7. Properties of Expectation

Expectation of a Sum

- Proposition:

$$\mathbb{E}\left[g(X,Y)\right] = \sum_x \sum_y g(x,y)p(x,y)$$

$$\mathbb{E}\left[g(X,Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy$$

- Using the proposition, we have

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

Similarly,

$$\mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)$$

Example: the binomial distribution.

Let $X \sim \text{Binomial}(n, p)$. Find $\mathbb{E}(X)$.

Solution. (A simpler method than direct calculation)

Example. Ten hunters are waiting for ducks to fly by. When a flock of ducks flies overhead, the hunters fire at the same time, but each chooses his target at random, independently of the others. If each hunter independently hits his target with probability p, compute the expected number of ducks that escape unhurt when a flock of size 10 flies overhead.

Solution.

Example. The coupon collecting problem.

There are N different baseball cards, and each cereal packet contains one card, which is equally like to be any type. Find the expected number of cereal packets required to collect all N baseball cards.

Solution.

Covariance

Definition: The **covariance** between X and Y is

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

Proposition 1.

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

Proof

Proposition 2. If X and Y are **independent**,

$$\mathbb{E}\left[g(X)h(Y)\right] = \mathbb{E}\left[g(X)\right]\mathbb{E}\left[h(Y)\right]$$

Proof.

Examples:

- If X and Y are **independent**,

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

- If X and Y are **independent**, $\mathrm{Cov}(X, Y) = 0$.

More Properties of Covariance

- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- Symmetry: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- Linearity in each variable:

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

$$\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$$

Exercise: derive this.

- Bilinearity as a function of both variables:

$$\mathrm{Cov}\Big(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j\Big) = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{Cov}(X_i, Y_j)$$

The Variance of a Sum of RVs

- For any random variables $X_1, \ldots, X_n$

$$\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

Proof.

- A consequence: if $X_1, \ldots, X_n$ are independent,

$$\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Example. Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables having expected value μ and variance σ^2. Let $\overline{X} = \sum_{i=1}^{n} X_i/n$ be the sample mean. Find

(a) $\mathrm{Var}(\overline{X})$

(b) $\mathrm{Cov}(X_i - \overline{X}, \overline{X})$

Example: variance of a binomial RV

Let $X \sim \text{Binomial}(n, p)$. Find $\text{Var}(X)$.

Solution. (A simpler method than direct calculation)

Correlation

Definition. The **correlation** of X and Y is

$$\mathrm{Corr}(X,Y) = \mathrm{Cov}(X,Y)/\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$$

- Correlation measures (linear) association of RVs on a scale of -1 and 1.
- Example: $\mathrm{Corr}(X, aX+b) = 1$ if $a > 0$, and $\mathrm{Corr}(X, aX+b) = -1$ if $a < 0$.

Definition. X and Y are **uncorrelated** if $\mathrm{Corr}(X,Y) = 0$, i.e. if $\mathrm{Cov}(X,Y) = 0$.

Independent RVs are uncorrelated, but the converse is not necessarily true.

Example: Let X and Y have joint p.m.f. $p(x, y)$ given by

	$y = 1$	$y = -1$	$y = 5$	$y = -5$
$x = 1$	$1/4$	$1/4$	0	0
$x = -1$	0	0	$1/4$	$1/4$

Show that X and Y are uncorrelated but not independent.

Solution.

Example. Covariance of indicator RVs.

For two events A and B, let

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{else} \end{cases}, \quad I_B = \begin{cases} 1 & \text{if } B \text{ occurs} \\ 0 & \text{else} \end{cases}$$

Show that $\mathrm{Cov}(I_A, I_B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$.

Solution

Note. I_A and I_B are independent if and only if they are uncorrelated.

Example. Let X be the number of 1's and Y the number of 2's, in n tosses of a die. Find $\mathrm{Cov}(X, Y)$ and $\mathrm{Corr}(X, Y)$.

Solution.

Example. A die is rolled twice. Let X be the value of the first roll, Y the value of the second, and $Z = X + Y$. Find the covariance of X and Z.

Solution. Method 1: Direct computation.

Solution. Method 2: Problems like this are often simpler using the identity
$\mathrm{Cov}(U, V + W) = \mathrm{Cov}(U, V) + \mathrm{Cov}(U, W)$.