Lecture 6.1

# Multiple Random Variables: Expectations

# Expectations for joint distributions

**Law of the Unconscious Statistician:**

- The expected value of a real-valued function $h$ of the discrete random vector $(X_1, \ldots, X_n)$ with pmf $f$ is

$$\mathbb{E}[h(X_1, \ldots, X_n)] = \sum_{x_1, \ldots, x_n} h(x_1, \ldots, x_n)\, f(x_1, \ldots, x_n),$$

  where the sum is taken over all possible values of $(x_1, \ldots, x_n)$.

- The expected value of a real-valued function $h$ of the continuous random vector $(X_1, \ldots, X_n)$ with joint pdf $f$ is

$$\mathbb{E}[h(X_1, \ldots, X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \ldots, x_n)\, f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n.$$

# Question

Suppose $(X, Y)$ has joint pdf

$$f_{X,Y}(x, y) = \begin{cases} \frac{6}{7}(1 + x^2 y), & (x, y) \in [0, 1] \times [0, 1] \\ 0, & \text{else.} \end{cases}$$

What is $\mathbb{E}(XY)$?

$$\mathbb{E}(xy) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) \mathrm{d}x \mathrm{d}y = \int_0^1 \int_0^1 xy \tfrac{6}{7}(1 + x^2 y) \mathrm{d}x \mathrm{d}y$$

$$= \tfrac{6}{7} \int_0^1 \int_0^1 xy + x^3 y^2 \mathrm{d}x \mathrm{d}y = \tfrac{6}{7} \int_0^1 \left[ \frac{x^2 y}{2} + \frac{x^4 y^2}{4} \right]_0^1 \mathrm{d}y = \ldots .$$

# Properties of the Expectation

Special cases for functions of random variable are their linear combination and their product.

Let $X_1, \ldots, X_n$ be random variables with expectations $\mu_1, \ldots, \mu_n$.

- The expectation of their linear combination is the linear combination of their individual expectations. That is, for all constants $a$, $b_1, \ldots, b_n$,

$$\mathbb{E}[a + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n] = a + b_1 \mu_1 + \cdots + b_n \mu_n$$

- If $X_1, X_2, \cdots X_n$ are <u>independent</u> random variables, then the expectation of their product is the product of their individual expectations. That is,

$$\mathbb{E}[X_1 X_2 \cdots X_n] = \mu_1 \, \mu_2 \cdots \mu_n \, .$$

# Question

We were told that if $X \sim \text{Bin}(n, p)$, then $\mathbb{E}X = np$. We can now show this is true.

We can view $X$ as the total number of successes in $n$ independent Bernoulli trials (coin flips) with success probability $p$. Introduce independent Bernoulli random variables $X_1, \ldots, X_n$, where $X_i = 1$ if the $i$th trial is a success (and $X_i = 0$ otherwise). Observe that

$$X = X_1 + \cdots + X_n \; .$$

Each Bernoulli random variable $X_i$ has expectation $p$ so

$$\mathbb{E}X = \mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}X_1 + \mathbb{E}X_2 + \cdots + \mathbb{E}X_n = np$$

# Question

Suppose $X_1, \ldots, X_{10}$ are random variables each with a $\mathcal{N}(1, 9)$ distribution. What is the expected value of the average

$$\bar{X} = \frac{1}{10}\left(X_1 + X_2 + \cdots + X_{10}\right)?$$

(a) 0

(b) 1/10

(c) 1 ✓

(d) 10

# Question - Hash tables

In a hash table, we map $n$ items independently to one of $k$ slots. Further, all $k$ slots are equally likely to be chosen for each item.

Suppose we have 20 items and 100 slots.

**Question (1):** What is the expected number of items that map to any one slot?

**Question (2):** What is the expected number of empty slots?

# Question - Hash tables

**Question (1):** What is the expected number of items that map to any one slot?

Without loss of generality, let's consider slot 1. Define

$$X_i = \begin{cases} 1, & \text{The } i^{\text{th}} \text{ item is mapped to slot 1,} \\ 0, & \text{else.} \end{cases}$$

Since all 100 slots are equally likely, we have $\mathbb{P}(X_i = 1) = 1/100$.

So,

$$\mathbb{E}(X_i) = 0 \times \mathbb{P}(X_i = 0) + 1 \times \mathbb{P}(X_i = 1) = \frac{1}{100}.$$

The total number of items mapped to slot 1 is $Y = \sum_{i=1}^{20} X_i$. So,

$$\mathbb{E}(Y) = \sum_{i=1}^{20} \mathbb{E}(X_i) = \frac{20}{100} = \frac{1}{5}.$$

# Question - Hash tables

**Question:** What is the expected number of empty slots?

Define $X_i = \begin{cases} 1, & \text{If } i^{\text{th}} \text{ slot is empty,} \\ 0, & \text{else.} \end{cases}$

The total number of empty slots is $Y = \sum_{i=1}^{100} X_i$. We have

$$\begin{aligned}
\mathbb{P}(X_i = 1) &= \mathbb{P}\left(\left\{\text{no item is mapped to } i^{\text{th}} \text{ slot}\right\}\right) \\
&= \mathbb{P}\left(\left\{\text{item 1 is not mapped to } i^{\text{th}} \text{ slot}\right\}\right. \\
&\qquad \cap \left\{\text{item 2 is not mapped to } i^{\text{th}} \text{ slot}\right\} \\
&\qquad\qquad \vdots \\
&\qquad \left. \cap \left\{\text{item 20 is not mapped to } i^{\text{th}} \text{ slot}\right\}\right)
\end{aligned}$$

## Question - Hash tables

$$\mathbb{P}(X_i = 1) = \mathbb{P}\left(\bigcap_{j=1}^{20} \left\{\text{item j is not mapped to } i^{\text{th}} \text{ slot}\right\}\right)$$
$$= \prod_{j=1}^{20} \mathbb{P}\left(\left\{\text{item j is not mapped to } i^{\text{th}} \text{ slot}\right\}\right)$$
$$= \prod_{j=1}^{20} \left(1 - \frac{1}{100}\right) = \left(1 - \frac{1}{100}\right)^{20}.$$

So,

$$\mathbb{E}X_i = 0 \times \mathbb{P}(X_i = 0) + 1 \times \mathbb{P}(X_i = 1) = \left(1 - \frac{1}{100}\right)^{20}.$$

Hence,

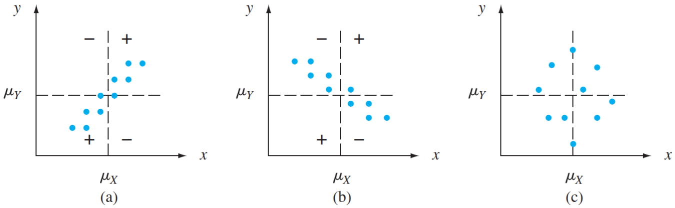
$$\mathbb{E}(Y) = \mathbb{E}X_1 + \mathbb{E}X_2 + \ldots + \mathbb{E}X_{100} = 100\left(1 - \frac{1}{100}\right)^{20}.$$

# Covariance

The **covariance** of two random variables $X$ and $Y$ with expectations $\mathbb{E}X = \mu_X$ and $\mathbb{E}Y = \mu_Y$ is defined as

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

The covariance measures the degree to which $X$ and $Y$ tend to be large at the same time in relation to their respective means, or the degree to which one tends to be large while the other is small, again in relation to their respective means,



# Properties of Covariance I

Recall the definition of the variance of $X$:

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mu_X)^2], \quad \text{where } \mu_X = \mathbb{E}X.$$

However, we usually calculate the variance as

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

There is a similar simplification of the covariance.

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right] \\
&= \mathbb{E}\left[XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y\right] \\
&= \mathbb{E}(XY) - \mathbb{E}(\mu_Y X) - \mathbb{E}(\mu_X Y) + \mathbb{E}(\mu_X \mu_Y) \\
&= \mathbb{E}(XY) - \mu_Y \mu_X - \mu_X \mu_Y + \mu_X \mu_Y \\
&= \mathbb{E}(XY) - \mu_X \mu_Y.
\end{aligned}$$

# Properties of Covariance II

The covariance is symmetric.

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right] \\ &= \mathbb{E}\left[(Y - \mu_Y)(X - \mu_X)\right] \\ &= \mathrm{Cov}(Y, X).\end{aligned}$$

The covariance of a random variable with itself is the variance.

$$\mathrm{Cov}(X, X) = \mathbb{E}\left[(X - \mu_X)(X - \mu_X)\right] = \mathbb{E}\left[(X - \mu_X)^2\right] = \mathrm{Var}(X).$$

# Properties of Covariance III

Recall that for independent random variables $X$ and $Y$,

$$\mathbb{E}\left[XY\right] = \mu_X \mu_Y,$$

where $\mu_X = \mathbb{E}X$ and $\mu_Y = \mathbb{E}Y$.

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

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

The converse is not necessarily true. For example, suppose $\mathbb{P}(X = 1) = \mathbb{P}(X = 0) = \mathbb{P}(X = -1) = 1/3$ and $Y = X^2$. Clearly, $X$ and $Y$ are dependent. But $\mathbb{E}(X) = 0$ and

$$\mathbb{E}(XY) = \mathbb{E}(X^3) = 1^3 \times \frac{1}{3} + 0^3 \times \frac{1}{3} + (-1)^3 \frac{1}{3} = 0,$$

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

# Properties of Covariance IV

Take any $a, b \in \mathbb{R}$ and random variables $X$ and $Y$. From linearity of expectations

$$\mathbb{E}(aX + bY) = a\mu_X + b\mu_Y.$$

Take another random variable $Z$. Then

$$\begin{aligned}
\operatorname{Cov}(aX + bY, Z) &= \mathbb{E}((aX + bY)Z) - (a\mu_X + b\mu_Y)\mu_Z \\
&= \mathbb{E}(aXZ + bYZ) - a\mu_X\mu_Z - b\mu_Y\mu_Z \\
&= a\,\mathbb{E}(XZ) - a\mu_X\mu_Z + b\,\mathbb{E}(YZ) - b\mu_Y\mu_Z \\
&= a\operatorname{Cov}(X, Z) + b\operatorname{Cov}(Y, Z)
\end{aligned}$$

By symmetry, the same expansion applies for $\operatorname{Cov}(Z, aX + bY)$.

# Properties of Covariance V

This property is very useful for calculating the variance of a sum of two random variables.

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{Cov}(X+Y, X+Y) \\
&= \mathrm{Cov}(X, X+Y) + \mathrm{Cov}(Y, X+Y) \\
&= \mathrm{Cov}(X, X) + \mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, X) + \mathrm{Cov}(Y, Y) \\
&= \mathrm{Var}(X) + 2\mathrm{Cov}(X, Y) + \mathrm{Var}(Y).
\end{aligned}$$

What is $\mathrm{Var}(X+Y)$ if $X$ and $Y$ are independent? Then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

# Linear Combinations of Random Variables

Let $X_1, \ldots, X_n$ be independent random variables with variances $\sigma_1^2, \ldots, \sigma_n^2$. Then,

$$\mathrm{Var}(a + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n) = b_1^2 \sigma_1^2 + \cdots + b_n^2 \sigma_n^2$$

for all constants $a$, $b_1, \ldots, b_n$.

What if the random variables $X_1, X_2, \ldots, X_n$ are not independent?

The above expression for the variance still holds if

$$\mathrm{Cov}(X_i, X_j) = 0, \qquad \text{for all } i \neq j.$$

# Correlation

Although $\text{Cov}(X, Y)$ gives a numerical measure of the degree to which $X$ and $Y$ vary together, the magnitude of $\text{Cov}(X, Y)$ is also influenced by the overall magnitudes of $X$ and $Y$.

**Correlation coefficient** is a scaled version of the covariance,

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

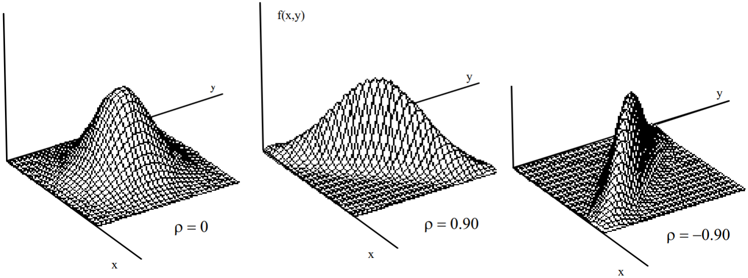
The correlation coefficient always lies between $-1$ and 1. This is a consequence of the Cauchy-Schwarz inequality

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)\text{Var}(Y)}.$$

# Correlation

The correlation is a measure of the amount of linear relationship between two random variables.

If the joint distribution of $X$ and $Y$ is relatively concentrated around a straight line in the $xy$-plane that has a positive (or negative) slope, then $\rho(X, Y)$ will typically be close to 1 (or $-1$).




In fact, $\rho(X, Y) = \pm 1$ if and only if $Y = aX + b$ for some $a \neq 0$ and $b$.

# Question

Suppose $X_1, \ldots, X_{10}$ are independent random variables each with a $\mathcal{N}(1, 9)$ distribution. What is the variance of the average

$$\bar{X} = \frac{1}{10}(X_1 + X_2 + \cdots + X_{10})?$$

(a) 9

(b) $\frac{3}{10}$

(c) $\frac{9}{10}$ ✓

(d) $\frac{9}{100}$

# Questions

Suppose $X_1$ and $X_2$ are independent random variables, each having $\text{Exp}(1/2)$ distribution.

Recall if $Y \sim \text{Exp}(\lambda)$, then $\mathbb{E}Y = \lambda^{-1}$ and $\text{Var}(Y) = \lambda^{-2}$.

Let $U = X_1 + 2X_2 + 3$. What is $\mathbb{E}(U)$?

(a) 4.5
(b) 6
(c) 9 ✓
(d) 15

## Questions

Suppose $X_1$ and $X_2$ are independent random variables, each having Exp(1/2) distribution.

Recall if $Y \sim \text{Exp}(\lambda)$, then $\mathbb{E}Y = \lambda^{-1}$ and $\text{Var}(Y) = \lambda^{-2}$.

Let $U = X_1 + 2X_2 + 3$. What is $\text{Var}(U)$?

(a) 6
(b) 20 ✓
(c) 23
(d) 29

# Questions

Suppose $X_1$ and $X_2$ are independent random variables, each having $\text{Exp}(1/2)$ distribution.

Recall if $Y \sim \text{Exp}(\lambda)$, then $\mathbb{E}Y = \lambda^{-1}$ and $\text{Var}(Y) = \lambda^{-2}$.

Let $U = X_1 + 2X_2 + 3$. What is $\text{Cov}(U, X_2)$?

(a) 2
(b) 4
(c) 7
(d) 8 ✓

# Questions

Suppose $X_1$, $X_2$, $X_3$, and $X_4$ are independent random variables, each having Exp(1/2) distribution.

Recall if $Y \sim \text{Exp}(\lambda)$, then $\mathbb{E}Y = \lambda^{-1}$ and $\text{Var}(Y) = \lambda^{-2}$.

Let $U = X_1 + 2X_2 + 3$ and $V = X_3 + X_4$. What is $\text{Cov}(U, V)$?

(a) 0 ✓

(b) 4

(c) 8

(d) 12

# Sums of Independent Random Variables

Recall that for independent random variables $X_1, \ldots, X_n$,

$$\mathbb{E}[h_1(X_1)h_2(X_2)\cdots h_n(X_n)] = \mathbb{E}[h_1(X_1)]\,\mathbb{E}[h_2(X_2)]\,\cdots\mathbb{E}[h_n(X_n)]\,.$$

This helps us to determine the MGF of a sum of independent random variables. If $X$ and $Y$ are independent random variables, then

$$M_{X+Y}(s) = \mathbb{E}e^{s(X+Y)} = \mathbb{E}\left(e^{sX}e^{sY}\right) = \mathbb{E}e^{sX}\mathbb{E}e^{sY} = M_X(s)\,M_Y(s)$$

In general, the MGF of $X_1 + \cdots + X_n$ is

$$M_{X_1+\cdots+X_n}(s) = M_{X_1}(s)\cdots M_{X_n}(s).$$

# Example

Suppose $X \sim \text{Bin}(n, p)$. Then $X = X_1 + X_2 + \cdots + X_n$ where $X_i$ are independent Bernoulli random variables with success probability $p$.

The MGF of the $\text{Ber}(p)$ distribution is

$$M_{X_i}(s) = \mathbb{E}e^{sX_i} = (1-p)\, e^{s \cdot 0} + pe^{s \cdot 1} = 1 - p + pe^{s}.$$

The MGF of the $\text{Bin}(n, p)$ distribution is

$$M_X(s) = M_{X_1}(s) \cdots M_{X_n}(s) = (1 - p + pe^{s})^{n}\,.$$

# Sums of Independent Normal Random Variables

**Recall:** If $X \sim \mathcal{N}(\mu, \sigma^2)$, $M_X(s) = \exp\left(\mu s + \frac{\sigma^2}{2} s^2\right)$, $s \in \mathbb{R}$.

Let $X_1, \ldots, X_n$ be independent rvs as $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$.

For any $a$, $b_1, \ldots, b_n$, consider

$$Y = a + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n.$$

What is the distribution of $Y$?

## Sums of Independent Normal Random Variables

$$
\begin{aligned}
M_Y(s) &= \mathbb{E}\left(e^{s(a+b_1X_1+b_2X_2+\cdots+b_nX_n)}\right) \\
&= \mathbb{E}\left(e^{sa}e^{sb_1X_1}e^{sb_2X_2}\ldots e^{sb_nX_n}\right) \\
&= e^{sa}M_{X_1}(sb_1)\ \ldots\ M_{X_n}(sb_n) \\
&= \exp(sa)\exp\left(\mu_1 sb_1 + \frac{\sigma_1^2}{2}s^2b_1^2\right)\ldots\exp\left(\mu_n sb_n + \frac{\sigma_n^2}{2}s^2b_n^2\right) \\
&= \exp\left((a+\mu_1 b_1+\ldots+\mu_n b_n)\,s + \frac{(\sigma_1^2b_1^2+\ldots+\sigma_n^2b_n^2)}{2}s^2\right)
\end{aligned}
$$

So,

$$
Y \sim \mathcal{N}\left(a+\sum_{i=1}^n b_i\mu_i, \sum_{i=1}^n b_i^2\sigma_i^2\right).
$$