ECE 203 – Section 6
Properties of Expectations

- Expectations of sums of random variables
- Covariance, variance of sums, and correlation
- Conditional expectations and variances
- Moment generating function and its properties
- Multivariate normal random variables

The slides have been prepared based on the lecture notes of Prof. Patrick Mitran.

- Recall that the mean value of X is

$$E[X] = \begin{cases} \sum_x x p_X(x), & \text{if } X \text{ is discrete} \\ \\ \int_{-\infty}^{\infty} x f_X(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

- Since $E[X]$ is a weighted sum of *all possible values* of X, when we have $P[a \le X \le b] = 1$, then $a \le E[X] \le b$. Why?

- Since $f_X(x) = 0$ for $x \notin [a, b]$, we have

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b x f_X(x) dx \ge$$

$$\ge \int_a^b a f_X(x) dx = a \int_a^b f_X(x) dx = a$$

- **Proposition:** Let X and Y be two random variables. Let $g(x,y)$ be a function. Then

$$E[g(X,Y)] = \begin{cases} \sum_y \sum_x g(x,y) p_{XY}(x,y) & \text{if } X \ \& \ Y \text{ are discrete} \\ \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y) dx dy & \text{if } X \ \& \ Y \text{ are continuous} \end{cases}$$

- Indeed, since $E[Z] = \int_0^{\infty} P[Z > t] dt$, we have

$$E[g(X,Y)] = \int_0^{\infty} P[g(X,Y) > t] dt = \int_0^{\infty} \iint_{(x,y):g(x,y)>t} f_{XY}(x,y) dx dy \, dt =$$

$$= \iint_{\mathbb{R}^2} \int_0^{g(x,y)} f_{XY}(x,y) dt \, dx dy = \iint_{\mathbb{R}^2} g(x,y) f_{XY}(x,y) dx dy$$

- The discrete case can be proven similarly.

Example

- Two persons are randomly positioned on a road of length L. If their positions are independent, and uniformly distributed on the length L, what is the mean distance between them?

- *Solution:* Let X and Y be the positions. Then

$$f_{XY}(x,y) = \begin{cases} \frac{1}{L^2}, & 0 < x < L, 0 < y < L \\ 0, & \text{otherwise} \end{cases}$$

We want

$$E[|X-Y|] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x-y| f_{XY}(x,y)dxdy = \frac{1}{L^2}\int_0^L\int_0^L |x-y|dxdy$$

Now,

$$\int_0^L |x-y|dx = \int_0^y (y-x)dx + \int_y^L (x-y)dx = \frac{L^2}{2} + y^2 - yL$$

So,

$$E[|X-Y|] = \frac{1}{L^2}\int_0^L \left(\frac{L^2}{2} + y^2 - yL\right) dy = \frac{L}{3}$$

- In the continuous case, use $g(x, y) = x + y$ to find $E[X + Y]$.
- *Solution:*

$$E[X+Y] = E[g(X,Y)] =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f_{XY}(x,y)dxdy =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)dydx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{XY}(x,y)dxdy$$

$$= \int_{-\infty}^{\infty} x f_X(x)dx + \int_{-\infty}^{\infty} y f_Y(y)dy = E[X] + E[Y]$$

- By induction, we also have

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$$

for any set of random variables X_1, ..., X_n.

The sample mean

- **Defintion:** Let $X_1, X_2, \ldots, X_n$ be iid with mean μ. The quantity

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  is called the **sample mean**.

- What is $E[\bar{X}]$?

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

Example

- In one of our earlier examples, n persons throw their hat into the centre of a room and pick a hat at random. Let $Z = \#$ people that get back their hat. What is $E[Z]$?

- *Solution:* Let

$$X_i = \begin{cases} 1, & \text{if person } i \text{ get back their hat} \\ 0, & \text{otherwise} \end{cases}$$

Then $Z = X_1 + \cdots + X_n$, while $P[X_i = 1] = 1/n$. Consequently,

$$E[Z] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = \frac{1}{n} + \cdots + \frac{1}{n} = 1$$

Another example

- You and 9 friends play a game at a carnival. You each have one ball, and must try to hit 10 moving targets. When the game starts, each of you picks your target randomly, independently of the others. Each of you has probability p of hitting your target. What is the expected number of targets not hit?

- *Solution:* Let $X_i = 1$ when target i is not hit, and 0 otherwise.

  Each person will, independently, hit target i with probability $p/10$. So,

$$P[X_i = 1] = \left(1 - \frac{p}{10}\right)^{10}$$

  Further,

$$E[X_i] = 1 \cdot P[X_i = 1] + 0 \cdot P[X_i = 0] = \left(1 - \frac{p}{10}\right)^{10}$$

  and, consequently,

$$E[X] = E[X_1 + \ldots + X_{10}] = E[X_1] + \ldots + E[X_{10}] = 10\left(1 - \frac{p}{10}\right)^{10}$$

- Suppose that for random variables X and Y, we have $X \geq Y$ (i.e., $X(s) \geq Y(s)$ for every $s \in S$).
- Then $Z = X - Y \geq 0$, implying $E[Z] \geq 0$.
- Equivalently $E[X - Y] = E[X] - E[Y] \geq 0$ and, thus, $E[X] \geq E[Y]$.
- To conclude,

$$X \geq Y \quad \Longrightarrow \quad E[X] \geq E[Y]$$

Boole's inequality

- Let $A_1, \ldots, A_n$ be events, and $X_1, \ldots, X_n$ be the indicator variables

$$X_i = \begin{cases} 1, & A_i \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$$

  Then $X = \sum_{i=1}^{n} X_i$ is the # events that have occurred.

- Let Y be another random variable that is equal to 1 when $X \geq 1$, and 0, when $X = 0$. Then $X \geq Y$ and, therefore, $E[X] \geq E[Y]$.

- Specifically,

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P[A_i]$$

$$E[Y] = P[\{\text{at least one } A_i \text{ occurs}\}] = P[A_1 \cup \cdots \cup A_n]$$

  and, therefore,

$$P[A_1 \cup \cdots \cup A_n] \leq \sum_{i=1}^{n} P[A_i]$$

- **Proposition:** If X and Y are independent, then for any functions g and h:

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

- Indeed,

$$E[g(X)h(Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y)dxdy =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy =$$

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

$$= E[h(Y)]E[g(X)]$$

- **Proposition:** If X and Y are independent, then for any functions g and h, $g(X)$ and $h(Y)$ are independent.

- To see that let $A = \{x \mid g(x) \leq a\}$ and $B = \{y \mid h(y) \leq b\}$. Then

$$P[g(X) \leq a, h(Y) \leq b] = P[X \in A, Y \in B] = P[X \in A]\; P[Y \in B] =$$

$$= P[g(X) \leq a]\; P[h(Y) \leq b]$$

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

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

- **Definition:** The *covariance* between X and Y, denoted $Cov[X, Y]$, is defined to be

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

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

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

$$= E\left[\ XY + (-E[X]Y) + (-E[Y]X) + E[X]E[Y])\ \right]$$

$$= E[XY] + E[\ -E[X]Y\ ] + E[\ -E[Y]X\ ] + E[\ E[X]E[Y]\ ]$$

$$= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y]$$

$$= E[XY] - E[X]E[Y]$$

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

Example

- Does $Cov[X, Y] = 0$ imply X and Y are independent? Actually, not always.

- To see that, let $P[X = 0] = P[X = 1] = P[X = -1] = 1/3$. Also, let

$$Y = \begin{cases} 0, & X \neq 0 \\ 1, & X = 0 \end{cases}$$

1. X and Y are not independent, since

$$P[X = 0, Y = 0] = 0 \neq P[X = 0]P[Y = 0]$$

while $P[X = 0] = 1/3$ and $P[Y = 0] = 2/3$.

2. $X\,Y = 0$, so $E[X\,Y] = 0$, and, since $E[X] = 0$, we have

$$Cov[X, Y] = E[XY] - E[X]E[Y] = 0 - 0 \cdot E[Y] = 0$$

- So, in this example, $Cov[X, Y] = 0$, while X and Y are dependent.

Properties of covariance

- Covariance has a number of important and useful properties which are summarized below.

- **Proposition:**

  1. $Cov[X,Y] = Cov[Y,X]$
  2. $Cov[X,X] = Var[X]$
  3. $Cov[aX,Y] = a\,Cov[X,Y]$
  4. $Cov\left[\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov[X_i, Y_j]$

- For proof, see the textbook.

Sample variance

- Recall the sample mean of iid $X_1, X_2, \ldots X_n$ (with mean μ and variance σ^2) is defined as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Then $X_i - \bar{X}$ is called the *i-th deviation*, and

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

  is called the **sample variance**.

- The variance of $\bar{X}$ is given by

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

$$= \frac{1}{n^2} Var\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var\left[X_i\right] = \frac{\sigma^2}{n}$$

- Note that

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

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

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

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

and, hence,

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

$$= \sum_{i=1}^{n} E[(X_i - \mu)^2] - nVar[\bar{X}] = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2$$

suggesting that $E[S^2] = \sigma^2$.

Example

- Compute the variance of $X \sim \text{Binomial}(n, p)$.

- *Solution:* $X = X_1 + \cdots + X_n$ where $X_1, ..., X_n$ are iid and distributed according to Bernoulli(p).

  Now $Var[X_i] = E[X_i^2] - (E[X_i])^2 = p - p^2 = p(1-p)$ and, therefore,

$$Var[X] = Var[X_1 + \ldots + X_n] = Var[X_1] + \ldots + Var[X_n] =$$
$$= nVar[X_1] = np(1-p)$$

- Thus, we have

$$X \sim \text{Binomial}(n, p) \quad \Longrightarrow \quad E[X] = np, \; Var[X] = np(1-p)$$

- The *correlation* of two random variables X and Y, denoted $\rho(X,Y)$, is defined to be

$$\rho(X,Y) = \frac{Cov[X,Y]}{\sqrt{Var[X]\,Var[Y]}}$$

with $-1 \leq \rho(X,Y) \leq 1$.

- The correlation coefficient is a measure of the degree of linearity between X and Y.
- $\rho(X,Y)$ close to ± 1 indicates high degree of linearity between X and Y.
- $\rho(X,Y) > 0$ indicates Y tends to increase when X does; we say X and Y are *positively correlated*.
- $\rho(X,Y) < 0$ indicates Y tends to decrease when X does; we say X and Y are *negatively correlated*.
- If $\rho(X,Y) = 0$ then X and Y are called *uncorrelated*.

Example

- Let I_A and I_B be indicator variables for events A and B, *viz.*

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

- In this case, we have

$$Cov[I_A, I_B] = E[I_A I_B] - E[I_A]E[I_B] = P[AB] - P[A]P[B] =$$

$$= P[A|B]P[B] - P[A]P[B] = P[B]\left(P[A|B] - P[A]\right)$$

- I_A and I_B are positively correlated when $P[A|B] > P[A]$.
- I_A and I_B are negatively correlated when $P[A|B] < P[A]$.
- I_A and I_B are uncorrelated when $P[A|B] = P[A]$.
- **Important:** The independence of X and Y always suggests that $Cov(X, Y) = \rho(X, Y) = 0$. The opposite direction, however, would generally be incorrect.

Another example

- Let $X_1, \ldots, X_n$ be iid with variance σ^2, and recall that

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is called the sample mean and $X_i - \bar{X}$ is called the i-th deviation. Show that

$$Cov[X_i - \bar{X}, \bar{X}] = 0$$

for each $i = 1, \ldots, n$.

- *Solution:*

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

$$= Cov[X_i, \frac{1}{n}\sum_{j=1}^{n} X_j] - Var[\bar{X}] = \frac{1}{n} Cov[X_i, \sum_{j=1}^{n} X_j] - \frac{\sigma^2}{n} =$$

$$= \frac{1}{n}\sum_{j=1}^{n} Cov[X_i, X_j] - \frac{\sigma^2}{n} = \frac{1}{n} Cov[X_i, X_i] - \frac{\sigma^2}{n} = \frac{1}{n}\sigma^2 - \frac{\sigma^2}{n} = 0$$

- For two discrete random variables X and Y with $P[Y = y > 0]$, we had

$$p_{X|Y}(x|y) = P[X = x|Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]} = \frac{p_{XY}(x, y)}{p_Y(y)}$$

- So, we can define the *conditional expectation* as

$$E[X|Y = y] = \sum_x xP[X = x|Y = y] = \sum_x xp_{X|Y}(x|y)$$

- Similarly, if X and Y are continuous, then provided $f_Y(y) > 0$, we have

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

and, consequently,

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

Example

- Suppose X and Y have joint pdf

$$f_{XY}(x,y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y}, & 0 < x < \infty,\ 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

Find $E[X|Y=y]$.

- *Solution:* For $x > 0$, we have

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{f_{XY}(x,y)}{\int_{-\infty}^{\infty} f_{XY}(x,y)dx} =$$

$$= \frac{\frac{1}{y}e^{-x/y}e^{-y}}{e^{-y}\int_0^{\infty}\frac{1}{y}e^{-x/y}dx} = \frac{\frac{1}{y}e^{-x/y}e^{-y}}{e^{-y}\cdot 1} = \frac{1}{y}e^{-x/y}$$

Therefore,

$$E[X|Y=y] = \int_0^{\infty}\frac{x}{y}e^{-x/y}dx = y$$

- Conditional expectations satisfy all the properties of ordinary expectation.

$$E[g(X)|Y=y] = \begin{cases} \sum_x g(x) p_{X|Y}(x|y), & \text{discrete case} \\ \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y), dx & \text{continuous case} \end{cases}$$

and

$$E\left[\sum_{i=1}^{n} X_i \mid Y=y\right] = \sum_{i=1}^{n} E[X_i|Y=y]$$

Computing Expectations by Conditioning

- $E[X|Y=y]$ is a function of y, say $g(y)$.
- Let $E[X|Y]$ be $g(Y)$ and, hence, in the last example we'd have had

$$E[X|Y=y]=y$$

and so $E[X|Y]=Y$.

- **Proposition:** $E[X]=E\left[E[X|Y]\right]$, namely

$$E[X]=\begin{cases}\sum_y E[X|Y=y]P[Y=y] & \text{[discrete case]}\\ \int_{-\infty}^{\infty} E[X|Y=y]f_Y(y)dy & \text{[continuous case]}\end{cases}$$

Indeed,

$$\int_{-\infty}^{\infty} E[X|Y=y]f_Y(y)dy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx\, f_Y(y)dy=$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y)f_Y(y)\,dxdy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)\,dxdy=$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)\,dydx=\int_{-\infty}^{\infty} x f_X(x)\,dx=E[X]$$

- You are in a room containing 3 doors. The first door exits the building after 3 minutes of travel. The second door returns you to where you are after 5 minutes of travel. The third door returns you to where you are after 7 minutes of travel. If each time you enter this room, you are equally likely to pick each of the 3 doors, what is the expected time until you leave the building?

- *Solution:* Let X be the time it takes to leave building and Y be the door choice.

$$E[X] = \sum_{k=1}^{3} E[X|Y=k]P[Y=k] = \frac{1}{3}\sum_{k=1}^{3} E[X|Y=k]$$

Also,

$$E[X|Y=1] = 3$$
$$E[X|Y=2] = 5 + E[X]$$
$$E[X|Y=1] = 7 + E[X]$$

Combining, we have

$$E[X] = \frac{1}{3}(3 + 5 + E[X] + 7 + E[X])$$

and, therefore, $E[X] = 15$.

- The number of people that enter a store in a day is a random variable with mean 50. The amounts spent by the persons that enters are iid with mean \$8. The amounts spent by the persons that enter are independent of the number of people that enter. What is the expected amount spent in the store in one day?

- *Solution:* Let $N = \#$ customers that enter store in one day. Let $X_i =$ amount spent by i-th customer. Therefore, the total amount spent is given by $Y = \sum_{i=1}^{N} X_i$.

$$E\left[\sum_{i=1}^{N} X_i\right] = E\left[E\left[\sum_{i=1}^{N} X_i \mid N\right]\right]$$

and

$$E\left[\sum_{i=1}^{N} X_i \mid N = n\right] = E\left[\sum_{i=1}^{n} X_i \mid N = n\right] =$$

$$= E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E\left[X_i\right] = nE[X_1] \quad \Longrightarrow \quad E\left[\sum_{i=1}^{N} X_i \mid N\right] = NE[X_1]$$

Hence,

$$E\left[\sum_{i=1}^{N} X_i\right] = E\left[NE[X_1]\right] = E[N]E[X_1] = 50 \cdot 8$$

Properties of conditional expectations

- Recall that X and Y are called *jointly Gaussian (normal)* or *bivariate Gaussian (normal)* with parameters μ_X, μ_Y, $\sigma_X > 0$, $\sigma_Y > 0$, and $-1 < \rho < 1$, when

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}$$

- We already know that

$$E[X] = \mu_X,\ E[Y] = \mu_Y,\ Var[X] = \sigma_X^2,\ Var[Y] = \sigma_Y^2$$

and, therefore,

$$\rho(X,Y) = \frac{Cov[X,Y]}{\sigma_X\sigma_Y} = \frac{E[XY] - \mu_X\mu_Y}{\sigma_X\sigma_Y}$$

- To determine $E[XY]$, recall that $f_{X|Y}(x|y)$ is a pdf for X where X has mean

$$\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$$

and the variance of X is $\sigma_X^2(1-\rho^2)$. So

$$E[X|Y=y] = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$$

- Now $E[XY] = E\left[E[XY|Y]\right]$ and

$$E[XY|Y=y] = E[Xy|Y=y] = yE[X|Y=y] =$$

$$= y\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right) = \mu_X y + \rho\frac{\sigma_X}{\sigma_Y}(y^2 - \mu_Y y)$$

implying

$$E[XY|Y] = \mu_X Y + \rho\frac{\sigma_X}{\sigma_Y}(Y^2 - \mu_Y Y)$$

Properties of conditional expectations (cont.)

- Consequently,

$$E[XY] = E\left[E[XY|Y]\right] = E\left[\mu_X Y + \rho\frac{\sigma_X}{\sigma_Y}(Y^2 - \mu_Y Y)\right] =$$

$$= \mu_X E[Y] + \rho\frac{\sigma_X}{\sigma_Y}(E[Y^2] - \mu_Y E[Y]) = \mu_X\mu_Y + \rho\frac{\sigma_X}{\sigma_Y}(E[Y^2] - \mu_Y^2) =$$

$$= \mu_X\mu_Y + \rho\frac{\sigma_X}{\sigma_Y}Var[Y] = \mu_X\mu_Y + \rho\frac{\sigma_X}{\sigma_Y}\sigma_Y^2 = \mu_X\mu_Y + \rho\sigma_X\sigma_Y$$

- Thus, finally, we obtain

$$\rho(X,Y) = \frac{E[XY] - \mu_X\mu_Y}{\sigma_X\sigma_Y} = \frac{\rho\sigma_X\sigma_Y}{\sigma_X\sigma_Y} = \rho$$

Computing Probabilities by Conditioning

- We can use conditioning to compute probabilities as well as expectations.
- Let A be an event, and

$$I_A = \begin{cases} 1, & A \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$$

- Then

$$E[I_A] = P[A]$$
$$E[I_A|Y = y] = P[A|Y = y]$$

and hence

$$P[A] = \begin{cases} \sum_y P[A|Y = y]P[Y = y], & Y \text{ is discrete} \\ \int_{-\infty}^{\infty} P[A|Y = y] f_Y(y) dy, & Y \text{ is continuous} \end{cases}$$

- Note that, if Y is discrete, then defining $B_i = \{Y = y_i\}$ yields

$$P[A] = \sum_i P[A|B_i]P[B_i]$$

where B_1, B_2, ... partition the sample space.

Example

- Say X and Y are independent random variables with densities $f_X(x)$ and $f_Y(y)$. Find $P[X < Y]$.

- *Solution:* One could always compute

$$P[X < Y] = \iint_{x<y} f_X(x) f_Y(y) dx dy$$

and then simplify. Alternatively,

$$P[X < Y] = \int_{-\infty}^{\infty} P[X < Y \mid Y = y] f_Y(y) dy =$$

$$= \int_{-\infty}^{\infty} P[X < y \mid Y = y] f_Y(y) dy = \int_{-\infty}^{\infty} P[X < y] f_Y(y) dy =$$

$$= \int_{-\infty}^{\infty} F_X(y) f_Y(y) dy$$

where

$$F_X(y) = \int_{-\infty}^{y} f_X(x) dx$$

- Say X and Y are independent random variables with densities $f_X(x)$ and $f_Y(y)$. Find the cdf and pdf of $X+Y$.

- *Solution:* Lets solve this by conditioning on Y.

$$P[X+Y \leq a] = \int_{-\infty}^{\infty} P[X+Y \leq a \mid Y=y] f_Y(y)dy =$$

$$= \int_{-\infty}^{\infty} P[X+y \leq a \mid Y=y] f_Y(y)dy = \int_{-\infty}^{\infty} P[X+y \leq a] f_Y(y)dy =$$

$$= \int_{-\infty}^{\infty} P[X \leq a-y] f_Y(y)dy = \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)dy$$

and, taking the derivatives, we obtain

$$f_{X+Y}(a) = \frac{d}{da} P[X+Y \leq a] = \frac{d}{da} \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)dy =$$

$$= \int_{-\infty}^{\infty} \frac{d}{da} F_X(a-y) f_Y(y)dy = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y)dy$$

- So far, we have defined expectation, conditional expectation and variance.

- **Definition:** The *conditional variance* of X given Y is defined as

$$Var[X|Y] = E[X^2|Y] - (\ E[X|Y]\ )^2$$

- In this case, we also have

$$E[\ Var[X|Y]\ ] = E\left[\ E[X^2|Y]\ \right] - E\left[\ (\ E[X|Y]\ )^2\ \right] =$$

$$= E[X^2] - E\left[\ (\ E[X|Y]\ )^2\ \right]$$

- Also, $E[X|Y] = g(Y)$ for some function g is a random variable, so

$$Var[\, g(Y)\,] = E\left[\,(\, g(Y)\,)^2\right] - (E[\, g(Y)\,])^2$$

$$Var[\, E[X|Y]\,] = E\left[\,(\, E[X|Y]\,)^2\right] - (E[\, E[X|Y]\,])^2 =$$

$$= E\left[\,(\, E[X|Y]\,)^2\right] - (E[X])^2$$

- Adding the above two results yields an important proposition.

- **Conditional Variance Formula:**

$$Var[X] = E[\, Var[X|Y]\,] + Var[\, E[X|Y]\,]$$

Example

- Let $X_1, X_2, \cdots$ be iid and let N be a non-negative random variable, independent of the X_i, $i = 1, 2, \ldots$ Let's compute $Var\left[\sum_{i=1}^{N} X_i\right]$ by conditioning on N.

- *Solution:* We know that

$$Var[X] = E[\, Var[X|Y]\,] + Var[\, E[X|Y]\,]$$

and we choose

$$X = \sum_{i=1}^{N} X_i, \quad Y = N$$

Then

$$E\left[\sum_{i=1}^{N} X_i \mid N = n\right] = E\left[\sum_{i=1}^{n} X_i \mid N = n\right] = E\left[\sum_{i=1}^{n} X_i\right] = nE[X_1]$$

(Since N is independent of the X_i) and hence

$$E\left[\sum_{i=1}^{N} X_i \mid N\right] = NE[X_1]$$

- For the same reason,

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

$$= Var\left[\sum_{i=1}^{n} X_i\right] = nVar\left[X_1\right]$$

and, therefore,

$$Var\left[\sum_{i=1}^{N} X_i \mid N\right] = NVar[X_1]$$

- Finally, by the conditional variance formula:

$$Var\left[\sum_{i=1}^{N} X_i\right] = E\left[\, NVar[X_1]\,\right] + Var\left[\, NE[X_1]\,\right] =$$

$$= E[N]\,Var[X_1] + (E[X_1])^2 Var[N]$$

- Say we observe a random variable $X = x$.
- We want to make an estimate $\hat{y} = g(x)$ of the outcome that the random variable Y will take.
- In this case $g(X)$ is called a **predictor (or estimator)** of Y.

- A common criterion to design $g(x)$ is to minimize the **mean squared error (MSE)** defined as

$$E[\ (Y - g(X))^2\ ]$$

- **Proposition:** The **minimum MSE (MMSE)** estimator $g(X)$ of Y is given by

$$g(x) = E[Y|X = x]$$

- To see that, we first note that

$$E[(Y-g(X))^2] = \int_{-\infty}^{\infty} E[(Y-g(X))^2 \mid X=x] f_X(x)dx =$$

$$= \int_{-\infty}^{\infty} E[(Y-g(x))^2 \mid X=x] f_X(x)dx =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y-g(x))^2 f_{Y|X}(y|x)dy \; f_X(x)dx$$

- Now, let's minimize the inner integral for each x, i.e., find $g(x)$ that minimizes

$$\int_{-\infty}^{\infty} (y-g(x))^2 f_{Y|X}(y|x)dy$$

- To do this, take the derivative with respect to $g(x)$ and set to 0:

$$0 = \frac{d}{dg(x)} \int_{-\infty}^{\infty} (y-g(x))^2 f_{Y|X}(y|x)dy =$$

$$= \int_{-\infty}^{\infty} 2(g(x)-y) f_{Y|X}(y|x)dy = 2g(x) \int_{-\infty}^{\infty} f_{Y|X}(y|x)dy -$$

$$-2 \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy = 2g(x) - 2E[Y|X=x]$$

- Two random variables X and Y are jointly Gaussian with parameters μ_X, μ_Y, $\sigma_X > 0$, $\sigma_Y > 0$, and $-1 < \rho < 1$. What is the minimum mean squared error estimate of Y given that $X = x$?
- *Solution:*

$$g(x) = E[Y|X = x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

- **Definition:** The *moment generating function* $M_X(t)$ of a random variable X is defined as

$$M_X(t) = E[e^{tX}]$$

- **Note:** there is also a closely related concept called *characteristic function* defined as

$$\phi_X(t) = E[e^{\imath tX}], \quad \text{with } \imath = \sqrt{-1}$$

- $M_X(t)$ is called moment generating function because we can find the *moments* $E[X^n]$ from it easily:

$$M_X^{(1)}(t) = \frac{dM_X(t)}{dt} = E\left[\frac{d}{dt}e^{tX}\right] = E\left[Xe^{tX}\right]$$

$$M_X^{(n)}(t) = E\left[X^n e^{tX}\right]$$

and, hence,

$$M_X^{(1)}(0) = E[X], \quad M_X^{(n)}(0) = E[X^n]$$

- Find $M_X(t)$ if $X \sim \text{Binomial}(n,p)$ and use this to find $E[X]$, $E[X^2]$, and $Var[X]$.

- *Solution:*

$$M_X(t) = E[e^{tX}] = \sum_{k=0}^{n} e^{tk} p_X(k) =$$

$$= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k \, (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k \, (1-p)^{n-k} = (pe^t + 1 - p)^n$$

$$M_X^{(1)}(t) = n(pe^t + 1 - p)^{n-1} pe^t$$

$$M_X^{(2)}(t) = n(n-1)(pe^t + 1 - p)^{n-2}(pe^t)^2 + n(pe^t + 1 - p)^{n-1} pe^t$$

Hence,

$$E[X] = M_X^{(1)}(0) = np, \quad E[X^2] = M_X^{(2)}(0) = n(n-1)p^2 + np$$

and, therefore,

$$Var[X] = E[X^2] - (E[X])^2 = np(1-p)$$

- Find $M_X(t)$ if $X \sim \text{Poisson}(\lambda)$ and use this to find $E[X]$, $E[X^2]$, and $Var[X]$.

- *Solution:*

$$M_X(t) = E[e^{tX}] = \sum_{n=0}^{\infty} e^{tn} p_X(n) = \sum_{n=0}^{\infty} e^{tn} \frac{\lambda^n}{n!} e^{-\lambda} =$$

$$= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = e^{-\lambda} \exp(\lambda e^t) = \exp(\ \lambda(e^t - 1)\ )$$

Consequently,

$$M_X^{(1)}(t) = \lambda e^t \exp(\ \lambda(e^t - 1)\ )$$

$$M_X^{(2)}(t) = (\lambda e^t)^2 \exp(\ \lambda(e^t - 1)\ ) + \lambda e^t \exp(\ \lambda(e^t - 1)\ )$$

and, hence,

$$E[X] = M_X^{(1)}(0) = \lambda, \quad E[X^2] = M_X'^{(2)}(0) = \lambda^2 + \lambda$$

finally yielding

$$Var[X] = E[X^2] - (E[X])^2 = \lambda$$

Yet another example

- Find $M_X(t)$ if $X \sim \text{Exponential}(\lambda)$ and use this to find $E[X]$, $E[X^2]$, and $Var[X]$.

- *Solution:*

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx =$$

$$= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{-(\lambda - t)x} dx = \frac{\lambda}{\lambda - t}$$

for $t < \lambda$. Thus,

$$M_X^{(1)}(t) = \frac{\lambda}{(\lambda - t)^2}, \quad M_X^{(2)}(t) = \frac{2\lambda}{(\lambda - t)^3}$$

and, therefore,

$$E[X] = M_X^{(1)}(0) = 1/\lambda, \quad E[X^2] = M_X^{(2)}(0) = 2/\lambda^2$$

yielding

$$Var[X] = E[X^2] - (E[X])^2 = 1/\lambda^2$$

- Find $M_X(t)$ if $X \sim \mathcal{N}(\mu, \sigma^2)$ and use this to find $E[X]$, $E[X^2]$, and $Var[X]$.

- *Solution:* Let $Z \sim \mathcal{N}(0,1)$. We first compute $M_Z(t)$:

$$M_Z(t) = E[e^{tZ}] = \int_{-\infty}^{\infty} e^{tz} f_Z(z) dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz} e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{z^2 - 2zt}{2}\right) dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(z-t)^2}{2} + \frac{t^2}{2}\right) dz =$$

$$= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(z-t)^2}{2}\right) dz = e^{t^2/2}$$

- Since $X = \mu + \sigma Z$:

$$M_X(t) = E[e^{tX}] = E[e^{t(\mu+\sigma Z)}] =$$

$$= E[e^{t\mu} e^{t\sigma Z}] = e^{t\mu} E[e^{t\sigma Z}] = e^{t\mu} M_Z(t\sigma) =$$

$$= e^{t\mu} e^{\frac{t^2\sigma^2}{2}} = \exp\left(\frac{t^2\sigma^2}{2} + \mu t\right)$$

and, thus, we have

$$M_X^{(1)}(t) = (\mu + t\sigma^2) \exp\left(\frac{t^2\sigma^2}{2} + \mu t\right)$$

$$M_X^{(2)}(t) = (\mu + t\sigma^2)^2 \exp\left(\frac{t^2\sigma^2}{2} + \mu t\right) + \sigma^2 \exp\left(\frac{t^2\sigma^2}{2} + \mu t\right)$$

yielding

$$E[X] = M_X^{(1)}(0) = \mu, \quad E[X^2] = M_X^{(2)}(0) = \mu^2 + \sigma^2$$

and, finally,

$$Var[X] = E[X^2] - (E[X])^2 = \sigma^2$$

- Let X and Y be *independent* random variables. Then

$$M_{X+Y}(t) = E\left[e^{t(X+Y)}\right] = E\left[e^{tX}e^{tY}\right] =$$

$$= E\left[e^{tX}\right]E\left[e^{tY}\right] = M_X(t)M_Y(t)$$

- Another useful fact is that *the distribution of X is uniquely determined by $M_X(t)$*. (The textbook has tables of MGF for various distributions.)

Example

- Let $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$ be independent. What is the distribution of $X + Y$?

- *Solution:*

$$M_{X+Y}(t) = M_X(t) M_Y(t) =$$

$$= \exp(\ \lambda_1(e^t - 1)\ ) \exp(\ \lambda_2(e^t - 1)\ ) = \exp(\ (\lambda_1 + \lambda_2)(e^t - 1)\ )$$

and, therefore, $X + Y$ is $\text{Poisson}(\lambda_1 + \lambda_2)$.

Another example

- Let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ be independent. What is the distribution of $X + Y$?

- *Solution:*

$$M_{X+Y}(t) = M_X(t)M_Y(t) = \exp\left(\frac{t^2\sigma_X^2}{2} + \mu_X t\right) =$$

$$= \exp\left(\frac{t^2(\sigma_X^2 + \sigma_Y^2)}{2} + (\mu_X + \mu_Y)t\right)$$

So $X + Y$ is $\sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

- **Definition:** For random variables $X_1, X_2, \ldots, X_n$, the *joint moment generating function* is defined as

$$M(t_1, \ldots, t_n) = E\left[e^{t_1 X_1 + t_2 X_2 + \cdots + t_n X_n}\right]$$

with

$$M_{X_i}(t) = E[e^{tX_i}] = M(0, 0, \ldots, t, 0, \ldots, 0)$$

- The joint MGF uniquely determines the joint pdf.
- If $X_1, \ldots, X_n$ are *independent* then

$$M(t_1, t_2, \ldots, t_n) = E\left[e^{t_1 X_1 + t_2 X_2 + \cdots + t_n X_n}\right] =$$

$$= E\left[e^{t_1 X_1}\right] E\left[e^{t_2 X_2}\right] \cdots E\left[e^{t_n X_n}\right] = M_{X_1}(t_1) M_{X_2}(t_2) \cdots M_{X_n}(t_n)$$

- Since the joint MGF uniquely specifies the joint distribution, then the mutual independence of $X_1, \ldots, X_n$ is equivalent to

$$M(t_1, t_2, \ldots, t_n) = M_{X_1}(t_1) M_{X_2}(t_2) \cdots M_{X_n}(t_n)$$

Example

- If X and Y are independent and $\sim \mathcal{N}(\mu, \sigma^2)$, show that $X+Y$ and $X-Y$ are independent.

- *Solution:*

$$E\left[e^{t(X+Y)+s(X-Y)}\right] = E\left[e^{(t+s)X+(t-s)Y}\right] = E\left[e^{(t+s)X}\right] E\left[e^{(t-s)Y}\right] =$$

$$= e^{\mu(t+s)+\sigma^2(t+s)^2/2} e^{\mu(t-s)+\sigma^2(t-s)^2/2} = e^{2\mu t+\sigma^2 t^2} e^{\sigma^2 s^2}$$

Consequently,

1. The 1st term is the MGF (in t) for a normal with mean 2μ and variance $2\sigma^2$.

2. The 2nd term is the MGF (in s) for a normal with mean 0 and variance $2\sigma^2$.

- Let $Z_1, Z_2, \ldots, Z_n$ be n independent standard normal random variables.
- Let $\mu_1, \mu_2, \ldots, \mu_m$ be m constants.
- Define $X_1, X_2, \ldots, X_m$ by

$$\begin{aligned} X_1 &= a_{11}Z_1 + \cdots + a_{1n}Z_n + \mu_1 \\ X_2 &= a_{21}Z_1 + \cdots + a_{2n}Z_n + \mu_2 \\ \vdots \quad & \quad \vdots \qquad\qquad\quad \vdots \qquad\quad \vdots \\ X_m &= a_{m1}Z_1 + \cdots + a_{mn}Z_n + \mu_m \end{aligned}$$

- Then we say that $X_1, \ldots, X_m$ are **multivariate normal (Gaussian)**.

- We can write this compactly in matrix form as

$$\boldsymbol{X} = A\boldsymbol{Z} + \boldsymbol{\mu}$$

where the matrix A is $m \times n$ and has a_{ij} as its (i, j) entry, and

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix}, \qquad \boldsymbol{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix}.$$

- Now, let B be a $k \times m$ matrix, and $\boldsymbol{\nu}$ a column vector of length k. Then

$$\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{\nu} = (BA)\boldsymbol{Z} + (B\boldsymbol{\mu} + \boldsymbol{\nu})$$

and, thus, $\boldsymbol{Y}$ is multivariate Gaussian too.

- To summarize, *an affine transformation of a multivariate Gaussian is a multivariate Gaussian again*.

- Since each X_i is a sum of independent Gaussian random variables, each X_i is Gaussian with

$$E[X_i] = E[a_{i1}Z_1 + \ldots + a_{in}Z_n + \mu_i] = \mu_i$$

and

$$Var[X_i] = Var[a_{i1}Z_1 + \ldots + a_{in}Z_n + \mu_i] = Var[a_{i1}Z_1 + \ldots + a_{in}Z_n] =$$

$$= a_{i1}^2 Var[Z_1] + \ldots + a_{in}^2 Var[Z_n] = a_{i1}^2 + \ldots + a_{in}^2$$

- Moreover,

$$M(t_1, \ldots, t_m) = E\left[e^{t_1X_1 + t_2X_2 + \ldots + t_mX_m}\right] = E\left[e^U\right]$$

where

$$U = t_1X_1 + t_2X_2 + \ldots + t_mX_m$$

- Now, U is a linear combination of $Z_1, \ldots, Z_n$, so U is Gaussian. Also,

$$E[U] = E[t_1X_1 + \cdots + t_mX_m] = \underbrace{t_1\mu_1 + \cdots + t_m\mu_m}_{\mu}$$

$$Var[U] = Cov\left[\sum_{i=1}^{m} t_iX_i, \sum_{j=1}^{m} t_jX_j\right] = \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{m} t_it_jCov\left[X_i, X_j\right]}_{\sigma^2}$$

- Since $U \sim \mathcal{N}(\mu, \sigma^2)$, we have

$$M(t_1, \ldots, t_n) = E[e^U] = M_U(t)|_{t=1} = \exp\left(\mu + \sigma^2/2\right) =$$

$$= \exp\left(\sum_{i=1}^{m} t_i\mu_i + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} t_it_jCov\left[X_i, X_j\right]\right)$$

- We see that the joint MGF of $X_1, \ldots, X_m$ depends only on the means $\mu_1, \ldots, \mu_m$ covariances $Cov\,[X_i, X_j]$.

- Since the MGF uniquely determines the joint distribution, then the joint distribution of a multivariate Gaussian (normal) depends only the means $E[X_i]$ and covariances $Cov[X_i, X_j]$.