

EE 325: Probability and Random Processes

Module 3: Expectations, Functions of a Random Variable, Higher Moments

D Manjunath

Networking Lab
Department of Electrical Engineering
Indian Institute of Technology

September 28, 2020

Topics in Module 3

- Expectation of random variable.
- Functions of a random variable of the form $g(X)$.
- Moments of a random variable.
- Moment generating functions and characteristic functions of random variables.
- Mostly from Chapter 5 of the text; Inequalities will be in the next module.

Expectation

- In a coin toss, let $H = 1$ and $T = -1$ be the values associated with heads and tails, i.e., the random variable $X \in \{-1, 1\}$
- Toss a coin (not necessarily a fair one, call it the Shakuni coin) N times; Let N_h be the number of heads in the N tosses.
- Sample realisation: H T T T H H H T H H T H H T H T
- The sample mean is

$$\begin{aligned}\frac{N_H - N_T}{N} &= \frac{N_H(+1) + N_T(-1)}{N} \\ &= \frac{N_H}{N}(+1) + \frac{N_T}{N}(-1)\end{aligned}$$

- As $N \rightarrow \infty$, $\frac{N_H}{N} \rightarrow p$ and $\frac{N_T}{N} \rightarrow (1 - p)$; average, denoted by \bar{X} , will be

$$p_{+1}(+1) + p_{-1}(-1)$$

- Generalising, for N samples of a discrete random variable, the sample mean would be

$$\bar{X} = \sum_i \frac{i N_i}{N} = \sum_i i \left(\frac{N_i}{N} \right)$$

- Formally, the expectation of a discrete random variable is defined as

$$E(X) = \sum_i i p_X(i)$$

- For continuous random variables

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

- It is possible that the summation, or the integration, is not finite. In this case

Example Expectations

- In a coin toss, let $H = 1$ and $T = -1$ be the values associated with heads and tails, i.e., the random variable $X \in \{-1, 1\}$

$$\Pr(X = 1) = 0.3 \quad \Pr(X = -1) = 0.7$$

$$E(X) = 0.3 - 0.7 = -0.4$$

- 0 – 1 Bernouilli Random Variable:

$$X \in \{0, 1\}$$

$$p_X(1) = \alpha$$

$$p_X(0) = 1 - \alpha$$

$$E(X) = \alpha$$

Example Expectations

- Two fair, six-sided dice are thrown; X is the sum of the values on the top face.

$p_X \downarrow X \rightarrow$	2	3	4	5	6	7
	1/36	2/36	3/36	4/36	5/36	6/36

$p_X \downarrow X \rightarrow$	8	9	10	11	12
	5/36	4/36	3/36	2/36	1/36

$$\begin{aligned} E(X) &= (2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12)/36 \\ &= 252/36 \end{aligned}$$

- Binomial Random Variable: Count the number of Heads from N independent coin tosses.

$$X \in \{0, 1, \dots, N\}$$

$$\Pr(X = k) = p_X(k) = \begin{cases} \binom{N}{k} \alpha^k (1 - \alpha)^{N-k} & \text{for } 0 \leq k \leq N \\ 0 & \text{otherwise} \end{cases}$$

Expectation of a Binomial RV

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^N k \binom{N}{k} \alpha^k (1-\alpha)^{N-k} \\&= \sum_{k=1}^N k \frac{N!}{k! (N-k)!} \alpha^k (1-\alpha)^{N-k} \\&= N\alpha \sum_{k=1}^N \frac{(N-1)!}{(k-1)! (N-k)!} \alpha^{k-1} (1-\alpha)^{N-k} \\&= N\alpha \sum_{k=1}^N \frac{(N-1)!}{(k-1)! (N-1)-(k-1)!} \alpha^{k-1} (1-\alpha)^{(N-1)-(k-1)} \\&= N\alpha \sum_{k_1=0}^{N-1} \frac{(N-1)!}{k_1! ((N-1)-k_1)!} \alpha^{k_1} (1-\alpha)^{(N-1)-k_1} \\&= N\alpha \left[\sum_{k_1=0}^{N-1} \binom{N-1}{k_1} \alpha^{k_1} (1-\alpha)^{(N-1)-k_1} \right] = N\alpha\end{aligned}$$

Expectation of a Geometric RV

$$X \in \{0, 1, \dots\}$$

$$\Pr(X = k) = p_X(k) = \begin{cases} (1 - \alpha)^k \alpha & \text{for } k \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$E(X) = \sum_{k=0}^{\infty} k(1 - \alpha)^k \alpha$$

An identity

$$\begin{aligned} \sum_{k=0}^{\infty} x^k &= \frac{1}{1 - x} \\ \sum_{k=0}^{\infty} kx^{k-1} &= \frac{1}{(1 - x)^2} \end{aligned}$$

$$E(X) = \alpha(1 - \alpha) \sum_{k=0}^{\infty} k(1 - \alpha)^{k-1} = \alpha(1 - \alpha) \frac{1}{\alpha^2} = \frac{1 - \alpha}{\alpha}$$

Expectation of a Poisson Random Variable

$$X \in \{0, 1, \dots\}$$

$$\Pr(X = k) = p_X(k) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda} & \text{for } k \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{k_1=0}^{\infty} \frac{\lambda^{k_1}}{(k_1)!} \\ &= \lambda \end{aligned}$$

Expectation of exponential distribution

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

To obtain the expectation from the definition, perform integration by parts with $u = x$ and $dv = \lambda e^{-\lambda x} dx$

$$\begin{aligned} E(X) &= \int_0^{\infty} x \lambda e^{-\lambda x} dx \\ &= x e^{-\lambda x} \Big|_0^{\infty} - \int_0^{\infty} -e^{-\lambda x} dx \\ &= x e^{-\lambda x} \Big|_0^{\infty} + \left(-\frac{1}{\lambda} \right) e^{-\lambda x} \Big|_0^{\infty} \\ &= \frac{1}{\lambda} \end{aligned}$$

Expectation of the Gaussian distribution

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Let $y = x - \mu$ and we get

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (y + \mu) e^{-\frac{y^2}{2\sigma^2}} dy \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} y e^{-\frac{y^2}{2\sigma^2}} dy + \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \mu e^{-\frac{y^2}{2\sigma^2}} dy \\ &= 0 + \mu \end{aligned}$$

- Conditional pdf and pmf can be used to define conditional expectation.

$$E(X|A) = \int_{-\infty}^{\infty} x f_{X|A}(x|A) dx$$

$$E(X|A) = \sum_{x=-\infty}^{\infty} x p_{X|A}(x|A)$$

Function of a Random Variable

- Recall the notion of a function: Map the elements of the *domain* to the elements of the *range* (or *co-domain*).
- Consider a function $g(\cdot) : \mathcal{R} \rightarrow \mathcal{R}$.
- Let X be a random variable and let Y be the variable for $g(\cdot)$, i.e., consider $Y = g(X)$.
- Can view $g(X)$ as a map from Ω to the set of numbers and hence we can say $Y = g(X)$ is a random variable.

Function of a Random Variable

- Consider $\Pr(Y \in A)$; this is the same as $\Pr(g(X) \in A)$.
- Let B be a subset of Ω defined as $B := \{\omega : g(X(\omega)) \in A\}$.
- Clearly, $\Pr(Y \in A) = \Pr(g(X) \in A) = \Pr(B)$.
- The event $\{Y \leq y\} = \{g(X) \leq y\} = \{X \in g^{-1}(-\infty, y)\}$.
- This essentially implies

$$F_Y(y) = \Pr(Y \leq y) = \Pr(X \in g^{-1}(-\infty, y)).$$

Function of a Random Variable

- $Y = g(X)$, g is not necessarily one-to-one.
- $f_Y(y) \approx \Pr(y < Y \leq y + dy)$.
- Need to obtain $f_Y(y)$ from $f_X(x)$.
- Assume $y = g(x)$ has k solutions: x_1, \dots, x_k .
- Event $\{y < Y \leq y + dy\}$ can occur if any the following occur:
 $\{x_1 < X \leq x_1 + dx_1\}, \dots, \{x_k < X \leq x_k + dx_k\}$
- These are mutually exclusive events. Thus

$$\Pr(y < Y \leq y + dy) = \Pr(x_1 < X \leq x_1 + dx_1) + \dots + \Pr(x_k < X \leq x_k + dx_k)$$

- Each of these events can be written as

$$\Pr(x_i < X \leq x_i + dx_i) = f_X(x_i)dx_i = f_X(x_i) \frac{dy}{|g'(x_i)|}$$

- Thus

$$f_Y(y)dy = \left(\sum_{i=1}^k \frac{f_X(x_i)}{|g'(x_i)|} \right) dy \implies f_Y(y) = \left(\sum_{i=1}^k \frac{f_X(x_i)}{|g'(x_i)|} \right)$$

- $Y = X^2$.
- Specialise this for the case when X is a unit normal random variable.
- $Y = |X|$
- X is uniform in $[-5, 5]$ and

$$Y = \begin{cases} -1 & \text{if } X < -1 \\ X & \text{if } -1 \leq X \leq 1 \\ 1 & \text{if } X > 1 \end{cases}$$

- X is exponential with parameter λ and $Y = \lfloor X \rfloor$.
- X is uniformly distributed in $(0, \pi/2)$. $Y = \sin X$.
- X is uniformly distributed in $(0, 1)$. $Y = \sin^{-1} X$.

Expectation of a Function of a Random Variable

$$E(g(X)) = \sum_k g(k) p_X(k)$$

$$E(g(X)) = \int g(x) f_X(x) dx$$

Analogously

$$E(g(X, Y)) = \sum_j \sum_k g(j, k) p_{X,Y}(j, k)$$

$$E(g(X, Y)) = \int \int g(x, y) f_{X,Y}(x, y) dx dy$$

More about Expectations

- Non negative random variables have non negative expectations

$$X \geq 0 \text{ i.e., } f_X(x) = 0 \text{ for } x \leq 0, \implies E(X) \geq 0$$

- Bounded random variables have bounded expectations

$$a \leq X \leq b \text{ i.e., } f_X(x) = 0 \text{ for } x \notin [a, b], \implies a \leq E(X) \leq b.$$

- Expectation of a constant is a constant, i.e., if $X = c$, then
- Expectation is a linear operation, i.e.

$$\begin{aligned} E(aX + b) &= \int_{-\infty}^{\infty} (ax + b)f_X(x) \\ &= a \int_{-\infty}^{\infty} x f_X(x) dx + b \int_{-\infty}^{\infty} f_X(x) dx \\ &= a\bar{X} + b \end{aligned}$$

- What is $E(X - \bar{X})$?

More about Expectations

- Compare $E(X - \bar{X})$ with $E(X - a)$ for a constant a .

$$\begin{aligned}E\left((X - a)^2\right) &= E\left(\left((X - \bar{X}) + (\bar{X} - a)\right)^2\right) \\&= E\left((X - \bar{X})^2 + (\bar{X} - a)^2 + 2(X - \bar{X})(\bar{X} - a)\right) \\&= E\left((X - \bar{X})^2\right) + E\left((\bar{X} - a)^2\right) + 2E\left((X - \bar{X})(\bar{X} - a)\right) \\&= E\left((X - \bar{X})^2\right) + (\bar{X} - a)^2 + 2(\bar{X} - a)E\left((X - \bar{X})\right) \\&= E\left((X - \bar{X})^2\right) + (\bar{X} - a)^2 \\E\left((X - a)^2\right) &\geq E\left((X - \bar{X})^2\right)\end{aligned}$$

More Expectations: Moments

- k -th moment:

$$E(X^k) = \sum_i i^k p_X(i)$$

$$E(X^k) = \int_{-\infty}^{\infty} x^k f_X(x) dx$$

- k -th central moment: $(X - \bar{X})^k$

$$E((X - \bar{X})^k) = \sum_i (i - \bar{X})^k p_{bX}(i)$$

$$E(X^k) = \int_{-\infty}^{\infty} (x - \bar{X})^k f_X(x) dx$$

Variance of a Random Variable

- Variance of a random variable X is defined as the expectation of $(X - E(X))^2$ and we will denote it by $\text{VAR}(X)$.

$$\begin{aligned}\text{VAR}(X) &= E\left((X - \bar{X})^2\right) \\ &= E(X^2 - 2X\bar{X} + \bar{X}^2) \\ &= E(X^2) - 2\bar{X}E(X) + \bar{X}^2 \\ &= E(X^2) - \bar{X}^2\end{aligned}$$

$$\begin{aligned}E(X^2) &= \sum_k k^2 p_X(k) \\ &= \int x^2 f_X(x) dx\end{aligned}$$

Variance: Examples

- In a coin toss, let $H = 1$ and $T = -1$ be the values associated with heads and tails, i.e., the random variable $X \in \{-1, 1\}$
 $\Pr(X = 1) = 0.3$ and $\Pr(X = -1) = 0.7$
- Toss two independent coins; Along the lines of the previous random variable, let $HH = 2$, $HT = TH = 0$ and $TT = -2$.
 $\Pr(X = 2) = 0.09$, $\Pr(X = 0) = 0.42$, and $\Pr(X = -1) = 0.49$

Variance of a Random Variable

- What is the variance of a constant?
- Variance of a scalar multiple of a random variable

$$\begin{aligned}\text{VAR}(aX) &= E\left((aX - E(aX))^2\right) \\ &= E\left((aX - aE(X))^2\right) \\ &= E\left(a^2 (X - E(X))^2\right) \\ &= a^2 E\left((X - E(X))^2\right) \\ &= a^2 \text{VAR}(X)\end{aligned}$$

- 0 – 1 Bernouilli Random Variable:

$$X \in \{0, 1\}$$

$$p_X(1) = \alpha, \quad p_X(0) = 1 - \alpha$$

$$E(X) = \alpha, \quad E(X^2) = \alpha$$

$$\text{VAR}(X) = \alpha - \alpha^2 = \alpha(1 - \alpha)$$

Variance of a Random Variable

- Geometric(α) random variable: $\frac{1-\alpha}{\alpha^2}$
- Binomial(N, α) random variable: $N\alpha(1 - \alpha)$
- Poisson(λ) random variable: λ
- Uniform(0, 1) random variable: $1/12$
- Exponential(λ) random variable: $1/\lambda^2$
- Gaussian(μ, σ) random variable: σ

More Expectations: Generating and Characteristic Functions

- If X is a discrete random variable, z^X : $E(z^X)$ is called the moment generating function (MGF), and also the probability generating function (PGF).

$$\mathcal{F}_X(z) := E(z^X) = \sum_i z^i p_X(i)$$

- If X is a continuous random variable, z^X : $E(e^{j\omega X})$ is called the characteristic function. It is also the MGF.

$$\phi_X(j\omega) := E(e^{j\omega X}) = \int_{-\infty}^{\infty} e^{j\omega x} f_X(x) dx$$

Often we use $j\omega = s$. Thus

$$\phi_X(s) := E(e^{sX}) = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx$$

Moment and Probability Generating Functions

- For discrete random variables, recall the definition

$$\mathcal{F}_X(z) := \mathbb{E}(z^X) = \sum_i z^i p_X(i).$$

- Evaluating $\mathcal{F}_X(z)$ at $z = 0, 1$

$$\mathcal{F}_X(0) = \Pr(X = 0) = p_0$$

$$\mathcal{F}_X(1) = \sum_i p_X(i) = 1$$

- Take the derivative of $\mathcal{F}_X(z)$ w.r.t. z and evaluating at 0 and 1

$$\mathcal{F}'_X(z) = \frac{d}{dz} \left(\sum_i z^i p_X(i) \right) = \sum_i i z^{i-1} p_X(i)$$

- $\mathcal{F}'_X(0) = p_X(1)$ $\mathcal{F}'_X(1) = \mathbb{E}(X)$

Moment and Probability Generating Functions

- More generally

$$\frac{\mathcal{F}_X^{(i)}(0)}{i!} = p_X(i)$$

$$\mathcal{F}_X^{(k)}(1) = E(X(X-1)(X-2)\dots(X-k+1))$$

The last LHS is also called the k -th factorial moment.

Moments from MGF of Continuous RVs

- First moment from the Moment Generating Function

$$\phi'_X(s) = \frac{d}{ds} \left(\int_{-\infty}^{\infty} e^{sx} f_X(x) dx \right) = \int_{-\infty}^{\infty} x e^{sx} f_X(x) dx$$

$$\phi'_X(0) = \int_{-\infty}^{\infty} x f_X(x) dx = E(X)$$

- Second moment from the Moment Generating Function

$$\phi''_X(s) = \frac{d^2}{ds^2} \left(\int_{-\infty}^{\infty} e^{sx} f_X(x) dx \right) = \int_{-\infty}^{\infty} x^2 e^{sx} f_X(x) dx$$

$$\phi''_X(0) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = E(X^2)$$

- In general

$$\phi_X^{(n)}(0) = \int_{-\infty}^{\infty} x^n f_X(x) dx = E(X^n)$$

Two or More Random Variables

- On the same sample space we can define multiple random variables; begin with two, say X and Y .
- Like joint probability of two events, can define joint pmf.
- Example: Roll a fair die till you get a 1. Let the total number rolls be the random variable X . Let the number of 2s before the first 1 be the random variable Y . Let the number of evens before the first 1, be the random variable Z . Verify the following conditional and joint pmfs.

$$p_{Z|X}(k|i) = \binom{i-1}{k} (3/5)^k (2/5)^{i-k} \quad \text{for } k = 0, 1, \dots, i-1$$

$$p_{XY}(i,j) = \binom{i-1}{j} (1/5)^j (4/5)^{i-1-j} (5/6)^{i-1} (1/6) \\ \text{for } i = 1, 2, \dots \text{ and } j = 0, 1, \dots, i-1$$

$$p_{XZ}(i,k) = \binom{i-1}{k} (3/5)^k (2/5)^{i-k} (5/6)^{i-1} (1/6) \\ \text{for } i = 1, 2, \dots \text{ and } k = 0, 1, \dots, i-1$$

Two or More Random Variables

- More generally, like the joint probability of two events and the joint pmf, can define joint cdf for X and Y .

$$F_{X,Y}(x,y) := \Pr(X \leq x, Y \leq y)$$

$$f_{X,Y}(x,y) := \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

- Informally, $f_{X,Y}(x,y) dx dy$ is the probability that X and Y lie in the infinitesimal area of $dx dy$ near (x,y)
- Some easy identities.

$$F_X(x) = F_{XY}(x, \infty)$$

$$F_Y(y) = F_{XY}(\infty, y)$$

$$\Pr(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1)$$

Two or More Random Variables

- More identities

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) \, dx \, dy = 1$$
$$\int_{-\infty}^{\infty} f_{XY}(x, y) \, dy = f_X(x)$$

- X and Y are independent if

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

This also implies

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

Two or More Random Variables: Tutorial Problems

- 1 Obtain the distribution of $Z = \min(X, Y)$
Key observation: $\{Z > z\} \implies \{X > z, Y > z\}$
- 2 Obtain the distribution of $Z = \max(X, Y)$
Key observation: $\{Z \leq z\} \implies \{X \leq z, Y \leq z\}$
- 3 Order statistics of N iid variables: Take the N realisations of the random variables and sort them in, say increasing order. The r -th element of the sorted list is called the r -th order statistics.
Use the results from either of the preceding examples.
- 4 Obtain the distribution of $Z = X + Y$.

Expectations of Functions of Two Random Variables

- Let $Z = g(X, Y)$. Using arguments identical to the mean of a function of a random variable, we can show

$$E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dx dy$$



$$\begin{aligned} E(X + Y) &= \int \int (x + y) f_{X,Y}(x, y) dx dy \\ &= \int \int x f_{X,Y}(x, y) dx dy + \int \int y f_{X,Y}(x, y) dx dy \\ &= \int x \int f_{X,Y}(x, y) dy dx + \int y \int f_{X,Y}(x, y) dx dy \\ &= \int x f_X(x) dx + \int y f_Y(y) dy \\ &= E(X) + E(Y) \end{aligned}$$

Observe that there were NO assumptions on X and Y .

Covariance of two Random Variables

- Covariance is a measure of two random variables, say X and Y are related. How they vary ‘together.’
- **Informally**
 - Want to know if high values of X means high values of Y **and** low values of X means low values of Y .
 - Could be the opposite: High values of X means low values of Y **and** low values of X means high values of Y .
 - They could ‘balance out’: high values of X could give rise to both high and low values of Y . Similarly for low values.
 - We now define the notions of positive correlation, negative correlation and uncorrelatedness between two random variables.

Covariance of two Random Variables

- For convenience we will write $E(X) = \mu_X$ and $E(Y) = \mu_Y$.
- Denote covariance of X and Y by $\text{COV}(X, Y)$ and define

$$\text{COV}(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

Thus

$$\begin{aligned}\text{COV}(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) \\&= E(XY - \mu_X Y - X\mu_Y + \mu_X \mu_Y) \\&= E(XY) - E(\mu_X Y) - E(X\mu_Y) + E(\mu_X \mu_Y) \\&= E(XY) - \mu_X E(Y) - E(X) \mu_Y + \mu_X \mu_Y \\&= E(XY) - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\&= E(XY) - \mu_X \mu_Y\end{aligned}$$

- What is $\text{COV}(X, X)$?

Covariance of two Independent Random Variables

Let X and Y be two independent random variables. Want to know $E(XY)$.

$$\begin{aligned} E(XY) &= \sum_j \sum_k j k p_{X,Y}(j, k) \\ &= \sum_j \sum_k j k p_X(j) p_Y(k) \\ &= \sum_j j p_X(j) \sum_k k p_Y(k) \\ &= \left(\sum_j j p_X(j) \right) \left(\sum_k k p_Y(k) \right) \\ &= E(X) E(Y) \end{aligned}$$

$$\text{COV}(X, Y) = ??$$

Thus independence implies uncorrelatedness. Does uncorrelatedness imply independence?

Covariance Identities

- $\text{COV}(X + Y, Z) = \text{COV}(X, Z) + \text{COV}(Y, Z)$
- $\text{COV}(X_1 + X_2, Y_1 + Y_2) = \text{COV}(X_1, Y_1) + \text{COV}(X_1, Y_2) + \text{COV}(X_2, Y_1) + \text{COV}(X_2, Y_2)$
- More generally,

$$\text{COV}\left(\sum_{i=1}^m X_i, \sum_{j=1}^n Y_j\right) = \sum_{i,j} \text{COV}(X_i, Y_j)$$

- Variance of the sum of random variables

$$\text{VAR}(X_1 + X_2) = \text{VAR}(X_1) + \text{VAR}(X_2) + 2\text{COV}(X_1, X_2)$$

- More generally

$$\text{VAR}\left(\sum_{i=1}^m X_i\right) = ???$$

Covariance Identities

- Variance of the sum of arbitrary random variables

$$\text{VAR}\left(\sum_{i=1}^m X_i\right) = \sum_{i=1}^m \text{VAR}(X_i) + \sum_{\substack{j=1 \\ j \neq i}}^m \text{COV}(X_i, X_j)$$

- Correlation coefficient of two random variables

$$\text{Corr}(X, Y) = \frac{\text{COV}(X, Y)}{\sqrt{\text{VAR}(X) \text{VAR}(Y)}}$$

- Can show that $|\text{Corr}(X, Y)| \leq 1$
- Notation: $\rho_{XY} = \text{Corr}(X, Y)$.

Conditional Distributions

- Conditional distributions

$$F_Y(y|X \leq x) = \frac{\Pr(Y \leq y, X \leq x)}{\Pr(X \leq x)} = \frac{F_{XY}(x, y)}{F_X(x)}$$

$$f_Y(y|X \leq x) = \frac{\partial F_{XY}(x, y) / \partial y}{F_X(x)}$$

$$F_Y(y|x_1 \leq X \leq x_2) = \frac{\Pr(Y \leq y, x_1 < X \leq x_2)}{\Pr(x_1 \leq X \leq x_2)} = \frac{F_{XY}(x_2, y) - F_{XY}(x_1, y)}{F_X(x_2) - F_X(x_1)}$$

$$f_Y(y|X = x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

- Total probability

$$f_X(x) = \int_{-\infty}^{\infty} f_X(x|y) f_Y(y) dy$$

- Bayes theorem

$$f_Y(y|x) = \frac{f_X(x|y) f_Y(y)}{\int_{-\infty}^{\infty} f_X(x|y) f_Y(y) dy}$$

Conditional Expectation

- Conditional expectation of X given $Y = y$ is

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_X(x|Y = y) dx$$

- Observe that this is a function of y and is a random variable—depends on the value of Y from the experiment.
- Can define its moments. Leading us to chain rule of expectations:

$$\begin{aligned} E(E(X|Y)) &= \int_{-\infty}^{\infty} E(X|Y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx \right) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X|Y}(x,y) f_Y(y) dy \right) dx \\ &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{XY}(x,y) dy \right) dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx = E(X) \end{aligned}$$