Lecture Notes #3

A Quick Review of Probability & Statistics

(1) Normal Distribution

Q. Who invented the normal distribution?

* Right: Johann Carl Friedrich Gauss (30 April 1777 – 23 February 1855)

<i>Probability Density Function (p.d.f.)

\[ X \sim N(\mu, \sigma^2) \]: X follows normal distribution of mean \( \mu \) and variance \( \sigma^2 \)
\[ f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty, x \in R \]

\[ P(a \leq X \leq b) = \int_a^b f(x)dx = \text{area under the pdf curve bounded by } a \text{ and } b \]

<ii> Cumulative Distribution Function (c.d.f.)

\[ F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt \]

\[ f(x) = \frac{d}{dx} F(x) \]

(2). Mathematical Expectation (Review).

Continuous random variable: \[ E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \]

Discrete random variable: \[ E[g(X)] = \sum_{x\in\Omega} g(x)P(X = x) \]
Properties of Expectations:

(1) \( E(c) = c \), where \( c \) is a constant

(2) \( E[c \cdot g(X)] = c \cdot E[g(X)] \), where \( c \) is a constant

(3) \( E[g(X) + h(Y)] = E[g(X)] + E[h(Y)] \), for any \( X \& Y \)

(4) \( E[g(X) \cdot h(Y)] = E[g(X)] \cdot E[h(Y)] \), if \( X \& Y \) are independent – otherwise it is usually not true.

Special case:

1) (population) Mean: \( \mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx \)

Note: \( E(aX+b) = aE(X)+b \), where \( a \& b \) are constants

2) (population) Variance:

\[ \text{Var}(X) = \sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) \, dx = E(X^2) - [E(X)]^2 \]

Note: \( \text{Var}(aX+b) = a^2 \text{Var}(X) \), where \( a \& b \) are constants

3) Moment generating function:

\[ M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x) \, dx \text{, when } X \text{ is continuous.} \]

\[ M_X(t) = E(e^{tX}) = \sum \text{all possible values of } x \cdot e^{tx} \cdot f(x) \]

For normal distribution, \( X \sim N(\mu, \sigma^2) \), \( f(x) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \), \( -\infty < x < \infty \)

\[ M_X(t) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x) \, dx = e^{t\mu + \frac{1}{2}t^2\sigma^2} \]

4) Moment:

1\(^{\text{st}}\) (population) moment: \( E(X) = \int x \cdot f(x) \, dx \)

2\(^{\text{nd}}\) (population) moment: \( E(X^2) = \int x^2 \cdot f(x) \, dx \)

...
$K^{th}$ (population) moment: $E(X^k) = \int x^k \cdot f(x) \, dx$

**Theorem.** If $X_1, X_2$ are independent, then

$$f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$$

**Theorem.** If $X_1, X_2$ are independent, then

$$M_{X_1, X_2}(t) = M_{X_1}(t) M_{X_2}(t)$$

**Theorem.** Under regularity conditions, there is 1-1 correspondence between the pdf and the mgf of a given random variable $X$. That is,

$$pdf \ f(x) \leftrightarrow mgf \ M_X(t).$$

**E.g.** Sgt. Jones wishes to select one army recruit into his unit. It is known that the IQ distribution for all recruits is normal with mean 180 and standard deviation 10. What is the chance that Sgt. Jones would select a recruit with an IQ of at least 200?

**Sol.)** Let $X$ represents the IQ of a randomly selected recruit.

$$X \sim N(\mu = 180, \sigma = 10)$$

$$P(X \geq 200) = P\left(\frac{X - 180}{10} \geq \frac{200 - 180}{10}\right) = P(Z \geq 2) = 2.28\%$$

**E.g.** Sgt. Jones wishes to select three army recruits into his unit. It is known that the IQ distribution for all recruits is normal with mean 180 and standard deviation 10. What is the chance that Sgt. Jones would select three recruits with an average IQ of at least 200?

**Sol.)** Let $X_1, X_2, X_3$ represents the IQ of three randomly selected recruits.

$$X_i \overset{i.i.d.}{\sim} N(\mu, \sigma^2), \quad i = 1, 2, 3$$

$$\bar{X} \sim N(\mu^* = 180, \sigma^* = \frac{10}{\sqrt{3}})$$
\[ P(X \geq 200) = P\left( \frac{X - 180}{10/\sqrt{3}} \geq \frac{200 - 180}{10/\sqrt{3}} \right) = P(Z \geq 2\sqrt{3}) = 0.03\% \]

e.g. (Another practical example of mathematical expectation) A gambler goes to bet. The dealer has 3 dice, which are fair, meaning that the chance that each face shows up is exactly 1/6.

The dealer says: "You can choose your bet on a number, any number from 1 to 6. Then I'll roll the 3 dice. If none show the number you bet, you'll lose $1. If one shows the number you bet, you'll win $1. If two or three dice show the number you bet, you'll win $3 or $5, respectively."

Is it a fair game? (*A fair game is such that the expected, or equivalently the long term average, winning is zero.)

Solution: Let \( X \) be # of dice show the number you bet, then \( X \sim Binomial(3, \frac{1}{6}) \) (Because the dealer roll 3 dice independently, and the probability for each dice show the number you bet is \( \frac{1}{6} \)). That is,

\[ P(X = x) = \binom{3}{x} \left( \frac{1}{6} \right)^x \left( \frac{5}{6} \right)^{3-x}, \quad x = 0, 1, 2, 3 \]

Let \( Y \) be the amount of money you get, then,

\[ P(Y = -1) = P(X = 0) = \binom{3}{0} \left( \frac{1}{6} \right)^0 \left( \frac{5}{6} \right)^{3-0} = \left( \frac{5}{6} \right)^3 = \frac{125}{216} \]

\[ P(Y = 1) = P(X = 1) = \binom{3}{1} \left( \frac{1}{6} \right)^1 \left( \frac{5}{6} \right)^{3-1} = 3 \cdot \frac{1}{6} \cdot \left( \frac{5}{6} \right)^2 = \frac{75}{216} \]

\[ P(Y = 3) = P(X = 2) = \binom{3}{2} \left( \frac{1}{6} \right)^2 \left( \frac{5}{6} \right)^{3-2} = 3 \cdot \left( \frac{1}{6} \right)^2 \cdot \frac{5}{6} = \frac{15}{216} \]

\[ P(Y = 5) = P(X = 3) = \binom{3}{3} \left( \frac{1}{6} \right)^3 \left( \frac{5}{6} \right)^{3-3} = \left( \frac{1}{6} \right)^3 = \frac{1}{216} \]

Thus, your expected winning is
\[
E(Y) = \sum_y y \cdot P(Y = y) = (-1) \cdot \frac{125}{216} + 1 \cdot \frac{75}{216} + 3 \cdot \frac{15}{216} + 5 \cdot \frac{1}{216} = 0
\]

Therefore, it is a fair game.

*Definition. Joint distribution, and independence*

Definition. The joint cdf of two random variables X and Y are defined as:

\[ F_{X,Y}(x, y) = F(x, y) = P(X \leq x, Y \leq y) \]

Definition. The joint pdf of two discrete random variables X and Y are defined as:

\[ f_{X,Y}(x, y) = f(x, y) = P(X = x, Y = y) \]

Definition. The joint pdf of two continuous random variables X and Y are defined as:

\[ f_{X,Y}(x, y) = f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) \]

Definition. The marginal pdf of the discrete random variable X or Y can be obtained by summation of their joint pdf as the following:

\[ f_X(x) = \sum_y f(x, y) ; f_Y(y) = \sum_x f(x, y) ; \]

Definition. The marginal pdf of the continuous random variable X or Y can be obtained by integration of the joint pdf as the following:

\[ f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy; \]
\[ f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx; \]

Definition. The conditional pdf of a random variable X or Y is defined as:
Definition. The joint moment generating function of two random variables $X$ and $Y$ is defined as

$$M_{X,Y}(t_1, t_2) = E(e^{t_1 X + t_2 Y})$$

Note that we can obtain the marginal mgf for $X$ or $Y$ as follows:

$$M_X(t_1) = M_{X,Y}(t_1, 0) = E(e^{t_1 X + 0 Y}) = E(e^{t_1 X}); \quad M_Y(t_2) = M_{X,Y}(0, t_2) = E(e^{0 X + t_2 Y})$$

Theorem. Two random variables $X$ and $Y$ are independent $\iff$ (if and only if)

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \iff f_{X,Y}(x, y) = f_X(x)f_Y(y) \iff M_{X,Y}(t_1, t_2) = M_X(t_1)M_Y(t_2)$$

Definition. The covariance of two random variables $X$ and $Y$ is defined as

$$COV(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

Theorem. If two random variables $X$ and $Y$ are independent, then we have

$$COV(X, Y) = 0.$$

(*Note: However, $COV(X, Y) = 0$ does not necessarily mean that $X$ and $Y$ are independent.)

Exercise

Q1. Let $X_1 \& X_2$ be independent $N(\mu, \sigma^2)$. Prove that $X_1 + X_2$ and $X_1 - X_2$ are independent in two approaches: (1) pdf, (2) mgf

Solution:

(1) The pdf approach: Let

$$W = X_1 + X_2$$

$$Z = X_1 - X_2$$

Then we have
This is a 1-1 transformation between \((X_1, X_2)\) and \((W, Z)\).

Define the Jacobian \(J\) of the transformation as:

\[
J = \begin{vmatrix}
\frac{\partial x_1}{\partial w} & \frac{\partial x_1}{\partial z} \\
\frac{\partial x_2}{\partial w} & \frac{\partial x_2}{\partial z}
\end{vmatrix} = \begin{vmatrix}\frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & \frac{1}{2}
\end{vmatrix} = -\frac{1}{2}
\]

Then the joint distribution of the new random variables is given by:

\[
f_{W,Z}(w,z) = f_{X_1,X_2}\left(\frac{w+z}{2}, \frac{w-z}{2}\right) |J|
\]

Since \(X_1\) and \(X_2\) are independent, we have:

\[
f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)
\]

Thus we find:

\[
f_{W,Z}(w,z) = f_{X_1}\left(\frac{w}{2}\right)f_{X_2}\left(\frac{w-z}{2}\right) |J|
\]

\[
= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^2 e^{-\frac{1}{2\sigma^2}(w^2-\mu^2)} \cdot \frac{1}{2\sigma^2}(w^2-\mu^2)^2 \cdot \frac{1}{2}
\]

\[
= \frac{1}{2} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^2 e^{-\frac{1}{4\sigma^2}(w-\mu)^2} \cdot \frac{1}{4\sigma^2}(w-\mu)^2
\]

\[
= \frac{1}{\sqrt{2\pi \times 2\sigma^2}} e^{-\frac{1}{4\sigma^2}(w-\mu)^2} \frac{1}{\sqrt{2\pi \times 2\sigma^2}} e^{-\frac{1}{4\sigma^2}(z)^2}
\]

\[
= f_Z(z)f_W(w)
\]
So Z and W are independent. Furthermore we see that \(X_1 + X_2 \sim N(2\mu, 2\sigma^2)\) and \(X_1 - X_2 \sim N(0, 2\sigma^2)\)

(2) The mgf approach: Now we prove that

\[
M_{X_1+X_2,X_1-X_2}(t_1, t_2) = M_{X_1+X_2}(t_1)M_{X_1-X_2}(t_2)
\]

**Solution:**

Let \(W = X_1 + X_2\) and \(Z = X_1 - X_2\)

* **W and Z are independent if and only if** \(M_{W,Z}(t_1, t_2) = M_{W}(t_1)M_{Z}(t_2)\)

\[
M_{W,Z}(t_1, t_2) = E(e^{t_1W+t_2Z}) \quad \text{Substitute for W & Z}
\]

\[
= E(e^{t_1(X_1+X_2)+t_2(X_1-X_2)})
\]

\[
E(e^{(t_1+t_2)X_1+(t_1-t_2)X_2})
\]

\[
= E(e^{(t_1+t_2)X_1})E(e^{(t_1-t_2)X_2}) \quad \text{because X_1 & X_2 are independent}
\]

* \(E(e^{(t_1+t_2)X_1}) = e^{\mu t_1+\mu t_2+\frac{\sigma^2 t_1^2}{2}+\sigma^2 t_2 t_2+\frac{\sigma^2 t_2^2}{2}}\)

* \(E(e^{(t_1-t_2)X_2}) = e^{\mu t_1-\mu t_2+\frac{\sigma^2 t_1^2}{2}-\sigma^2 t_2 t_2+\frac{\sigma^2 t_2^2}{2}}\)

\[
\therefore E(e^{(t_1+t_2)X_1+(t_1-t_2)X_2}) = e^{\mu t_1+\mu t_2+\frac{\sigma^2 t_1^2}{2}+\sigma^2 t_2 t_2+\frac{\sigma^2 t_2^2}{2}}e^{\mu t_1-\mu t_2+\frac{\sigma^2 t_1^2}{2}-\sigma^2 t_2 t_2+\frac{\sigma^2 t_2^2}{2}}
\]

\[
= \exp\left(\mu t_1 + \mu t_2 + \frac{\sigma^2 t_1^2}{2} + \sigma^2 t_2 t_2 + \frac{\sigma^2 t_2^2}{2} + \mu t_1 - \mu t_2 + \frac{\sigma^2 t_1^2}{2} - \sigma^2 t_1 t_2 + \frac{\sigma^2 t_2^2}{2}\right)
\]

\[
= \exp\left(2\mu t_1 + \frac{1}{2}(2\sigma^2)t_1^2 + \frac{1}{2}(2\sigma^2)t_2^2\right)
\]

\[
e^{2\mu t_1+\frac{1}{2}(2\sigma^2)t_1^2}e^{\sigma^2 t_2+\frac{1}{2}(2\sigma^2)t_2^2} = M_{X_1+X_2}(t_1)M_{X_1-X_2}(t_2)
\]

\[
= M_{W}(t_1)M_{Z}(t_2)
\]

\[
\therefore M_{W,Z}(t_1, t_2) = M_{W}(t_1)M_{Z}(t_2)
\]
Q2. Let \( Z \) be \( N(0,1) \), (1) please derive the covariance between \( Z \) and \( Z^2 \); (2) Are \( Z \) and \( Z^2 \) independent?

**Solution:**

(1) \( \text{cov}(Z, Z^2) = E \left[ (Z - E(Z)) \left( Z^2 - E(Z^2) \right) \right] = E[(Z - 0)(Z^2 - 1)] = E[Z^3] = \frac{d^3}{dt^3}[\exp(t^2)]\big|_{t=0} = 0 \)

(2) \( Z \) and \( Z^2 \) are not independent. You can do it in many ways, using (a) pdf, (b) cdf/probabilities, or (c) mgf.

For example, \( P(Z > 1, Z^2 > 1) \neq P(Z > 1) \ast P(Z^2 > 1) = [P(Z > 1)]^2 \)

Q3. Let \( \sim N(3,4) \), please calculate \( P(1 < X < 3) \)

**Solution:**

\[
P(1 < X < 3) = P \left( \frac{1 - 3}{2} < \frac{X - 3}{2} < \frac{3 - 3}{2} \right) = P(-1 < Z < 0) = P(0 < Z < 1) = 0.3413
\]

*Note: In the above, \( Z \sim N(0,1) \)*

Q4. Students A and B plans to meet at the SAC Seawolf market between 12noon and 1pm tomorrow. The one who arrives first will wait for the other for 30 minutes, and then leave. What is the chance that the two friends will be able to meet at SAC
during their appointed time period assuming that each one will arrive at SAC independently, each at a random time between 12noon and 1pm?

**Solution:** 3/4 as illustrated by the following figure.

Let X and Y denote the arrival time (in hour where 0 indicates 12noon and 1 indicates 1pm) of A and B respectively. Our assumption implies that X and Y each follows Uniform[0,1] distribution, and furthermore, they are independent. Thus the joint pdf of X and Y follows the uniform distribution with f(x,y) = 1, if (x,y) is in the square region, and f(x,y) = 0, outside the square region. A and B will be able to meet iff |X − Y| ≤ 1/2 as represented by the red area M bounded by the lines: X − Y = −1/2 and X − Y = +1/2.

That is:

\[ P(\text{A & B will meet}) = P\left(|X − Y| ≤ \frac{1}{2}\right) = \int_{-\frac{1}{2}}^{\frac{1}{2}} 1 \, dx \, dy = 1 - \frac{1}{8} - \frac{1}{8} = \frac{3}{4} \]

**Definitions:** population correlation & sample correlation

Definition: The population correlation coefficient ρ is defined as:

\[ \rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}} \]

Definition: Let \((X_1, Y_1), ..., (X_n, Y_n)\) be a random sample from a given bivariate population, then the sample correlation coefficient r is defined as:
\[ r = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{[\Sigma(X_i - \bar{X})^2][\Sigma(Y_i - \bar{Y})^2]}} \]

**Definition:** Bivariate Normal Random Variable

\((X, Y) \sim BN(\mu_X, \sigma_X^2; \mu_Y, \sigma_Y^2; \rho)\) where \(\rho\) is the correlation between \(X\) & \(Y\)

The joint p.d.f. of \((X, Y)\) is

\[
f_{X,Y}(x, y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1 - \rho^2}} \exp\left\{ -\frac{1}{2(1 - \rho^2)} \left[ \left( \frac{x - \mu_X}{\sigma_X} \right)^2 - 2\rho \left( \frac{x - \mu_X}{\sigma_X} \right) \left( \frac{y - \mu_Y}{\sigma_Y} \right) + \left( \frac{y - \mu_Y}{\sigma_Y} \right)^2 \right] \right\}
\]

Exercise: Please derive the mgf of the bivariate normal distribution.

**Q5.** Let \(X\) and \(Y\) be random variables with joint pdf

\[
f_{X,Y}(x, y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1 - \rho^2}} \exp\left\{ -\frac{1}{2(1 - \rho^2)} \left[ \left( \frac{x - \mu_X}{\sigma_X} \right)^2 - 2\rho \left( \frac{x - \mu_X}{\sigma_X} \right) \left( \frac{y - \mu_Y}{\sigma_Y} \right) + \left( \frac{y - \mu_Y}{\sigma_Y} \right)^2 \right] \right\}
\]

Where \(-\infty < x < \infty, -\infty < y < \infty\). Then \(X\) and \(Y\) are said to have the bivariate normal distribution. The joint moment generating function for \(X\) and \(Y\) is

\[
M(t_1, t_2) = \exp\left\{ t_1\mu_X + t_2\mu_Y + \frac{1}{2} (t_1^2\sigma_X^2 + 2\rho t_1 t_2 \sigma_X \sigma_Y + t_2^2\sigma_Y^2) \right\}
\]

(a) Find the marginal pdf’s of \(X\) and \(Y\);

(b) Prove that \(X\) and \(Y\) are independent if and only if \(\rho = 0\).

(Here \(\rho\) is indeed, the <population> correlation coefficient between \(X\) and \(Y\).)

(c) Find the distribution of \((X + Y)\).
(d) Find the conditional pdf of $f(x|y)$, and $f(y|x)$

**Solution:**

(a)  

The moment generating function of $X$ can be given by 

$$M_X(t) = M(t, 0) = \exp \left[ \mu_X t + \frac{1}{2} \sigma_X^2 t^2 \right].$$

Similarly, the moment generating function of $Y$ can be given by 

$$M_Y(t) = M(t, 0) = \exp \left[ \mu_Y t + \frac{1}{2} \sigma_Y^2 t^2 \right].$$

Thus, $X$ and $Y$ are both marginally normal distributed, i.e.,

$$X \sim N(\mu_X, \sigma_X^2), \text{ and } Y \sim N(\mu_Y, \sigma_Y^2).$$

The pdf of $X$ is 

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X}} \exp \left[ -\frac{(x - \mu_X)^2}{2\sigma_X^2} \right].$$

The pdf of $Y$ is 

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y}} \exp \left[ -\frac{(y - \mu_Y)^2}{2\sigma_Y^2} \right].$$

(b) 

If $\rho = 0$, then 

$$M(t_1, t_2) = \exp \left[ \mu_X t_1 + \mu_Y t_2 + \frac{1}{2} (\sigma_X^2 t_1^2 + \sigma_Y^2 t_2^2) \right] = M(t_1, 0) \cdot M(0, t_2)$$
Therefore, X and Y are independent.

If X and Y are independent, then

\[ M(t_1, t_2) = M(t_1, 0) \cdot M(0, t_2) = \exp \left[ \mu_X t_1 + \mu_Y t_2 + \frac{1}{2} (\sigma_X^2 t_1^2 + \sigma_Y^2 t_2^2) \right] \]

Therefore, \( \rho = 0 \)

(c)

\[ M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX+ty}] \]

Recall that \( M(t_1, t_2) = E[e^{t_1X+t_2Y}] \), therefore we can obtain \( M_{X+Y}(t) \) by \( t_1 = t_2 = t \) in \( M(t_1, t_2) \)

That is,

\[ M_{X+Y}(t) = M(t, t) = \exp \left[ (\mu_X + \mu_Y)t + \frac{1}{2} (\sigma_X^2 t^2 + 2\rho\sigma_X\sigma_Y t^2 + \sigma_Y^2 t^2) \right] \]

\[ \therefore X + Y \sim N(\mu = \mu_X + \mu_Y, \sigma^2 = \sigma_X^2 + 2\rho\sigma_X\sigma_Y + \sigma_Y^2) \]

(d)

The conditional distribution of X given Y=y is given by

\[ f(x|y) = \frac{f(x, y)}{f(y)} = \frac{1}{\sqrt{2\pi\sigma_X\sqrt{1-\rho^2}}} \exp \left\{ - \frac{(x - \mu_X - \frac{\sigma_X}{\sigma_Y}(y - \mu_Y))^2}{2(1-\rho^2)\sigma_X^2} \right\}. \]
Similarly, we have the conditional distribution of $Y$ given $X=x$ is

\[
f(y|x) = \frac{f(x,y)}{f(x)} = \frac{1}{\sqrt{2\pi \sigma_y \sqrt{1 - \rho^2}}} \exp \left\{-\frac{\left(\frac{y - \mu_y}{\sigma_y} - \frac{\sigma_y}{\sigma_x} \rho (x - \mu_x)\right)^2}{2(1 - \rho^2) \sigma_y^2}\right\}.
\]

Therefore:

\[
X|Y = y \sim N\left(\mu_x + \rho \frac{\sigma_x}{\sigma_y} (y - \mu_y), (1 - \rho^2) \sigma_x^2\right)
\]

\[
Y|X = x \sim N\left(\mu_y + \rho \frac{\sigma_y}{\sigma_x} (x - \mu_x), (1 - \rho^2) \sigma_y^2\right)
\]