1 Relationships Between Variables

1.1 Covariance

The covariance measures the joint variability of two random variables.

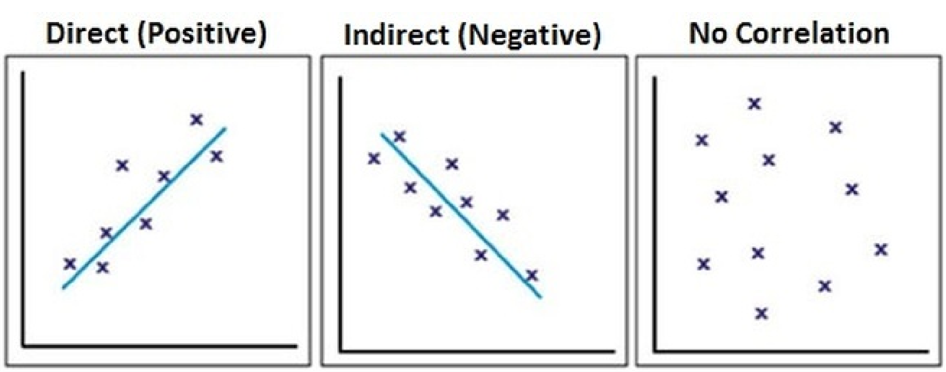
Definition 1. For two random variables X and Y, we define

$$\text{Cov}(X,Y) = \mathbb{E}\left[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right].$$

as the covariance between X and Y, provided the expression exists.

1.1.1 Properties

1. *Relationship with Variance:* $\text{Cov}(X,X) = \text{Var}(X)$.
2. *Equivalent formula:*
$$\text{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$
3. *Relationship with independence I:* If X and Y are independent,
$$\text{Cov}(X,Y) = 0.$$
The converse of this statement is false!. There are pairs of random variables that have zero covariance, but are dependent (see Problem 1.10).
4. *Relationship with Independence II:* If X and Y have zero covariance, then
$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y].$$
5. *Cauchy–Schwarz Inequality:* For any random variables X and Y,
$$|\,\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}.$$
6. *The Sign of the Covariance*: Suppose X,Y are positively related (when X large, Y likely large; when X small, Y likely small), then
$$\text{Cov}(X,Y) > 0$$
Conversely, suppose X,Y are negatively related (when X large, Y likely small; when X small, Y likely large), then
$$\text{Cov}(X,Y) < 0.$$



1.2 Correlation

The correlation measures how linearly related two random variables are.

Definition 2. The *correlation* of X and Y, denoted $\text{corr}(X, Y)$, is defined by

$$\rho = \text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{SD(X)SD(Y)}.$$

We say that X and Y are *uncorrelated* if $\text{Cov}(X, Y) = 0$ (or equivalently $\text{corr}(X, Y) = 0$). We have implicitly assumed that X and Y have non-zero variance in this definition

1.2.1 Properties

1. $\rho = \text{corr}(X, Y)$ has the same sign as $\text{Cov}(X, Y)$
2. $-1 \leq \rho \leq 1$
3. $|\rho| = 1 \iff X = aY + b$. If $a > 0$, then $\rho = 1$, and if $a < 0$, then $\rho = -1$.
4. X, Y independent $\Rightarrow \text{corr}(X, Y) = 0$
5. $\text{corr}(X, Y) = 0 \not\Rightarrow X, Y$ independent in general
6. $\text{corr}(X, X) = \text{Cov}(X, X)/SD(X)^2 = \text{Var}(X)/\text{Var}(X) = 1$
7. *Correlation does not imply causation:* Two variables being correlated does not always imply that one variable causes another to behave in certain ways.

1.3 Linear Combinations of Random Variables

Definition 3. A *linear combination* of the random variables $X_1, ..., X_n$ is any random variable of the form

$$\sum_{i=1}^{n} a_i X_i$$

where $a_1, ..., a_n \in \mathbb{R}$.

Example 1. Many common statistics are given by linear combinations of random variables.

1. *The Total:* Taking $a_i = 1$ for all i gives us the total of $X_1, \ldots, X_n$

$$T = \sum_{i=1}^{n} X_i.$$

2. *The Sample Mean:* Taking $a_i = \frac{1}{n}$ for all i gives us the sample mean of $X_1, \ldots, X_n$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

1.3.1 Properties

Let $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}$. We have the following properties about linear combinations of random variables.

1. **Linearity of Expectation:** For any random variables $X_1, \dots, X_n$,

$$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \, \mathbb{E}[X_i].$$

2. **Bi-Linearity of Covariance:** For any random variables $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$,

$$\mathrm{Cov}\left[\sum_{i=1}^{n} a_i X_i, \sum_{i=1}^{m} b_i Y_i\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathrm{Cov}(X_i, Y_j).$$

In particular, for random variables X, Y, U, V be random variables, and $a, b, c, d \in \mathbb{R}$. Then,

$$\begin{aligned}&\mathrm{Cov}(aX + bY, cU + dV)\\&= ac\mathrm{Cov}(X, U) + ad\mathrm{Cov}(X, V) + bc\mathrm{Cov}(Y, U) + bd\mathrm{Cov}(Y, V)\end{aligned}$$

3. **Variance of Linear Combinations:** The following result shows how the variance of a linear combination is "broken down" into pieces:

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \, \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} a_i a_j \mathrm{Cov}(X_i, X_j).$$

In particular, for random variables X, Y, and $a, b \in \mathbb{R}$. Then,

$$\mathrm{Var}(aX + bY) = a^2 \, \mathrm{Var}(X) + b^2 \, \mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y).$$

If the $X_1, \dots, X_n$ are independent, then they are uncorrelated, so in this case

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \, \mathrm{Var}(X_i)$$

4. **Stability of Normally Distributed Random Variables:** The linear combination of independent normally distributed random variables are normally distributed. Let $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, n$ be independent then,

$$\sum_{i=1}^{n}(a_i X_i + b_i) \sim \mathrm{N}\left(\sum_{i=1}^{n} a_i \mu_i + b_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

In particular, if $X \sim \mathrm{N}(\mu, \sigma^2)$ and $Y = aX + b$, where $a, b \in \mathbb{R}$, then we have the following standardization result

$$Y \sim \mathrm{N}(a\mu + b, a^2\sigma^2).$$

Remark 1. Notice that by linearity, we have

$$\mathbb{E}\left[\sum_{i=1}^{n}(a_i X_i + b_i)\right] = \sum_{i=1}^{n}(a_i \, \mathbb{E}[X_i] + b_i) = \sum_{i=1}^{n} a_i \mu_i + b_i$$

and by independence

$$\mathrm{Var}\left(\sum_{i=1}^{n}(a_i X_i + b_i)\right) = \mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \, \mathrm{Var}(X_i) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

which precisely matches the mean and variance of the linear combination.

1.4 Indicator Random Variables

A random variable taking values in natural numbers can often be expressed as a sum of indicator random variables. Linearity of expectation provides a powerful tool to compute expected values and variances of sums of indicator random variables.

Definition 4. Let $A \subset S$ be an event. We say that $\mathbb{1}_A$ is the *indicator* random variable of the event A. $\mathbb{1}_A$ is defined by:

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \in A^c \end{cases}$$

Remark 2. The random variable $\mathbb{1}_A(\omega)$ is a Bernoulli random variable where a success is the occurrence of the event A.

1.4.1 Properties

1. The products of indicator random variables is the indicator of the intersection of events

$$\mathbb{1}_A \mathbb{1}_B = \begin{cases} 1 & \omega \in A \cap B, \\ 0 & \omega \in (A \cap B)^c \end{cases}$$

2. $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$
3. $\mathrm{Var}(\mathbb{1}_A) = \mathbb{P}(A)(1 - \mathbb{P}(A))$
4. $\mathrm{Cov}(\mathbb{1}_A, \mathbb{1}_B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\,\mathbb{P}(B)$

1.5 Example Problems

1.5.1 Applications

Problem 1.1. Let

$$(X_1, X_2, X_3) \sim \mathrm{Mult}(10, 0.5, 0.3, 0.2).$$

Compute $\mathrm{Cov}(X_1, X_2)$.

Solution 1.1. From the properties of the multinomial distribution (see Week 10), we know that if $(X_1, \dots, X_k) \sim \mathrm{Mult}(n, p_1, \dots, p_k)$ then

$$\mathbb{E}[X_i X_j] = n(n-1)p_i p_j, \qquad \mathbb{E}[X_i] = np_i.$$

Applied to this problem using the equivalent formula for the covariance,

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2] = n(n-1)p_1 p_2 - np_1 np_2 = -np_1 p_2 = -10 \cdot 0.5 \cdot 0.3 = -1.5.$$

Problem 1.2. In a manufacturing process, two pieces of metal are combined to form a new piece of metal. Due to variations in the production process, we assume that the lengths of the two pieces, say L_1 and L_2, follow continuous uniform distributions as $L_1 \sim U(0.9, 1.1)$ and $L_2 \sim U(1.5, 1.7)$. Furthermore, due to variations joining process of the two pieces, the length of the new piece is not exactly $L_1 + L_2$, but instead $L = L_1 + L_2 + \varepsilon$ where $\varepsilon \sim N(0, 0.1^2)$. Compute the expected total length, $E(L)$.

Solution 1.2. From the formula sheet, we see $\mathbb{E}(L_1) = \frac{0.9+1.1}{2} = 1$, $\mathbb{E}(L_2) = \frac{1.5+1.7}{2} = 1.6$ and $\mathbb{E}(\varepsilon) = 0$. By linearity,

$$\mathbb{E}(L) = \mathbb{E}(L_1 + L_2 + \varepsilon) = \mathbb{E}(L_1) + \mathbb{E}(L_2) + \mathbb{E}(L_3) = 1 + 1.6 + 0 = 2.6.$$

Problem 1.3. Let X, Y be independent random variables with $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$. What is $\mathrm{Var}(X - Y)$?

Solution 1.3. By independence, $\mathrm{Cov}(X, Y) = 0$, so the variance for linear combinations formula implies

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + (-1)^2\,\mathrm{Var}(Y) - 2\mathrm{Cov}(X, Y) = 1 + 1 = 2.$$

Problem 1.4. In a certain cooking process, the target temperature, say C, follows a normal distribution (in celsius) with mean 57 and standard deviation 2. Your American friend asks you: What is the distribution of the target temperature in Fahrenheit?

Aside: The relationship between the temperature in Celsius c and Fahrenheit f is $f = c \cdot 9/5 + 32$.

Solution 1.4. By the stability property, $C \cdot 9/5 + 32 \sim \mathrm{N}(57 \cdot 9/5 + 32, (9/5)^2 \cdot 2^2) = \mathrm{N}(134.6, 12.96)$.

Problem 1.5. Let $X \sim \mathrm{N}(\mu_1, \sigma^2)$ be independent of $Y \sim \mathrm{N}(\mu_2, \sigma^2)$. What is the distribution of $X - Y$?

Solution 1.5. By the stability property, we have $\mathbb{E}[X - Y] = \mu_1 - \mu_2$ and $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 2\sigma^2$, so

$$X - Y \sim \mathrm{N}(\mu_1 - \mu_2, 2\sigma^2).$$

Problem 1.6. Three cylindrical parts are joined end to end to make up a shaft in a machine: 2 type-A parts and 1 type-B part. The lengths of the parts vary a little, and have the following distributions:

$$A \sim N(6, 0.4), \quad B \sim N(35.2, 0.6).$$

The overall length of the assembled shaft must lie between 46.8 and 47.5 or else the shaft has to be scrapped. Assume the lengths of different parts are independent. What percentage of assembled shafts has to be scrapped?

Solution 1.6. Let A_1, A_2 and B denote the three independent parts. The total length is $L = A_1 + A_2 + B$ satisfies

$$L \sim N(6 + 6 + 35.2, 0.4 + 0.4 + 0.6) \quad \Rightarrow \quad L \sim N(47.2, 1.4)$$

The part is scrapped if $L < 46.8$ or $L > 47.5$, so

$$\begin{aligned}
\mathbb{P}(\text{“scrapped”}) &= \mathbb{P}(L < 46.8) + \mathbb{P}(L > 47.5) \\
&= \mathbb{P}\left(Z < \frac{46.8 - 47.2}{\sqrt{1.4}}\right) + \mathbb{P}\left(Z > \frac{47.5 - 47.2}{\sqrt{1.4}}\right) \\
&= \Phi(-0.37) + (1 - \Phi(0.27)) \\
&= (1 - \Phi(0.37)) + (1 - \Phi(0.27)) \\
&= 0.749.
\end{aligned}$$

Remark 3. A common mistake is to say that $A_1 + A_2 + B$ is the same as $L = 2A_1 + B$ (A_1 and A_2 have the same distribution, after all), and conclude

$$L = 2A_1 + B \sim N(2 \cdot 6 + 35.2, 2^2 \cdot 0.4 + 0.6) \Rightarrow L \sim N(47.2, 2.2).$$

The linearity of expectation (which holds even if the random variables are dependent) is not affected by this mistake; but the variance is affected by this mistake. This is because $A_1 + A_2$ and $2A_1$ are very different objects since the first is a sum of two independent random variables and the latter is the sum of two very dependent random variables.

Problem 1.7. Suppose that the height of adult males in Canada is normally distributed with a mean of 70 inches and variance of 4^2 inches, and let $X_1, ..., X_{10}$ denote the heights of a random sample of adult males. Suppose $\bar{X}_{10}$ denotes the sample mean of these heights.

Let

$$p_1 = \mathbb{P}(68 \leq X_1 \leq 72)$$

and

$$p_{10} = \mathbb{P}(68 \leq \bar{X}_{10} \leq 72).$$

Which of the following is true?

1. $p_1 > p_{10}$
2. $p_1 = p_{10}$
3. $p_1 < p_{10}$

Solution 1.7. The interval contains the mean, so this result should be intuitive because a larger sample means less variance, so p_{10} should be bigger. To reinforce this, we can compute this explicitly.

We find

$$\begin{aligned} p_1 &= \mathbb{P}(68 \leq X_1 \leq 72) \\ &= \mathbb{P}\left(\frac{68-70}{4} \leq Z \leq \frac{72-70}{4}\right), \quad Z \sim N(0,1) \\ &= \Phi(0.5) - \Phi(-0.5) = 2\Phi(0.5) - 1 \\ &= 2 \cdot 0.69146 - 1 = 0.38292 \end{aligned}$$

Next,

$$\bar{X}_{10} = \frac{1}{10}\sum_{i=1}^{10} X_i \sim N\left(\frac{1}{10}\sum_{i=1}^{10} 70, \frac{1}{10^2}\sum_{i=1}^{10} 4^2\right) \Rightarrow \bar{X}_{10} \sim N(70, 1.6)$$

so

$$\begin{aligned} p_{10} &= \mathbb{P}(68 \leq \bar{X}_{10} \leq 72) \\ &= \mathbb{P}\left(\frac{68-70}{\sqrt{1.6}} \leq Z \leq \frac{72-70}{\sqrt{1.6}}\right), \quad Z \sim \mathrm{N}(0,1) \\ &= \Phi(1.58) - \Phi(-1.58) = 2\Phi(1.58) - 1 \\ &= 2 \cdot 0.94295 - 1 = 0.8859 \end{aligned}$$

Problem 1.8. Let A be an event and $p = \mathbb{P}(A)$ the probability of A. At which value of p is $\mathrm{Var}(\mathbb{1}_A)$ maximized?

Solution 1.8. Using the variance of a Bernoulli random variable, we have

$$\mathrm{Var}(\mathbb{1}_A) = f(p) = p(1-p) = p - p^2$$

This is a downward facing parabola, so we can find the critical point

$$f'(p) = 1 - 2p = 0 \implies p = \frac{1}{2}$$

which maximizes the variance, $f(1/2) = 1/4$.

Remark 4. Intuitively, the variance is maximized when there's no tendency for either heads (1) or tails (0), so when $p = 1/2$. If $p = 3/4$, the random variable is less variable, as there's a tendency for success, for instance.

Problem 1.9. N passengers board a plane with N seats, where $N > 1$. Despite every passenger having an assigned seat, when they board the plane they sit in one of the remaining available seats at random. Show that the mean and variance of the number of people sitting in the correct seat once everyone is on board are both 1 (independent of the number N of passengers, weirdly enough).

Solution 1.9. This is a classical problem called the matching problem. Let X denote the number of people sitting in the correct of seat once everyone is on board, and let A_i be the event that the ith passenger is in the correct seat. We have

$$\mathbb{1}_{A_i} = \begin{cases} 1 & \text{the } i\text{th passenger is in the correct seat} \\ 0 & \text{the } i\text{th passenger is not in the correct seat} \end{cases}.$$

Clearly, $X = \sum_{i=1}^n \mathbb{1}_{A_i}$. We can now compute the mean and variance.

Expected Value: By linearity of expectation

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[\mathbb{1}_{A_i}] = \sum_{i=1}^n \mathbb{P}(A_i).$$

By symmetry, we have that the probability that the ith passenger is in the correct seat is

$$\mathbb{P}(A_i) = \frac{1}{n}$$

since the seat the ith passenger sits in is uniform over the n possible seats. Therefore,

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[\mathbb{1}_{A_i}] = \sum_{i=1}^n \mathbb{P}(A_i) = \sum_{i=1}^n \frac{1}{n} = 1.$$

Variance: By the linearity of expectation

$$\mathbb{E}[X^2] = \mathbb{E}\left[\left(\sum_{i=1}^n \mathbb{1}_{A_i}\right)^2\right] = \sum_{i,j=1}^n \mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_j}].$$

We have two cases

1. $i = j$: Suppose that $i = j$. Since $\mathbb{1}(A_i)\mathbb{1}(A_i) = 1$ if an only if A_i happens, so we have

   $$\mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_i}] = \mathbb{E}[\mathbb{1}_{A_i}] = \mathbb{P}(A_i) = \frac{1}{n}$$

   as we computed before.

2. $i \neq j$: Suppose that $i \neq j$. Since $\mathbb{1}_{A_i}\mathbb{1}_{A_j} = 1$ if an only if A_i and A_j happens

$$\mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_j}] = \mathbb{P}(A_i \cap A_j) = \frac{1}{n(n-1)}.$$

Note that the events A_i and A_j are not independent, so we can't simply multiply the probabilities. Instead, we can use the fact that sets the i and j passengers sit in are uniform over the $n(n-1)$ possible seats for two passengers.

Since there are $n(n-1)$ ways to pick indices $i \neq j$ and n ways to pick indices $i = j$, we have

$$\mathbb{E}[X^2] = \sum_{i,j=1}^{n} \mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_j}] = \sum_{i=j} \mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_i}] + \sum_{i\neq j} \mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_j}] = \frac{n}{n} + \frac{n(n-1)}{n(n-1)} = 2.$$

Therefore,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 2 - 1 = 1.$$

Remark 5. Instead of using the uniform distribution and symmetry, we could argue that

$$\mathbb{P}(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

since there are $(n-1)!$ seating patterns where the ith passenger is in the right seat and $n!$ total seating patterns (all of which are equally likely).

Likewise, we have

$$\mathbb{P}(A_i \cap A_j) = \frac{1}{n(n-1)} = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}.$$

since there are $(n-2)!$ seating patterns where the ith and jth passenger is in the right seat and $n!$ total seating patterns (all of which are equally likely).

Yet another way to compute the probability is to argue sequentially using the chain rule,

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i \mid A_j)\,\mathbb{P}(A_j) = \frac{1}{n-1} \cdot \frac{1}{n} = \frac{1}{n(n-1)},$$

since the probability the jth passenger sits in the right seat is $\frac{1}{n}$ and the probability the ith passenger sits in the right seat given that the jth passenger is in the right seat is $\frac{1}{n-1}$ since the jth passenger is already in the correct seat so there are $n-1$ seats left.

Remark 6. We could have also used the formula for the variance of a linear combination, but the notation is slightly more cumbersome. Using expected values is simpler because computing probabilities are easier than computing variances of Bernoulli random variables.

1.5.2 Proofs and Derivations

Problem 1.10. Suppose that X and Y are independent. Show that

$$\mathrm{Cov}(X, Y) = 0.$$

Show that the converse is false by providing a counterexample.

Solution 1.10. Suppose that X and Y are independent. We know that $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$. Therefore, using the equivalent formula,

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0.$$

Counterexample: Let $X \sim U(-1,1)$, and let $Y = X^2$. X and Y are not independent because

$$0 = \mathbb{P}\left(X > \frac{1}{2}, Y < \frac{1}{4}\right) \neq \mathbb{P}\left(X > \frac{1}{2}\right)\mathbb{P}\left(Y < \frac{1}{4}\right) > 0$$

since $X > \frac{1}{2} \implies X^2 > \frac{1}{4}$ so it is impossible that $Y = X^2 < \frac{1}{4}$ as well. However, we can compute the covariance,

$$\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \mathbb{E}[X^3] - \mathbb{E}[X]\,\mathbb{E}[X^2] = 0$$

since the PDF of X is symmetric, so $\mathbb{E}[X^3] = 0$ and $\mathbb{E}[X] = 0$.

Problem 1.11. Prove the Cauchy–Schwarz inequality,

$$|\,\mathbb{E}[XY]| \le \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}.$$

Equality holds if and only if $Y = aX$ for some constant a.

Solution 1.11. Notice that the statement is trivial if either $X = 0$ or $Y = 0$, so we consider the non-trivial cases.

For any $t \in \mathbb{R}$, we have

$$0 \le \mathbb{E}[(tX - Y)^2] = at^2 - 2bt + c$$

where $a = \mathbb{E}[X^2]$, $b = \mathbb{E}[XY]$ and $c = \mathbb{E}[Y^2]$. A quadratic polynomial $at^2 - 2bt + c$ is non-negative if and only if it has at most one root, which happens if the discriminant satisfies

$$D = 4b^2 - 4ac \le 0 \implies b^2 \le ac \implies |b| \le \sqrt{ac}$$

so $|\,\mathbb{E}[XY]| \le \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}$. This proves the first part of the statement.

We now consider the equality case. Suppose now that we have equality $|\,\mathbb{E}[XY]| = \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$, so $|b| = \sqrt{ac}$. This implies that $D = 0$, so the quadratic polynomial has exactly one real root. Let $\lambda = \frac{b}{a}$ denote the value of this root, so

$$\mathbb{E}[(\lambda X - Y)^2] = a\lambda^2 - 2b\lambda + c = 0.$$

We have that $\mathbb{E}[(\lambda X - Y)^2] = 0$ if and only if $\lambda X - Y = 0$ with probability one, so $Y = \lambda X = \frac{\mathbb{E}[XY]}{\mathbb{E}[X^2]} X$ with probability 1. Therefore, if $X \neq aY$ for any a, then $Y \neq \frac{\mathbb{E}[XY]}{\mathbb{E}[X^2]} X$ so $|\,\mathbb{E}[XY]| \neq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$.

To prove the converse, suppose that $X = aY$. We have

$$|\,\mathbb{E}[XY]| = |a||\,\mathbb{E}[Y^2]| = \sqrt{\mathbb{E}((aY)^2)}\sqrt{\mathbb{E}(Y^2)} = |a|\,\mathbb{E}[Y^2]$$

so equality holds.

Problem 1.12. Show that $\rho = \operatorname{corr}(X,Y)$ satisfies $|\rho| \le 1$ and $|\rho| = 1$ if and only $Y = aX + b$ for some constants a and b.

Solution 1.12. By the Cauchy–Schwarz inequality, applied to $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$ we have

$$|\operatorname{Cov}(X,Y)| = |\,\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]| \le \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}\sqrt{\mathbb{E}[(Y - \mathbb{E}[Y])^2]} = \sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}.$$

Rearranging terms implies that

$$|\operatorname{corr}(X,Y)| = \frac{|\operatorname{Cov}(X,Y)|}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}} \le 1.$$

Next, we have that equality happens if and only if $Y - \mathbb{E}[Y] = a(X - \mathbb{E}[X])$ for some constant a. This means that there must be a linear relation between Y and X if equality were to hold. To see that any linear relation achieves equality, suppose that $Y = aX + b$ for some constants a and b, so by bilinearity

$$|\mathrm{Cov}(X,Y)| = |\mathrm{Cov}(X, aX+b)| = |a\mathrm{Cov}(X,X) + b\mathrm{Cov}(X,1))| = |a|\,\mathrm{Var}(X)$$

and

$$\sqrt{\mathrm{Var}(Y)} = \sqrt{\mathrm{Var}(aX+b)} = |a|\sqrt{\mathrm{Var}(X)},$$

so

$$|\mathrm{corr}(X,Y)| = 1.$$

Remark 7. We can repeat the second computation without the absolute values to conclude that $\mathrm{corr}(X,Y) = 1$ implies that $Y = aX + b$ for some constant $a > 0$ and $\mathrm{corr}(X,Y) = -1$ implies that $Y = aX + b$ for some constant $a < 0$

Problem 1.13. Prove the binlinearity property of covariances

$$\mathrm{Cov}\left[\sum_{i=1}^n a_i X_i, \sum_{i=1}^n b_i Y_i\right] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathrm{Cov}(X_i, Y_j).$$

Solution 1.13. This is a direct consequence of linearity of expectation and the distributive property of numbers

$$\sum_{i=1}^n a_i \times \sum_{j=1}^m b_j = \sum_{i=1}^n \sum_{j=1}^m a_i b_j.$$

By the definition of the covariance,

$$\begin{aligned}
\mathrm{Cov}\left[\sum_{i=1}^n a_i X_i, \sum_{i=1}^n b_i Y_i\right] &= \mathbb{E}\left[\left(\sum_{i=1}^n a_i X_i - \mathbb{E}\left[\sum_{i=1}^n a_i X_i\right]\right)\left(\sum_{i=1}^m b_i Y_i - \mathbb{E}\left[\sum_{i=1}^m b_i Y_i\right]\right)\right] \\
\text{linearity of expectation} &= \mathbb{E}\left[\left(\sum_{i=1}^n a_i (X_i - \mathbb{E}[X_i])\right)\left(\sum_{i=1}^m b_i (Y_i - \mathbb{E}[Y_i])\right)\right] \\
\text{distributive property} &= \mathbb{E}\left[\sum_{i=1}^n \sum_{i=1}^m a_i b_i (X_i - \mathbb{E}[X_i])(Y_i - \mathbb{E}[Y_i])\right] \\
\text{linearity of expectation} &= \sum_{i=1}^n \sum_{i=1}^m a_i b_i\, \mathbb{E}[(X_i - \mathbb{E}[X_i])(Y_i - \mathbb{E}[Y_i])] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathrm{Cov}(X_i, Y_j).
\end{aligned}$$

Problem 1.14. Prove the formula for the variance of linear combinations of random variables,

$$\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2\, \mathrm{Var}(X_i) + 2\sum_{1 \le i < j \le n} a_i a_j \mathrm{Cov}(X_i, X_j).$$

Solution 1.14. Since $\text{Var}(X) = \text{Cov}(X, X)$, the proof follows directly from the bilinearity of covariance. We have

$$\begin{aligned}
\text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) &= \text{Cov}\left[\sum_{i=1}^{n} a_i X_i, \sum_{i=1}^{n} a_i X_i\right] \\
\text{bilinearity} \quad &= \sum_{i,j=1}^{n} a_i a_j \text{Cov}(X_i, X_j) \\
\text{split into diagonal and offdiagonal} \quad &= \sum_{i}^{n} a_i^2 \text{Cov}(X_i, X_i) + \sum_{i \neq j} a_i a_j \text{Cov}(X_i, X_j) \\
\text{Cov}(X,Y) = \text{Cov}(Y,X), \text{Var}(X) = \text{Cov}(X,X) \quad &= \sum_{i}^{n} a_i^2 \,\text{Var}(X_i) + 2 \sum_{i<j} a_i a_j \text{Cov}(X_i, X_j).
\end{aligned}$$

Problem 1.15. ($\star$) Let $X_i \sim \text{N}(\mu_i, \sigma_i^2),\ i = 1, 2, \dots, n$ be independent then,

$$\sum_{i=1}^{n} (a_i X_i + b_i) \sim \text{N}\left(\sum_{i=1}^{n} a_i \mu_i + b_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

Solution 1.15. The easiest proof of this fact uses moment generating functions, which will be explained a later week. In the meantime, we will present a geometric proof of this result using a multidimensional change of variables. For simplicity, we consider the case that $n = 2$ and the means of the random variables are 0 and $a_i = 1$ and $b_i = 1$. The general case can be recovered by an induction argument and a standardization argument.

Let $X_1 \sim \text{N}(0, \sigma_1)$ and $X_2 \sim \text{N}(0, \sigma_2)$ be independent. We need to show that

$$X_1 + X_2 \sim \text{N}(0, \sigma_1^2 + \sigma_2^2).$$

By standardization, $X_1 = \sigma_1 Z_1$ and $X_2 = \sigma_2 Z_2$. We want to find the PDF of

$$F_{X_1+X_2}(t) = \mathbb{P}(\sigma_1 Z_1 + \sigma_2 Z_2 \le t) = \frac{1}{2\pi} \iint_{\sigma_1 z_1 + \sigma_2 z_2 \le t} e^{-\frac{z_1^2 + z_2^2}{2}} dz_1 dz_2.$$

Using the change of variables corresponding to the rotation of the half plane to one perpendicular to the z_2 axis,

$$w_1 = \frac{\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} z_1 + \frac{\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} z_2 \qquad w_2 = \frac{\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} z_1 - \frac{\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} z_2$$

we have

$$\{\sigma_1 z_1 + \sigma_2 z_2 \le t\} = \left\{ w_1 \le \frac{t}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right\}$$

so

$$\frac{1}{2\pi} \iint_{\sigma_1 z_1 + \sigma_2 z_2 \le t} e^{-\frac{z_1^2 + z_2^2}{2}} dz_1 dz_2 = \frac{1}{2\pi} \iint_{w_1 \le \frac{t}{\sqrt{\sigma_1^2 + \sigma_2^2}}} e^{-\frac{w_1^2 + w_2^2}{2}} dw_1 dw_2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{t}{\sqrt{\sigma_1 + \sigma_2}}} e^{-\frac{w_1^2}{2}} dw_1.$$

We conclude that

$$F_{X_1+X_2}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{t}{\sqrt{\sigma_1 + \sigma_2}}} e^{-\frac{w_1^2}{2}} dw_1 = \mathbb{P}\left(Z \le \frac{t}{\sqrt{\sigma_1 + \sigma_2}}\right) = \mathbb{P}(\sqrt{\sigma_1 + \sigma_2} Z \le t)$$

which is the CDF of a $\text{N}(0, \sigma_1^2 + \sigma_2^2)$ random variable.

Remark 8. Essentially we have exploited the rotational invariance of the standard normal. Notice that the PDF is rotational symmetric since

$$f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{2\pi}e^{-\frac{z_1^2+z_2^2}{2}} = \frac{1}{2\pi}e^{-\frac{r^2}{2}} = f_{Z_1,Z_2}(r)$$

is a function of $r = \sqrt{z_1^2+z_2^2}$, which means that the density only depends on the distance from points to the origin. The special change of variables rotated the half plane $\{\sigma_1 z_1 + \sigma_2 z_2 \leq t\}$ to be perpendicular to the z_2 axis to reduce the problem to computing the probability $\sqrt{\sigma_1^2+\sigma_2^2}Z \leq t$.

Problem 1.16. Let $X_1,\dots,X_n$ be independent and $X_i \sim N(\mu,\sigma^2)$ for all $i = 1,\dots,n$. Show that

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Solution 1.16. By the Gaussian stability, we have

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$$

is normally distributed. We just have to compute the mean and variance. By linearity,

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{n\mu}{n} = \mu.$$

and the variance of a linear combination (the covariance is 0 by independence) gives us

$$\operatorname{Var}(\bar{X}_n) = \sum_{i=1}^n \frac{1}{n^2}\operatorname{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Remark 9. As n increases, the variance σ^2/n decreases, so the distribution of $\bar{X}_n$ becomes more concentrated around μ. This is intuitive because if think of estimating the average of the midterm (or any other event that is normally distributed), then asking 5 people gives us a less reliable result than asking 50 people.

Problem 1.17. Show that

1. $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$
2. $\operatorname{Var}(\mathbb{1}_A) = \mathbb{P}(A)(1-\mathbb{P}(A))$
3. $\operatorname{Cov}(\mathbb{1}_A, \mathbb{1}_B) = \mathbb{P}(A\cap B) - \mathbb{P}(A)\,\mathbb{P}(B)$

Solution 1.17. The proof is somewhat straightforward, and it relies on the observation that

$$\mathbb{1}_A\mathbb{1}_B = \begin{cases} 1 & \omega \in A\cap B, \\ 0 & \omega \in (A\cap B)^c \end{cases}$$

We can now compute the required objects

1. $$\mathbb{E}(\mathbb{1}_A) = 1 \cdot \mathbb{P}(\mathbb{1}_A = 1) + 0 \cdot P(\mathbb{1}_A = 0) = \mathbb{P}(A)$$

2. We have $\mathbb{1}_A^2 = 1$ if and only if $\omega \in A$, so
$$\mathbb{E}(\mathbb{1}_A^2) = 1 \cdot \mathbb{P}(\mathbb{1}_A^2 = 1) + 0 \cdot \mathbb{P}(\mathbb{1}_A^2 = 0) = 1 \cdot \mathbb{P}(A) + 0 \cdot \mathbb{P}(A^c) = \mathbb{P}(A)$$
so
$$\mathrm{Var}(\mathbb{1}_A) = \mathbb{E}(\mathbb{1}_A^2) - \mathbb{E}(\mathbb{1}_A)^2 = \mathbb{P}(A) - \mathbb{P}(A)^2 = \mathbb{P}(A)(1 - \mathbb{P}(A))$$

3. Similarly, we have $\mathbb{1}_A \mathbb{1}_B = 1$ if and only if $\omega \in A \cap B$, so
$$\mathbb{E}(\mathbb{1}_A \cdot \mathbb{1}_B) = 1 \cdot \mathbb{P}(\mathbb{1}_A \mathbb{1}_B = 1) + 0 \cdot \mathbb{P}(\mathbb{1}_A \mathbb{1}_B = 0) = 1 \cdot \mathbb{P}(A \cap B) + 0 \cdot \mathbb{P}((A \cap B)^c) = \mathbb{P}(A \cap B)$$
giving us
$$\mathrm{Cov}(\mathbb{1}_A, \mathbb{1}_B) = \mathbb{E}(\mathbb{1}_A \cdot \mathbb{1}_B) - \mathbb{E}(\mathbb{1}_A)\,\mathbb{E}(\mathbb{1}_B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\,\mathbb{P}(B).$$