

IEOR 4576 - DATA-DRIVEN METHODS IN FINANCE
COLUMBIA UNIVERSITY - IEO

Data-Driven Methods in Finance Homework 1

FALL 2024



Autor: Edward LUCYSZYN
Professor: Naftali COHEN

Questions

1. A fair die is rolled up to three times. After each roll, you can choose to stop or continue rolling based on the outcomes observed so far. The amount of money you win is equal to the value shown on the final roll. What is the optimal stopping strategy to maximize your expected winnings? Calculate the expected value of your winnings for this strategy.

Let's consider the case where we throw a third die. The expected value of the die, denoted by the random variable D_3 , is:

$$\mathbb{E}[D_3] = \sum_{i=1}^6 \frac{i}{6} = \frac{7}{2}.$$

Now let's consider the second roll. Let $h_2 : \llbracket 1, 6 \rrbracket \rightarrow \{0, 1\}$ be the decision function, whose value is 1 if we decide to roll the third die, or 0 otherwise. For each value of D_2 , we can either keep the value $i \in \llbracket 1, 6 \rrbracket$, or we can roll the other die whose expected value is $\mathbb{E}[D_3] = \frac{7}{2}$. By denoting S_2 as the random variable giving the value that a player will obtain by applying the strategy h_2 , we have:

$$S_2 = h_2(D_2)D_3 + (1 - h_2(D_2))D_2.$$

Hence, using the fact that D_2 and D_3 are independent,

$$\mathbb{E}[S_2] = \sum_{i=1}^6 h_2(i)\mathbb{E}[D_3] + (1 - h_2(i))i.$$

Since we only want to maximize the expected value, without considering the variance, it is clear that we want to keep the value of the second roll if it is above $\mathbb{E}[D_3] = \frac{7}{2}$. This means that we have:

$$\forall i \in \llbracket 1, 3 \rrbracket, h_2(i) = 1 \text{ and } \forall i \in \llbracket 4, 6 \rrbracket, h_2(i) = 0.$$

Thus,

$$\mathbb{E}[S_2] = \sum_{i=1}^3 \frac{7}{2} + \sum_{i=4}^6 \frac{i}{6} = \frac{17}{4}.$$

Let's now consider the first roll. We define D_1 as the value of the first die, S_1 as the value we will obtain at the end of the game with our strategy, and h_1 as the decision function. Using the same arguments as before, we have:

$$S_1 = h_1(D_1)S_2 + (1 - h_1(D_1))D_1.$$

So,

$$\mathbb{E}[S_1] = \sum_{i=1}^6 h_1(i)\mathbb{E}[S_2] + (1 - h_1(i))i.$$

We want to keep the value of the first roll if it is above $\mathbb{E}[S_2] = \frac{17}{4}$. This means that we have:

$$\forall i \in \llbracket 1, 4 \rrbracket, h_1(i) = 1 \text{ and } \forall i \in \llbracket 5, 6 \rrbracket, h_1(i) = 0.$$

Finally, we obtain:

$$\boxed{\mathbb{E}[S_1] = \sum_{i=1}^4 \frac{17}{4} + \sum_{i=5}^6 \frac{i}{6} = \frac{28}{6}.$$

2. An amoeba lives alone in a pond. After one minute, it will either die, divide into two amoebas, or remain unchanged, each with equal probability. Determine the expected number of amoebas in the pond and the variance after one minute.

Let's note X the random variable representing the number of amoebas after 1 minute. We have:

$$X = \begin{cases} 2 & \text{with probability } \frac{1}{3} \\ 1 & \text{with probability } \frac{1}{3} \\ 0 & \text{with probability } \frac{1}{3} \end{cases}.$$

Firstly,

$$\mathbb{E}[X] = 2 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = 1.$$

Now, let's compute the variance, by using:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

We have,

$$\mathbb{E}[X^2] = 2^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} = \frac{5}{3}.$$

Then,

$$\text{Var}(X) = \frac{5}{3} - 1^2 = \frac{2}{3}.$$

3. Calculate the mean and variance of a discrete uniform random variable defined on the set $1, 2, \dots, n$.

Let's note this random variable X :

$$\mathbb{E}[X] = \sum_{i=1}^n \frac{i}{n} = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Also, still by using $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$:

$$\mathbb{E}[X^2] = \sum_{i=1}^n \frac{i^2}{n} = \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}.$$

Thus,

$$\text{Var}(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = (n+1) \left[\frac{4n+2}{12} - \frac{3n+3}{12} \right] = \frac{n^2-1}{12}.$$

4. In a small town with 100 families, 30 families have 1 child, 50 families have 2 children, and 20 families have 3 children. The birth rank of a child is defined as 1 for the firstborn, 2 for the secondborn, and 3 for the thirdborn (if applicable).

a. A family is selected at random (with equal probability), and then a child from that family is chosen at random (with equal probability). Determine the probability mass function (PMF), mean, and variance of the child's birth rank.

Let X be the random variable representing the birth rank of the selected child. The number of children in each family affects the probabilities:

- If a family with 1 child is selected (probability $\frac{30}{100}$), the child's birth rank is always 1.
- If a family with 2 children is selected (probability $\frac{50}{100}$), the birth rank could be 1 or 2, with equal probability.
- If a family with 3 children is selected (probability $\frac{20}{100}$), the birth rank could be 1, 2, or 3, each with equal probability.

Thus, the probabilities are:

$$\begin{aligned}\mathbb{P}(X = 1) &= \frac{30}{100} \cdot 1 + \frac{50}{100} \cdot \frac{1}{2} + \frac{20}{100} \cdot \frac{1}{3} = \frac{185}{300}, \\ \mathbb{P}(X = 2) &= \frac{50}{100} \cdot \frac{1}{2} + \frac{20}{100} \cdot \frac{1}{3} = \frac{95}{300}, \\ \mathbb{P}(X = 3) &= \frac{20}{100} \cdot \frac{1}{3} = \frac{20}{300}.\end{aligned}$$

Firstly, the expectation:

$$\mathbb{E}[X] = 1 \cdot \frac{185}{300} + 2 \cdot \frac{95}{300} + 3 \cdot \frac{20}{300} = \frac{435}{300} = 1.45.$$

Next, the variance using $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. Hence:

$$\mathbb{E}[X^2] = 1^2 \cdot \frac{185}{300} + 2^2 \cdot \frac{95}{300} + 3^2 \cdot \frac{20}{300} = \frac{745}{300}.$$

Thus, the variance is:

$$\text{Var}(X) = \frac{745}{300} - \frac{435^2}{300^2} \approx 0.381.$$

b. A child is selected randomly from all the children in the town (with equal probability). Determine the PMF, mean, and variance of the child's birth rank.

In this part, each child is selected with equal probability. The total number of children in the town is:

$$30 \cdot 1 + 50 \cdot 2 + 20 \cdot 3 = 30 + 100 + 60 = 190.$$

Let Y be the random variable representing the birth rank of the selected child.

$$\begin{aligned}\mathbb{P}(Y = 1) &= \frac{30 + 50 + 20}{190} = \frac{100}{190} = \frac{10}{19}, \\ \mathbb{P}(Y = 2) &= \frac{50 + 20}{190} = \frac{70}{190} = \frac{7}{19}, \\ \mathbb{P}(Y = 3) &= \frac{20}{190} = \frac{2}{19}.\end{aligned}$$

Then,

$$\mathbb{E}[Y] = 1 \cdot \frac{10}{19} + 2 \cdot \frac{7}{19} + 3 \cdot \frac{2}{19} = \frac{30}{19} \approx 1.579.$$

Also, we have,

$$\mathbb{E}[Y^2] = 1^2 \cdot \frac{10}{19} + 2^2 \cdot \frac{7}{19} + 3^2 \cdot \frac{2}{19} = \frac{56}{19}$$

Thus, the variance is:

$$\text{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \frac{56}{19} - \left(\frac{30}{19}\right)^2 = \frac{56}{19} - \frac{900}{361} \approx 0.454.$$

5. Show that

$$\mathbb{E}((Y - \mathbb{E}(Y|X))^2|X) = \mathbb{E}(Y^2|X) - (\mathbb{E}(Y|X))^2$$

We have,

$$\begin{aligned} \mathbb{E}((Y - \mathbb{E}(Y|X))^2|X) &= \mathbb{E}(Y^2 - 2Y\mathbb{E}(Y|X) + (\mathbb{E}(Y|X))^2|X) \\ &= \mathbb{E}(Y^2|X) - 2\mathbb{E}(Y\mathbb{E}(Y|X)|X) + \mathbb{E}((\mathbb{E}(Y|X))^2|X). \end{aligned}$$

But, the term $\mathbb{E}((\mathbb{E}(Y|X))^2|X)$ is a function of X , so $\mathbb{E}((\mathbb{E}(Y|X))^2|X) = (\mathbb{E}(Y|X))^2$. Also, the term $\mathbb{E}(Y\mathbb{E}(Y|X)|X)$ simplifies to $\mathbb{E}(Y|X) \cdot \mathbb{E}(Y|X) = (\mathbb{E}(Y|X))^2$, since $\mathbb{E}(Y|X)$ is known given X . Thus, we get:

$$\begin{aligned} \mathbb{E}((Y - \mathbb{E}(Y|X))^2|X) &= \mathbb{E}(Y^2|X) - 2(\mathbb{E}(Y|X))^2 + (\mathbb{E}(Y|X))^2 \\ &= \mathbb{E}(Y^2|X) - (\mathbb{E}(Y|X))^2. \end{aligned}$$

7. Show that if $\mathbb{E}(Y|X) = c$ is a constant, then X and Y are uncorrelated.

Hint: Use Adam's law to find $\mathbb{E}(Y)$ and $\mathbb{E}(XY)$.

Using the law of total expectation (Adam's law):

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(c) = c.$$

Then, using the fact that X is $\sigma(X)$ -measurable (where $\sigma(X)$ represents the σ -algebra generated by X):

$$\mathbb{E}(XY) = \mathbb{E}(X\mathbb{E}(Y|X)) = \mathbb{E}(cX) = c\mathbb{E}(X).$$

Thus,

$$\boxed{\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = c\mathbb{E}(X) - c\mathbb{E}(X) = 0.}$$

8. Let X and Y be random variables with finite variances, and let $W = Y - \mathbb{E}(Y|X)$. This is a residual: the difference between the true value of Y and the predicted value of Y based on X .**a. Compute $\mathbb{E}(W)$ and $\mathbb{E}(W|X)$.**

First, using Adam's Law:

$$\boxed{\mathbb{E}[W] = \mathbb{E}[Y] - \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y] - \mathbb{E}[Y] = 0.}$$

Then, since $\mathbb{E}[Y|X]$ is $\sigma(X)$ -measurable,

$$\boxed{\mathbb{E}[W|X] = \mathbb{E}[Y - \mathbb{E}[Y|X]|X] = \mathbb{E}[Y|X] - \mathbb{E}[\mathbb{E}[Y|X]|X] = \mathbb{E}[Y|X] - \mathbb{E}[Y|X] = 0.}$$

b. Compute $\text{Var}(W)$, for the case that $W|X \sim N(0, X^2)$ with $X \sim N(0, 1)$.

We know that:

$$\text{Var}(W|X) = X^2.$$

Next, using the law of total variance, we have:

$$\text{Var}(W) = \mathbb{E}[\text{Var}(W|X)] + \text{Var}(\mathbb{E}[W|X]).$$

Since $\mathbb{E}[W|X] = 0$, the second term vanishes, so:

$$\text{Var}(W) = \mathbb{E}[\text{Var}(W|X)] = \mathbb{E}[X^2].$$

And then, with the fact that $X \sim N(0, 1)$:

$$\boxed{\text{Var}(W) = \mathbb{E}[X^2] = \text{Var}(X) + (\mathbb{E}[X])^2 = 1 + 0 = 1.}$$

10. Let X and Y be random variables. Is the statement

$$\max(X, Y) + \min(X, Y) = X + Y$$

correct? Also, is it correct to say

$$\text{Cov}(\max(X, Y), \min(X, Y)) = \text{Cov}(X, Y)$$

since either the maximum is X and the minimum is Y , or vice versa, and covariance is symmetric? Explain.

The first statement is correct because for any two real numbers X and Y , one of them will be the maximum and the other will be the minimum. Specifically, we have:

$$\max(X, Y) = \begin{cases} X & \text{if } X \geq Y \\ Y & \text{if } Y > X \end{cases},$$

and

$$\min(X, Y) = \begin{cases} Y & \text{if } X \geq Y \\ X & \text{if } Y > X \end{cases}.$$

In both cases, the sum of $\max(X, Y)$ and $\min(X, Y)$ is always $X + Y$.

For the second statement, this one is not generally true, here is a counter example. Let X and Y be independent random variables, where both X and Y follow a Bernoulli distribution with parameter $\frac{1}{2}$, i.e.,

$$X, Y \sim \mathcal{B}\left(\frac{1}{2}\right).$$

Since X and Y are independent, we have $\text{Cov}(X, Y) = 0$. Now, we compute the covariance of $\max(X, Y)$ and $\min(X, Y)$.

$$\max(X, Y) = \begin{cases} 0 & \text{if } (X, Y) = (0, 0), \\ 1 & \text{if } (X, Y) = (0, 1) \text{ or } (1, 0) \text{ or } (1, 1), \end{cases}$$

and

$$\min(X, Y) = \begin{cases} 0 & \text{if } (X, Y) = (0, 1) \text{ or } (1, 0) \text{ or } (0, 0), \\ 1 & \text{if } (X, Y) = (1, 1). \end{cases}$$

Then, we have:

$$\mathbb{E}[\max(X, Y)] = 0 \cdot 0.25 + 1 \cdot 0.75 = 0.75,$$

$$\mathbb{E}[\min(X, Y)] = 0 \cdot 0.75 + 1 \cdot 0.25 = 0.25.$$

For the product $\max(X, Y) \cdot \min(X, Y)$, it equals 1 only when $X = 1$ and $Y = 1$, so:

$$\mathbb{E}[\max(X, Y) \cdot \min(X, Y)] = 1 \cdot 0.25 = 0.25.$$

Thus,

$$\begin{aligned} \text{Cov}(\max(X, Y), \min(X, Y)) &= \mathbb{E}[\max(X, Y) \cdot \min(X, Y)] - \mathbb{E}[\max(X, Y)]\mathbb{E}[\min(X, Y)] \\ &= 0.25 - (0.75 \cdot 0.25) \\ &\neq 0. \end{aligned}$$

11. Two fair six-sided dice are rolled (one green and one orange), with outcomes X and Y corresponding to the green and orange dice.

a. Compute the covariance of $X + Y$ and $X - Y$.

By the linearity and symmetry of covariance, and the fact that X and Y are identically distributed:

$$\begin{aligned}\text{Cov}(X + Y, X - Y) &= \text{Cov}(X, X) - \text{Cov}(X, Y) + \text{Cov}(Y, X) - \text{Cov}(Y, Y) \\ &= \text{Var}(X) - \text{Cov}(X, Y) + \text{Cov}(X, Y) - \text{Var}(Y) \\ &= \text{Var}(X) - \text{Var}(Y) \\ &= 0.\end{aligned}$$

Thus, $\boxed{\text{Cov}(X + Y, X - Y) = 0.}$

b. Are $X + Y$ and $X - Y$ independent?

Covariance being zero does not necessarily imply independence. Suppose $X = 1$ and $Y = 6$. In this case: $X + Y = 7$ and $X - Y = -5$. The probability for this outcome is:

$$\mathbb{P}(X + Y = 7, X - Y = -5) = \mathbb{P}(X = 1, Y = 6) = \frac{1}{36}.$$

But,

$$\mathbb{P}(X + Y = 7) = \mathbb{P}((X, Y) = (1, 6) \text{ or } (2, 5) \text{ or } (3, 4) \text{ or } (4, 3) \text{ or } (5, 2) \text{ or } (6, 1)) = \frac{6}{36} = \frac{1}{6},$$

and

$$P(X - Y = -5) = P(X = 1 \text{ and } Y = 6) = \frac{1}{36}.$$

Since we have:

$$\mathbb{P}(X + Y = 7)\mathbb{P}(X - Y = -5) = \frac{1}{6} \cdot \frac{1}{36} = \frac{1}{216} \neq \frac{1}{36} = \mathbb{P}(X + Y = 7, X - Y = -5),$$

$\boxed{X + Y \text{ and } X - Y \text{ are not independent.}}$

12. Let X and Y be independent random variables. Show that:

$$\text{Var}(XY) = \text{Var}(X)\text{Var}(Y) + (\mathbb{E}(X))^2\text{Var}(Y) + (\mathbb{E}(Y))^2\text{Var}(X).$$

Firstly,

$$\text{Var}(XY) = \mathbb{E}((XY)^2) - (\mathbb{E}(XY))^2.$$

Since X and Y are independent, X^2 and Y^2 are independent. Thus,

$$\text{Var}(XY) = \mathbb{E}((X^2Y^2)) - (\mathbb{E}(XY))^2 = \mathbb{E}(X^2)\mathbb{E}(Y^2) - (\mathbb{E}(X)\mathbb{E}(Y))^2.$$

Recall that $\mathbb{E}(X^2) = \text{Var}(X) + (\mathbb{E}(X))^2$ and $\mathbb{E}(Y^2) = \text{Var}(Y) + (\mathbb{E}(Y))^2$. Therefore,

$$\begin{aligned}\text{Var}(XY) &= (\text{Var}(X) + (\mathbb{E}(X))^2)(\text{Var}(Y) + (\mathbb{E}(Y))^2) - (\mathbb{E}(X))^2(\mathbb{E}(Y))^2 \\ &= \text{Var}(X)\text{Var}(Y) + \text{Var}(X)(\mathbb{E}(Y))^2 + \text{Var}(Y)(\mathbb{E}(X))^2.\end{aligned}$$