn}(\cdot)$ is the sign function given by $\operatorname{sgn}(x) = 1$, if $x > 0$, $\operatorname{sgn}(x) = -1$, if $x < 0$, and $\operatorname{sgn}(0) = 0$.

Answer: If $\rho = 1$, then

$$\mathbb{E}\left[\operatorname{sgn}(X)\operatorname{sgn}(Y)\right] = \mathbb{E}\left[\operatorname{sgn}(Z)^2\right] = \mathbb{E}\left[1\right] = 1, \tag{1.15}$$

where Z is the standard normal variable, and, if $\rho = -1$,

$$\mathbb{E}\left[\operatorname{sgn}(X)\operatorname{sgn}(Y)\right] = \mathbb{E}\left[-\operatorname{sgn}(Z)^2\right] = \mathbb{E}\left[-1\right] = -1. \tag{1.16}$$

If $\rho \in (-1, 1)$, we obtain that

$$\begin{aligned} & \mathbb{E}\left[\operatorname{sgn}(X)\operatorname{sgn}(Y)\right] \\ = \; & \mathbb{P}\left[X > 0, Y > 0\right] + \mathbb{P}\left[X < 0, Y < 0\right] \\ & - \mathbb{P}\left[X > 0, Y < 0\right] - \mathbb{P}\left[X < 0, Y > 0\right]. \end{aligned} \tag{1.17}$$

Note that

$$\begin{aligned} \mathbb{P}\left[X > 0, Y > 0\right] &= \mathbb{P}\left[X < 0, Y < 0\right]; & (1.18) \\ \mathbb{P}\left[X > 0, Y < 0\right] &= \mathbb{P}\left[X < 0, Y > 0\right], & (1.19) \end{aligned}$$

due to symmetry, and therefore (1.17) can be written using (1.18–1.19) as

$$\begin{aligned} & \mathbb{E}\left[\operatorname{sgn}(X)\operatorname{sgn}(Y)\right] \\ = \; & 2\,\mathbb{P}\left[X > 0, Y > 0\right] - 2\,\mathbb{P}\left[X > 0, Y < 0\right] \end{aligned} \tag{1.20}$$

Moreover,

$$\begin{aligned} & \mathbb{P}[X>0,Y>0]+\mathbb{P}[X<0,Y<0] \\ & +\mathbb{P}[X>0,Y<0]+\mathbb{P}[X<0,Y>0] \\ = \; & 1. \end{aligned} \qquad (1.21)$$

Using (1.18–1.19) in (1.21), we find that

$$2\,\mathbb{P}[X>0,Y>0]+2\,\mathbb{P}[X>0,Y<0] \;=\; 1$$

and therefore

$$\mathbb{P}[X>0,Y<0] \;=\; \frac{1}{2}-\mathbb{P}[X>0,Y>0]\,. \qquad (1.22)$$

By substituting (1.22) in (1.20), we obtain that

$$\mathbb{E}[\mathrm{sgn}(X)\mathrm{sgn}(Y)] \;=\; 4\,\mathbb{P}[X>0,Y>0]-1. \qquad (1.23)$$

To compute $\mathbb{P}[X>0,Y>0]$, recall that, if X and Y are standard normal variables with joint normal distribution with correlation ρ, then there exist two independent standard normal variables Z_1 and Z_2 such that

$$\begin{pmatrix} X \\ Y \end{pmatrix} \;=\; \begin{pmatrix} Z_1 \\ \rho Z_1+\sqrt{1-\rho^2}Z_2 \end{pmatrix}. \qquad (1.24)$$

Let $\widetilde{\rho}=\sqrt{1-\rho^2}$. From (1.24), we obtain that

$$Y \;=\; \rho X+\widetilde{\rho}Z, \qquad (1.25)$$

where we denoted Z_2 by Z for simplicity. Note that $X=Z_1$ and $Z=Z_2$ are independent standard normals.

Then, from (1.25) and using the fact that X and Z are independent standard normal variables, it follows that

$$\begin{aligned} & \mathbb{P}[X>0,Y>0] \\ = \; & \mathbb{P}[X>0,\rho X+\widetilde{\rho}Z>0] \\ = \; & \mathbb{P}\left[X>0,Z>-\frac{\rho}{\widetilde{\rho}}X\right] \\ = \; & \frac{1}{2\pi}\int_0^\infty\int_{-\frac{\rho}{\widetilde{\rho}}x}^\infty e^{-\frac{x^2+z^2}{2}}\;dzdx. \end{aligned} \qquad (1.26)$$

We use a polar coordinates change of variables to compute the integral (1.26). Let

$$x = r\cos(\theta); \quad z = r\sin(\theta),$$

and recall that

$$dzdx \;=\; rd\theta dr. \tag{1.27}$$

Note that

$$\begin{aligned} & -\frac{\rho}{\widetilde{\rho}}x \;<\; z \;<\; \infty \\ \iff & -\frac{\rho}{\widetilde{\rho}} \;<\; \tan(\theta) \;<\; \infty \\ \iff & \alpha \;<\; \theta \;<\; \frac{\pi}{2}, \end{aligned} \tag{1.28}$$

where

$$\alpha \;=\; \arctan\left(-\frac{\rho}{\widetilde{\rho}}\right).$$

Note that α is the signed angle between the x-axis and the straight line $\rho x + \widetilde{\rho}z = 0$ on the (x,z) plane.

From (1.26) and using (1.27) and (1.28), we obtain that

$$\begin{aligned} & \mathbb{P}[X>0, Y>0] \\ = & \frac{1}{2\pi}\int_0^\infty \int_{-\frac{\rho}{\widetilde{\rho}}x}^\infty e^{-\frac{x^2+z^2}{2}}\ dzdx \\ = & \frac{1}{2\pi}\int_0^\infty \int_\alpha^{\frac{\pi}{2}} e^{-\frac{r^2}{2}} r\ d\theta dr \\ = & \frac{1}{2\pi}\left(\frac{\pi}{2}-\alpha\right)\int_0^\infty re^{-\frac{r^2}{2}}\ dr \\ = & \left(\frac{1}{4}-\frac{\alpha}{2\pi}\right)\left(-e^{-\frac{r^2}{2}}\right)\Big|_0^\infty \\ = & \frac{1}{4}-\frac{\alpha}{2\pi}. \end{aligned} \tag{1.29}$$

From (1.23) and (1.29), we conclude that

$$\begin{aligned}
\mathbb{E}\left[\operatorname{sgn}(X)\operatorname{sgn}(Y)\right] &= 4\,\mathbb{P}\left[X>0, Y>0\right]-1 \\
&= 4\left(\frac{1}{4}-\frac{\alpha}{2\pi}\right)-1 \\
&= -\frac{2\alpha}{\pi}. \qquad (1.30)
\end{aligned}$$

Formulas (1.15) and (1.16) for $\mathbb{E}\left[\operatorname{sgn}(X)\operatorname{sgn}(Y)\right]$ corresponding to $\rho = 1$ and $\rho = -1$, respectively, can be obtained from the general formula (1.30) as limiting cases when ρ goes to 1 and to -1. For example,

$$\begin{aligned}
&\lim_{\rho\searrow -1} \mathbb{E}\left[\operatorname{sgn}(X)\operatorname{sgn}(Y)\right] \\
&= \lim_{\rho\searrow -1}\left(-\frac{2\alpha}{\pi}\right) \\
&= -\frac{2}{\pi}\lim_{\rho\searrow -1}\arctan\left(-\frac{\rho}{\widetilde{\rho}}\right) \\
&= -\frac{2}{\pi}\lim_{\rho\searrow -1}\arctan\left(-\frac{\rho}{\sqrt{1-\rho^2}}\right) \\
&= -\frac{2}{\pi}\cdot\frac{\pi}{2} \\
&= -1,
\end{aligned}$$

which is the same as (1.16), since

$$\lim_{\rho\searrow -1}\left(-\frac{\rho}{\sqrt{1-\rho^2}}\right) = \infty$$

and therefore

$$\lim_{\rho\searrow -1}\arctan\left(-\frac{\rho}{\sqrt{1-\rho^2}}\right) = \frac{\pi}{2}. \quad \square$$

Question 12. How do you create a long Gamma, short vega options trading strategy?

Answer: Both Gamma and vega are highest for options around at–the–money (ATM). However, the Gamma of ATM options is higher for options with shorter maturity (i.e., for short–dated options), while the vega of ATM options is higher for longer maturity options (i.e., for long–dated options); see Figure 1.1 and Figure 1.2, respectively.

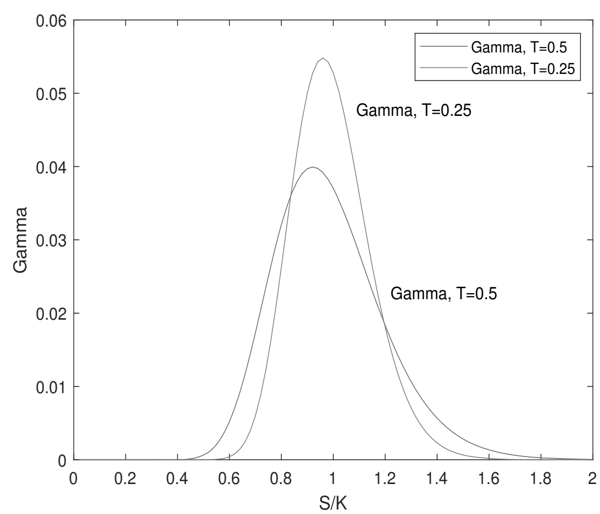

Figure 1.1: Dependence of Gamma on time to maturity

A trader who buys short–dated ATM options and sells the same number of long–dated ATM options will be long Gamma and short vega.

Note that calls and puts with the same strike have

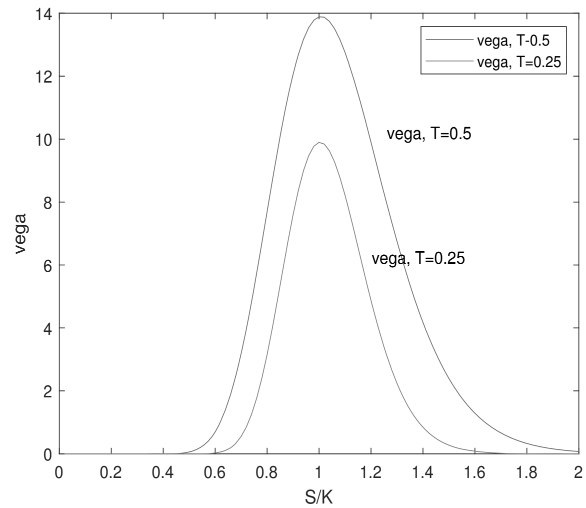

Figure 1.2: Dependence of vega on time to maturity

the same Gamma and vega, a consequence of the Put–Call parity, so you can take positions in either call or put options.

Also, the long Gamma, short vega portfolio can be made Delta–neutral by taking an appropriate position in the underlying asset. The delta of short–dated ATM options is smaller than the delta of long–dated ATM options. Then, the delta of the long Gamma, short vega portfolio is negative and therefore the trader will have to purchase units of the underlying asset in order to make the portfolio Delta–neutral. □

Question 13. Let X_t and Y_t be geometric Brownian motions driven by

$$\frac{dX_t}{X_t} = \mu_X dt + \sigma_X dW_t; \tag{1.31}$$

$$\frac{dY_t}{Y_t} = \mu_Y dt + \sigma_Y dB_t, \tag{1.32}$$

where W_t and B_t are correlated Brownian motions with constant correlation ρ. Show that

$$Z_t = \frac{X_t}{Y_t}$$

is also a geometric Brownian motion and determine its drift and volatility coefficients.

Answer: Let

$$f(x,y) = \frac{x}{y}.$$

By applying Itô's lemma to $Z_t = \frac{X_t}{Y_t}$, we obtain that

$$\begin{aligned} dZ_t &= d\left(\frac{X_t}{Y_t}\right) = d\,f(X_t, Y_t) \\ &= \frac{\partial f}{\partial x}(X_t, Y_t)\, dX_t + \frac{\partial f}{\partial y}(X_t, Y_t)\, dY_t \\ &\quad + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X_t, Y_t)\, d[X]_t + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(X_t, Y_t)\, d[Y]_t \\ &\quad + \frac{\partial^2 f}{\partial x \partial y}(X_t, Y_t)\, d[X,Y]_t. \end{aligned}$$

Note that

$$\frac{\partial f}{\partial x}(x,y) = \frac{1}{y}; \quad \frac{\partial f}{\partial y}(x,y) = -\frac{x}{y^2};$$

$$\frac{\partial^2 f}{\partial x^2}(x,y) = 0; \quad \frac{\partial^2 f}{\partial y^2}(x,y) = \frac{2x}{y^3}; \quad \frac{\partial^2 f}{\partial x \partial y}(x,y) = -\frac{1}{y^2}$$

and therefore

$$dZ_t = \frac{1}{Y_t}\, dX_t - \frac{X_t}{Y_t^2}\, dY_t + \frac{X_t}{Y_t^3}\, d[Y]_t - \frac{1}{Y_t^2}\, d[X,Y]_t. \tag{1.33}$$

Since

$$\begin{aligned} d[Y]_t &= \sigma_Y^2 Y_t^2 dt; \\ d[X,Y]_t &= \rho\sigma_X\sigma_Y X_t Y_t dt, \end{aligned}$$

it follows from (1.33) that

$$\begin{aligned} dZ_t &= \frac{1}{Y_t}\,dX_t - \frac{X_t}{Y_t^2}\,dY_t + \frac{X_t}{Y_t}\sigma_Y^2\,dt - \frac{X_t}{Y_t}\rho\sigma_X\sigma_Y\,dt \\ &= \frac{X_t}{Y_t}\,\frac{dX_t}{X_t} - \frac{X_t}{Y_t}\,\frac{dY_t}{Y_t} + \frac{X_t}{Y_t}\sigma_Y^2\,dt - \frac{X_t}{Y_t}\rho\sigma_X\sigma_Y\,dt \\ &= Z_t\,\frac{dX_t}{X_t} - Z_t\,\frac{dY_t}{Y_t} + Z_t\sigma_Y^2\,dt - Z_t\rho\sigma_X\sigma_Y\,dt, \end{aligned}$$

which can be written as

$$\frac{dZ_t}{Z_t} = \frac{dX_t}{X_t} - \frac{dY_t}{Y_t} + \left(\sigma_Y^2 - \rho\sigma_X\sigma_Y\right)\,dt. \tag{1.34}$$

By substituting (1.31) and (1.32) in (1.34), we obtain that

$$\begin{aligned} \frac{dZ_t}{Z_t} &= (\mu_X dt + \sigma_X dW_t) - (\mu_Y dt + \sigma_Y dB_t) \\ &\quad + \left(\sigma_Y^2 - \rho\sigma_X\sigma_Y\right)\,dt \\ &= \left(\mu_X - \mu_Y + \sigma_Y^2 - \rho\sigma_X\sigma_Y\right)\,dt \\ &\quad + (\sigma_X dW_t - \sigma_Y dB_t) \\ &= \mu_Z\,dt \; + \; (\sigma_X dW_t - \sigma_Y dB_t)\,, \end{aligned} \tag{1.35}$$

where

$$\mu_Z \;=\; \mu_X - \mu_Y + \sigma_Y^2 - \rho\sigma_X\sigma_Y.$$

Note that $\widetilde{W}_t$ given by

$$d\widetilde{W}_t \;=\; \frac{\sigma_X dW_t - \sigma_Y dB_t}{\sqrt{\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2}}$$

is a Brownian motion, and let

$$\sigma_Z \;=\; \sqrt{\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2}.$$

Then,

$$\sigma_X dW_t - \sigma_Y dB_t \;=\; \sigma_Z d\widetilde{W}_t, \qquad (1.36)$$

and we conclude from (1.35) and (1.36) that Z_t satisfies the SDE

$$\frac{dZ_t}{Z_t} \;=\; \mu_Z\, dt \;+\; \sigma_Z\, d\widetilde{W}_t.$$

Thus, Z_t is a geometric Brownian motion with drift μ_Z and volatility σ_Z, where

$$\begin{aligned} \mu_Z &= \mu_X - \mu_Y + \sigma_Y^2 - \rho\sigma_X\sigma_Y; \\ \sigma_Z &= \sqrt{\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2}. \quad \Box \end{aligned}$$

Question 14. Find the k–th largest element in an unsorted array. Assume that k is always valid, i.e., $k \geq 1$ and k is less than or equal to the length of the array.

Note: You are looking for the k–th largest element in the sorted order, not the k–th distinct element of the array.

Example 1:

```
Input: [3,2,1,5,6,4] and k = 2
Output: 5
```

Example 2:

```
Input: [3,2,3,1,2,4,5,5,6] and k = 4
Output: 4
```

Answer:
Solution 1: Use a max heap data structure as follows (sample code in C++):

```
class Solution {
public:
  int findKthLargest(vector<int>& nums, int k) {
    std::priority_queue<int> max_heap;
    for (int i = 0; i < nums.size(); ++i){
      max_heap.push(nums[i]);
```

```
    }
    int j = 0;
    while (j++ < k - 1){
      max_heap.pop();
    }
    return max_heap.top();
  }
};
```

Solution 2: Use a quick selection algorithm as follows (sample code in C++):

```
class Solution {
public:
  int findKthLargest(vector<int>& nums, int k) {
    const int size_n = nums.size();
    int left = 0, right = size_n;
    while (left < right) {
      int i = left, j = right - 1, pivot = nums[left];
      while(i <= j) {
        while (i <= j && nums[i] >= pivot) i++;
        while (i <= j && nums[j] < pivot) j--;
        if (i < j)
          swap(nums[i++], nums[j--]);
      }
      swap(nums[left], nums[j]);
      if (j == k - 1) return nums[j];
      if (j < k - 1) left = j + 1;
      else right = j;
    }
  }
};
```

Question 15. Given an array *nums*, there is a sliding window of size k which is moving from the very left of the array to the very right of the array. You can only see the k numbers in the window. Each time the sliding window moves right by one position. Assume that k is always valid, i.e., $k \geq 1$ and k is less than or equal to the size of the input array size for non-empty arrays.

Write an algorithm that returns the maximum of the sliding window.

Example:

```
Input: nums = [1,3,-1,-3,5,3,6,7], and k = 3
Output: [3,3,5,5,6,7]
```

Explanation:

```
Window Position                Max
--------------------------    -----
[1  3  -1] -3  5  3  6  7       3
 1 [3  -1  -3] 5  3  6  7       3
 1  3 [-1  -3  5] 3  6  7       5
 1  3  -1 [-3  5  3] 6  7       5
 1  3  -1  -3 [5  3  6] 7       6
 1  3  -1  -3  5 [3  6  7]      7
```

Answer: Use a deque (double-ended queue) data structure as follows (sample code in C++):

```
class Solution {
public:
vector<int> maxSlidingWindow(vector<int>& nums, int k) {
 int n = nums.size();
 vector<int> res;
 if (n == 0) return res;
 if (k == 1) return nums;
 deque<int> myDeque;
 for (int i = 0; i < n; ++i){
  if (myDeque.empty()) myDeque.push_back(i);
  else {
   if (i - myDeque.front() == k) myDeque.pop_front();
   if (i - myDeque.front() < k){
    if (nums[myDeque.back()] > nums[i]) myDeque.push_back(i);
    if (nums[myDeque.back()] < nums[i]){
     while (!myDeque.empty() && nums[i] > nums[myDeque.back()]){
      myDeque.pop_back();
     }
     myDeque.push_back(i);
    }
   }
  }
```

```
  if (i >= k-1) res.push_back(nums[myDeque.front()]);
 }
 return res;
}
};
```