1 Expectation

1.1 The general results

Proposition 1.1. If $X, Y$ have a joint probability mass function $p(x, y)$ then

$$E(g(X, Y)) = \sum_{x, y} g(x, y)P(X = x, Y = y).$$

If $X, Y$ have a joint density function $f(x, y)$ then

$$E(g(X, Y)) = \int_{-\infty}^\infty \int_{-\infty}^\infty g(x, y)f(x, y)dxdy.$$

Corollary 1.2. Let $X, Y$ be either two discrete or continuous RVs. Then

$$E(X + Y) = E(X) + E(Y).$$

By induction,

$$E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i).$$

1.2 Examples

Example 1.3. Accident location $X$ is uniformly distributed on a road of length $L$. At the time of the accident, the ambulence location is also independently uniformly distributed on the same road. What is the expected distance between the ambulence and the accident?
Ans:
\[ E(|X - Y|) = \int_0^L \int_0^L \frac{1}{L^2} |x - y| \, dx \, dy = \frac{L}{3}. \]

**Example 1.4. Sample mean**

Let \( X_1, \cdots, X_n \) be identically distributed. Let \( \overline{X} := \frac{1}{n} \sum_i X_i \). Compute \( E(\overline{X}) \).

Ans:
\[ E(\overline{X}) = \frac{1}{n} \sum_i E(X_i) = E(X_1). \]

**Example 1.5. Mean of a hypergeometric**

\( n \) balls are selected without replacement from an urn with \( N \) balls, \( m \) of which are white. Find the expectation of the number of white balls in the sample.

Ans: Let \( X_i, i = 1, \cdots, n \) be RVs such that
\[ X_i = \begin{cases} 1, & \text{if \ ith \ ball \ is \ white} \\ 0, & \text{otherwise.} \end{cases} \]

Then \( X = \sum_{i=1}^n X_i \) represents number of white balls in the sample. Thus
\[ E(X) = \sum_{i=1}^n E(X_i) = nP( \text{ith ball is white} ) = \frac{nm}{N}. \]

**Example 1.6. Hat selecting problem**

\( N \) people select their hats from a pile of \( N \) hats. Find the expected number of people selecting their own hat.

Ans: Let \( X_i, i = 1, \cdots, N \) be RVs such that
\[ X_i = \begin{cases} 1, & \text{if \ ith \ person \ selects \ his \ own \ hat} \\ 0, & \text{otherwise.} \end{cases} \]

Then \( X = \sum_{i=1}^N X_i \) represents number of people selecting their own hat. Thus
\[ E(X) = \sum_{i=1}^N E(X_i) = NP( \text{ith ball is white} ) = \frac{N}{N} = 1. \]
Example 1.7. Coupon selecting problem

Suppose there are $N$ types of coupons, each time one collects, any type is equally likely. Find the expected number of coupons one needs to collect before having a complete set of each type.

Ans: Let $X_i, i = 0, 1, N - 1$ be the number of additional coupons that need to be obtained after the $i$ distinct types have been collected in order to obtain another distinct type. Then the total number of coupons required is

$$X = \sum_{i=0}^{N-1} X_i.$$ 

Each $X_i$ has a Geometric($p_i$) distribution with

$$p_i = \frac{N - i}{N}.$$ 

Therefore $E(X_i) = \frac{1}{p_i} = \frac{N}{N - i}$. Thus $E(X) = \sum_{i=0}^{N-1} \frac{N}{N - i}$.

Example 1.8. Duck shooting problem

Ten hunters are shooting at the ducks, each choosing his target at random. Suppose each also independently hits his target with probability $p$, and there are 10 ducks. Find the expected number of ducks that escape unhurt.

Let $X_i, i = 1, \cdots, N$ be RVs such that

$$X_i = 1, \text{ if } \text{ith duck escapes unhurt}$$
$$= 0, \text{ otherwise.}$$

Then $X = \sum_{i=1}^{N} X_i$ represents number of ducks escaped unhurt.

Let us compute the probability of the $i$th duck being hurt. Let $E_{ij}$ be the event that the $i$th duck got hit by the $j$th hunter. Also let $F_{ij}$ be the event that the $j$th hunter locks on the $i$th duck. Then

$$P(E_{ij}) = P(E_{ij}|F_{ij})P(F_{ij}) = \frac{p}{10}.\]

(Note that this problem is unlike the hat-selection problem in that two hunters can hit the same duck. Thus if we let $E_i$ be the event that the $i$th duck got hurt, even though this is still true:

$$P(E_i) = \sum_{i=1}^{10} P(E_i|F_{ij})P(F_{ij}),$$

3
we do NOT have $P(E_i|F_{ij}) = p$ since even if the jth hunter locks on the ith duck, it can get hurt by other hunters as well. )

By assumption, $E_{i1}, E_{i2}, \cdots, E_{i10}$ are independent. Thus

$$P(E_{ij}^c) = (1 - \frac{p}{10})^{10}.$$ 

Thus $E(X) = 10(1 - \frac{p}{10})^{10}$.

**Example 1.9.** The inclusion exclusion principle

Let $E_1, E_2, \cdots, E_n$ be events and $X_i$ be random variables such that

$$X_i = 1, \text{ if } E_i \text{ occurs}$$

$$= 0, \text{ otherwise.}$$

Let $X = 1 - \prod_{i=1}^n (1 - X_i)$. Then

$$X = 1, \text{ if at least one of } E_i \text{ occurs}$$

$$= 0, \text{ non of } E_i \text{ occurs.}$$

Then $E(X) = P(\bigcup_{i=1}^n E_i)$. On the other hand, note that

$$1 - \prod_{i=1}^n (1 - X_i) = \sum_{i=1}^n X_i - \sum_{i<j} X_i X_j + \cdots + (-1)^{n+1} X_1 \cdots X_n.$$

From this we derive the inclusion exclusion formula.

## 2 Covariance and Variance

### 2.1 General results

**Definition 2.1.** The covariance between $X, Y$, denoted as $Cov(X, Y)$ is defined as

$$Cov(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - E(X)E(Y).$$

Remark, by the following proposition:

**Proposition 2.2.** If $X, Y$ are independent then $E(g(X)h(Y)) = E(g(X))E(h(Y))$,

we have that if $X, Y$ are independent then $Cov(X, Y) = 0$.

However, note that $Cov(X, Y) = 0$ does NOT imply that $X, Y$ are independent, as the following example shows.

**Example 2.3.** Let $X = -1, 0, 1$ with probability 1/3 each. Let $Y = 0$ if $X \neq 0$ and 1 if $X = 0$. Then $E(XY) = E(X) = 0$. Thus $Cov(X, Y) = 0$. But clearly $X, Y$ are NOT independent.
2.2 Some properties

Proposition 2.4. We have:
1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
2. $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$;
3. $\text{Cov}(\sum_i X_i, \sum_j Y_j) = \sum_{i,j} \text{Cov}(X_i, Y_j)$;
4. $\text{Cov}(X, X) = \text{Var}(X)$.
5. If $X, Y$ are independent, $\text{Cov}(X, Y) = 0$.

From properties 3 and 5, we have the following result about the variance of sum of independent RVs:

Proposition 2.5. If $X_1, X_2, \cdots, X_n$ are independent then

$$\text{Var}\left(\sum_i X_i\right) = \sum_i \text{Var}(X_i).$$

2.3 Examples

Example 2.6. Variance of sample mean and sample variance

Let $X_1, \cdots, X_n$ be independent and identically distributed. Let $\bar{X} := \frac{1}{n} \sum_i X_i$. Compute $\text{Var}(\bar{X})$. Also denote

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}.$$  

Compute $E(S^2)$.

Ans:

$$\text{Var}(\bar{X}) = \frac{1}{n^2} \sum_i \text{Var}(X_i) = \frac{\text{Var}(X)}{n}.$$  

$$E(S^2) = \frac{1}{n-1} E\left(\sum_{i=1}^n (X_i - \bar{X})^2\right)$$

$$= \frac{1}{n-1} E\left(\sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2)\right)$$

$$= \frac{1}{n-1} E\left(\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right)$$

$$= \frac{1}{n-1} E\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right)$$
Denoting $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$ we see that $E(X_i^2) = \mu^2 + \sigma^2$ and $E(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2$. Thus

$$E(S^2) = \frac{1}{n-1}[n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)] = \sigma^2 = \text{Var}(X_1).$$

We denote $\rho(X,Y) := \frac{\text{Cov}(X,Y)}{\text{Var}(X)\text{Var}(Y)}$ as the correlation between $X, Y$. It is a number between $[-1,1]$ where $-1$ denote perfect negative correlation, 0 non-correlation and 1 perfect positive correlation.

**Example 2.7. Correlation and conditional probability**

Let $A, B$ be 2 events and $I_A, I_B$ RVs such that

$$I_A(I_B) = 1 \text{ if } A (B) \text{ occurs}$$
$$= 0 \text{ otherwise.}$$

Then $E(I_A) = P(A), E(I_B) = P(B), E(I_AI_B) = P(AB)$. Thus

$$\text{Cov}(I_A, I_B) = P(AB) - P(A)P(B) = P(B)(P(A|B) - P(A)).$$

Thus we see that $A, B$ are positively correlated if event $B$ happenning makes it more likely for $A$ to happen, negatively correlated if $B$ happenning makes it less likely for $A$ to happen, and not correlated if $B$ happenning does not influence $A$.

**Example 2.8. Zero-correlation between sample deviation and sample mean**

Let $X_1, \cdots, X_n$ be independent and identically distributed. Show that

$$\text{Cov}(X_i - \overline{X}, \overline{X}) = 0.$$

Ans:

$$\text{Cov}(X_i - \overline{X}, \overline{X}) = \text{Cov}(X_i, \overline{X}) - \text{Var}(\overline{X})$$
$$= \frac{1}{n} \sum_j \text{Cov}(X_i, X_j) - \frac{\sigma^2}{n}$$
$$= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0.$$

Remark: It follows from this computation that

$$\text{Cov}(\frac{\sum_i X_i - \overline{X}}{n-1}, \overline{X}) = 0.$$
However, we CANNOT conclude that $\text{Cov}(S^2, \bar{X}) = 0$ from the above, because

$$\sum_{i}(X_i - \bar{X})^2 \over n - 1,$$

and we cannot “pass the square inside the sum.”

There is a special case worth noticing. That is, when $X, Y$ have joint Normal distribution (see next lecture for the definition), then $X, Y$ are independent if and only if $\text{Cov}(X, Y) = 0$. Thus, if $\bar{X}, X_i - \bar{X}$ have joint Normal distribution (which is the case if, e.g., $X_i$ are i.i.d. Normals) then it follows from the above computation that $\bar{X}, X_i - \bar{X}$ are also independent. Then we CAN conclude that $S^2, \bar{X}$ are independent because $X, Y$ independent implies $X^2, Y$ independent.