31. Properties of Expectations

Covariance, Variance of Sums [Ross S7.4]

Proposition 31.1 *If X and Y are independent, then for any functions $g(x)$ and $h(y)$:*

$$\begin{aligned} i) &\quad E[g(X)h(Y)] = E[g(X)]E[h(Y)] \\ ii) &\quad g(X) \text{ and } h(Y) \text{ are independent.} \end{aligned}$$

Why?

$$\begin{aligned} i) \quad E[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty} g(x)f_X(x)dx \int_{-\infty}^{\infty} h(y)f_Y(y)dy \\ &= E[h(Y)]E[g(X)] \end{aligned}$$

ii) Let $A = \{x \in \mathbb{R} \mid g(x) \leq a\}$ and $B = \{y \in \mathbb{R} \mid h(y) \leq b\}$. Then:

$$\begin{aligned} &P[g(X) \leq a, h(Y) \leq b] \\ &\qquad = P[X \in A, Y \in B] \\ &\qquad = P[X \in A]\ P[Y \in B] && \text{since } X \text{ and } Y \text{ are independent} \\ &\qquad = P[g(X) \leq a]\ P[h(Y) \leq b] \end{aligned}$$

For a single random variable X, its mean and variance give us some information about X.

For two random variables X and Y, **covariance** (and **correlation**) will give us information about the relationship between the pair X and Y.

Definition 31.1: The **covariance** between X and Y, denoted $Cov[X, Y]$, is defined as

$$Cov[X, Y] = E[\ (X - E[X])(Y - E[Y])\ ]$$

Just as $Var[X] = E[X^2] - (E[X])^2$, we also have:

$$\begin{aligned}
Cov[X, Y] &= E[\ (X - E[X])(Y - E[Y])\ ] \\
&= E[\ XY - E[X]Y - E[Y]X + E[X]E[Y])\ ] \\
&= E[XY] + E[\ -E[X]Y\ ] + E[\ -E[Y]X\ ] + E[\ E[X]E[Y]\ ] \\
&= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] \\
&= E[XY] - E[X]E[Y]
\end{aligned}$$

Note: If X and Y are independent, then $E[XY] = E[X]E[Y]$ so $Cov[X, Y] = 0$.

Example 31.1: Does $Cov[X, Y] = 0$ imply X and Y are independent?

Solution:

Proposition 31.2

i) $Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = Cov[Y, X]$

ii) $Cov[X, X] = E[(X - \mu_X)(X - \mu_X)] = Var[X]$

iii) $Cov[aX, Y] = E[(aX - a\mu_X)(Y - \mu_Y)] = aCov[X, Y] = Cov[X, aY]$

iv) $Cov\left[\sum_{i=1}^{n} X_i,\ \sum_{j=1}^{m} Y_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov[X_i, Y_j]$

Why?

For iv), let $\quad U = \sum_{i=1}^{n} X_i \qquad V = \sum_{j=1}^{m} Y_i$

$$E[X_i] = \mu_i \qquad E[Y_j] = \nu_j$$

Then: $\quad E\left[U\right] = \sum_{i=1}^{n} \mu_i \qquad E\left[V\right] = \sum_{j=1}^{m} \nu_j$

So, $\quad Cov\left[U,\ V\right] = E\left[\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right)\left(\sum_{j=1}^{m} Y_j - \sum_{j=1}^{m} \nu_j\right)\right]$

$$= E\left[\sum_{i=1}^{n}(X_i - \mu_i)\sum_{j=1}^{m}(Y_j - \nu_j)\right]$$

$$= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i - \mu_i)(Y_j - \nu_j)\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} E\left[(X_i - \mu_i)(Y_j - \nu_j)\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} Cov[X_i, Y_j]$$

Now,

$$Var\left[\sum_{i=1}^{n} X_i\right] = Cov\left[\sum_{i=1}^{n} X_i,\ \sum_{j=1}^{n} X_j\right] \qquad \text{by ii) of Prop. 31.2}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} Cov[X_i, X_j] \qquad \text{by iv) of Prop. 31.2}$$

$$= \sum_{\substack{i,j \\ j=i}} Cov[X_i, X_j] + \sum_{\substack{i,j \\ j\neq i}} Cov[X_i, X_j]$$

$$= \sum_{i=1}^{n} Var[X_i] + \sum_{\substack{i,j \\ j\neq i}} Cov[X_i, X_j] \qquad \text{by ii) of Prop. 31.2}$$

$$= \sum_{i=1}^{n} Var[X_i] + 2\sum_{\substack{i,j \\ i<j}} Cov[X_i, X_j] \qquad \text{by i) of Prop. 31.2}$$

If for $i \neq j$, each pair X_i, X_j are independent, then:

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i]$$

Example 31.2: Compute the variance of $X \sim \mathsf{Binomial}(n, p)$.

Solution:

Example 31.3: Recall (from Example 30.3) that $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called the **sample mean**. Let

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

be the **sample variance**.

Let $X_1, \ldots, X_n$ be iid with (common) mean μ and variance σ^2.

Find a) $Var[\bar{X}]$ and b) $E[S^2]$. [(b) is hard]

Solution:

Note: Since $E[S^2] = \sigma^2$ then S^2 is an unbiased estimator of variance.

Example 31.4: Let $X_1, \ldots, X_n$ be iid with variance σ^2. Recall $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean. $X_i - \bar{X}$ is called the i**th deviation**.

Compute $Cov[X_i - \bar{X}, \bar{X}]$.

Solution: