1179: Probability
Lecture 23 — MGFs, Covariance, and Properties of Bivariate Normal

Ping-Chun Hsieh (謝秉均)

December 8, 2021

Announcements

- HW3 is now available on E3!
  - Due on 12/17 (Friday), 1pm

Quick Review

$f_{V_1V_2}(v_1,v_2) = \underbrace{\frac{1}{\det(A)}}\cdot \underbrace{f_{U_1U_2}\left(A^{-1}\begin{bmatrix}v_1\\v_2\end{bmatrix}\right)}$

$U_1, U_2, f_{U_1U_2}$

- Linear transformation of 2 random variables? $\begin{bmatrix}V_1\\V_2\end{bmatrix} = \underbrace{\begin{bmatrix}a & b\\ c & d\end{bmatrix}}_{= A}\begin{bmatrix}U_1\\U_2\end{bmatrix}$

- How to find the distribution of $\underbrace{X+Y}_{= Z}$? (X, Y independent)

- "Convolution Theorem"

$$P_Z(z) = \sum_{\text{all } x} P_X(x)\cdot P_Y(z-x)$$

-

This Lecture

1. Moment Generating Functions

2. Covariance and Correlation Coefficient

3. Nice Properties of Bivariate Normal

- Reading material: Chapter 10.2-10.5 and 11.1

Recall: There are still a few remaining questions about bivariate normal…

(Q1) Is X_2 a normal random variable? What is the PDF?

MGF and sum of independent random variables

(Q2) What is "ρ" in the joint PDF of bivariate normal?

Covariance and correlation coefficient

(Q3) Why is bivariate normal useful? Any nice properties?

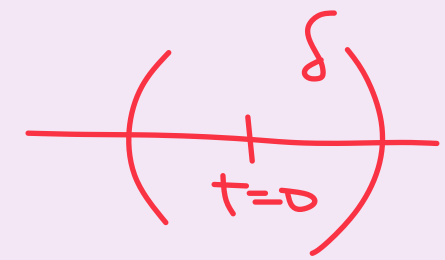
4 nice properties

(Q1) Moment Generating Functions

Moment Generating Function (Formally)

- **Moment Generating Function (MGF)**: For a random variable X, define

$$M_X(t) = E[e^{tX}], \quad t \in \mathbb{R}$$



If there exists $\delta > 0$ such that $M_X(t) < \infty$ for all $t \in (-\delta, \delta)$, then $M_X(t)$ is called the moment generating function of X

- Remark: If X is discrete with PMF $p_X(x)$, then

$$M_X(t) = \sum_{\text{all } x} p_X(x) \cdot e^{tx}$$

- Remark: If X is continuous with PDF $f_X(x)$, then

$$M_X(t) = \int_{-\infty}^{+\infty} f_X(x) \cdot e^{tx} \, dx$$

Nice Properties of MGF?

- **Let X_1, X_2 be two random variables**:

1. ______ Suppose $M_{X_1}(t) = M_{X_2}(t)$, for all $t \in \mathbb{R}$. Do X_1 and X_2 always have the same distribution (i.e., the same CDF)?

2. ______ Could we find moments $E[X_1^n]$ by using $M_{X_1}(t)$?

3. ______ Suppose X_1, X_2 are independent. Could we express $M_{X_1+X_2}(t)$ in terms of $M_{X_1}(t), M_{X_2}(t)$?

Nice Property (I): MGF Uniqueness Theorem

- **MGF Uniqueness Theorem**: Let X_1 and X_2 be two random variables with MGFs $M_{X_1}(t)$ and $M_{X_2}(t)$, respectively. If $M_{X_1}(t) = M_{X_2}(t)$ for all t in some interval $(-\alpha, \alpha)$, then X_1 and X_2 follow the same distribution, i.e.

$$P(X_1 \le u) = P(X_2 \le u), \text{ for all } u \in \mathbb{R}$$

- Remark: More details in the following reference
  - J. H. Curtiss, "A note on the theory of moment generating functions," 1942
  - https://projecteuclid.org/download/pdf_1/euclid.aoms/1177731541

Example: Find CDF from MGF

- Example: Suppose the MGF of a random variable X is

$$M_X(t) = \frac{1}{6}e^{-2t} + \frac{1}{3}e^{-t} + \frac{1}{4}e^{t} + \frac{1}{4}e^{2t}$$

  - Question: $P(|X| \leq 1) = ?$

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

$$= \begin{cases} \sum_{\text{all } x} p_X(x) \cdot e^{tx} \\ \int f_X(x) \cdot e^{tx}\,dx \end{cases}$$

$$p_X(x) = \begin{cases} \frac{1}{6}, & \text{if } x = -2 \\ \frac{1}{3}, & \text{if } x = -1 \\ \frac{1}{4}, & \text{if } x = 1 \\ \frac{1}{4}, & \text{if } x = 2 \\ 0, & \text{otherwise} \end{cases}$$

(By Uniqueness Theorem)

$$P(|X| \leq 1) = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$$

MGF of Special Random Variables

Distribution	Moment-generating function $M_X(t)$
Degenerate δ_a	e^{ta}
Bernoulli $P(X=1)=p$	$1-p+pe^t$
Geometric $(1-p)^{k-1}p$	$\frac{pe^t}{1-(1-p)e^t}$ $\forall t < -\ln(1-p)$
Binomial $B(n,p)$	$(1-p+pe^t)^n$
Negative Binomial $NB(r,p)$	$\frac{(1-p)^r}{(1-pe^t)^r}$
Poisson $Pois(\lambda)$	$e^{\lambda(e^t-1)}$
Uniform (continuous) $U(a,b)$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Uniform (discrete) $DU(a,b)$	$\frac{e^{at}-e^{(b+1)t}}{(b-a+1)(1-e^t)}$
Laplace $L(\mu,b)$	$\frac{e^{t\mu}}{1-b^2t^2}$, $\lvert t\rvert < 1/b$
Normal $N(\mu,\sigma^2)$	$e^{t\mu+\frac{1}{2}\sigma^2t^2}$

$\rightarrow e^{t\cdot\mu+\frac{1}{2}\sigma^2t^2}$

(By Uniqueness Theorem)

- **Example:** If $M_X(t) = \frac{1}{2}e^{0\cdot t} + \frac{1}{2}e^t$, then what kind of r.v. is X?

  $X \sim \text{Bernoulli}(p = \frac{1}{2})$

- **Example:** If $M_Z(t) = e^{2t^2-t}$, then what kind of r.v. is Z?

  $Z \sim N(\mu = -1, \sigma^2 = 4)$

Nice Property (II): From Sum to Product

- **MGF and Sum of 2 Independent Random Variables**: Given 2 independent random variables X_1, X_2 with MGFs $M_{X_1}(t)$ and $M_{X_2}(t)$, the MGF of $X_1 + X_2$ is

$$M_{X_1+X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(t)$$

$E[XY]$

$= E[X] \cdot E[Y]$

- Proof:

$$M_{X_1+X_2}(t) = E\left[e^{t(X_1+X_2)}\right]$$

$$= E\left[e^{tX_1} \cdot e^{tX_2}\right]$$

independence of X_1, X_2

$$= E\left[e^{tX_1}\right] \cdot E\left[e^{tX_2}\right]$$

$$= M_{X_1}(t) \cdot M_{X_2}(t).$$

Example: MGF of Sum of 2 Normal R.V.s

- Example: $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
  - X_1 and X_2 are assumed to be independent
  - Question: What is the MGF of $X_1 + X_2$? What is the PDF of $X_1 + X_2$?

$$M_{X_1}(t) = e^{t \cdot \mu_1 + \frac{1}{2}\sigma_1^2 t^2}$$

$$M_{X_2}(t) = e^{t \cdot \mu_2 + \frac{1}{2}\sigma_2^2 t^2}$$

$$M_{X_1+X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(t) = e^{t(\mu_1+\mu_2) + \frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}$$

$$X_1 + X_2 \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$

$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

$\vdots$

$X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$

$X_1, X_2, \cdots, X_n$ are independent

$(X_1+X_2) + X_3 + \cdots + X_n \sim \mathcal{N}\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right)$

Nice Property (III): Why Is $M_X(t)$ Called the Moment Generating Function?

$M_X(t)$ needs to be differentiable around $t=0$

- **Recall:** What is the "n-th moment" of X?

- **Use MGF to Find Moments**: Let X be a random variable with MGF $M_X(t)$. Then, for every $n \in \mathbb{N}$, we have

$$E[X^n] = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$$

(Suppose X is discrete)

- **Proof:** $\frac{d}{dt} M_X(t)\Big|_{t=0} = \frac{d}{dt} E[e^{tX}]\Big|_{t=0} = \frac{d}{dt}\left(\sum_{\text{all } x} P_X(x) \cdot e^{tx}\right)\Big|_{t=0}$

$= \sum_{\text{all } x} P_X(x) \cdot \frac{d e^{tx}}{dt}$ $x \cdot e^{tx}\Big|_{t=0}$

$= E[X]$

$\frac{d^n}{dt^n} M_X(t)\Big|_{t=0} = \sum_{\text{all } x} P_X(x) \cdot \frac{d^n}{dt^n} e^{tx}\Big|_{t=0} = E[X^n]$

$\frac{d^n}{dt^n} e^{tx} = x^n e^{tx}$

Example: Moments of Exp(λ)

t

- Example: Suppose $X \sim \text{Exp}(\lambda)$ $\Rightarrow$ $f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0 \\ 0, & \text{else} \end{cases}$
  - What is the MGF of X?
  - Use MGF to verify that $E[X] = \dfrac{1}{\lambda}$ and $\text{Var}[X] = \dfrac{1}{\lambda^2}$?

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{+\infty} f_X(x) \cdot e^{tx}\,dx = \int_0^{\infty} \lambda e^{-\lambda x} \cdot e^{tx}\,dx = \frac{\lambda}{-(\lambda - t)} e^{-(\lambda-t)x} \Big|_0^{\infty} = \begin{cases} \dfrac{\lambda}{\lambda - t}, & \text{if } \lambda > t \\ \text{does not exist}, & \text{else} \end{cases}$$

$$e^{-\lambda x} \cdot e^{tx} = e^{-(\lambda - t)x}$$

$$E[X] = \frac{d}{dt}\left(\frac{\lambda}{\lambda - t}\right)\Big|_{t=0} = \frac{\lambda}{(\lambda - t)^2}\Big|_{t=0} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

$$E[X^2] = \frac{d^2}{dt^2}\left(\frac{\lambda}{\lambda - t}\right)\Big|_{t=0} = \frac{2\lambda}{(\lambda - t)^3}\Big|_{t=0} = \frac{2}{\lambda^2}$$

$$\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{1}{\lambda^2}$$

"ρ"

(Q2) Covariance

Motivating Example for “Covariance”

- Example: Bus #2 (NCTU - Mackay - Train Station)
  - $X =$ traveling time from NCTU to Mackay
  - $Y =$ traveling time from Mackay to Train Station
  - We want to know Var$[X + Y]$
  - Question: Given Var$[X]$ and Var$[Y]$, can we get Var$[X + Y]$?

Covariance and Where to Find Them

- Property:

$$\text{Var}[aX + bY] = a^2\text{Var}[X] + b^2\text{Var}[Y] + \overbrace{2abE[(X - E[X])(Y - E[Y])]}^{\text{"Covariance"}}$$

$$\text{Var}[\underbrace{aX+bY}_{= Z}] = E\left[\left((aX+bY) - E[aX+bY]\right)^2\right]$$

$$= E\left[\left((aX - E[aX]) + (bY - E[bY])\right)^2\right]$$

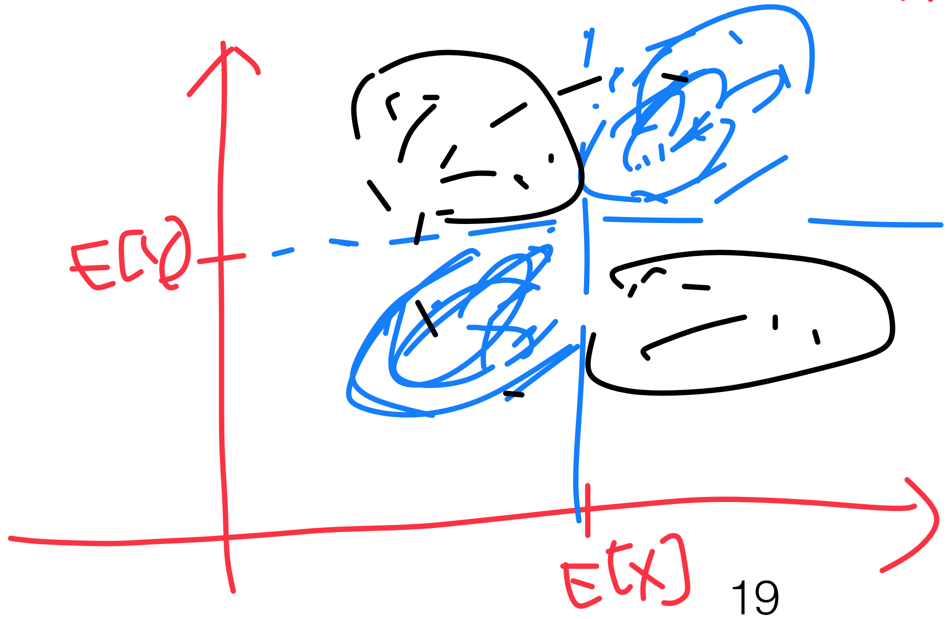
$$= E\left[(aX - E[aX])^2 + (bY - E[bY])^2 + 2 \cdot (aX - E[aX])(bY - E[bY])\right]$$

Covariance

- **Covariance**: Let X, Y be two random variables. Then, the covariance of X and Y is defined as

$$\mathrm{Cov}(X, Y) = E[\underbrace{(X - E[X])}\underbrace{(Y - E[Y])}]$$

- $\mathrm{Cov}(X, X) =$ Var[X]
- $\mathrm{Cov}(X, Y) = 0$: X, Y are said to be uncorrelated
- $\mathrm{Cov}(X, Y) > 0$: X, Y are said to be positively correlated
- $\mathrm{Cov}(X, Y) < 0$: X, Y are said to be negatively correlated
- Intuition:



Another Expression of Covariance

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

- Let X, Y be two random variables. Then, the covariance of X and Y can also be written as 乘積的平均 平均的乘積

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

- Question: How to show this?

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

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

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

$P(X\in A, Y\in B)$

$= P(X\in A)\cdot P(Y\in B)$

- Question: If X, Y are independent, then $\text{Cov}(X, Y) =$ 0
- Question: How about the converse argument?

?

$\text{Cov}(X,Y)=0 \Rightarrow X, Y$ independent

Example: Uncorrelated $\nRightarrow$ Independence

- Example: The pair of random variables (X, Y) takes the values $(1,0), (0,1), (-1,0), (0,-1)$, each with probability $\frac{1}{4}$
  - $\text{Cov}(X, Y)$ = ?
  - Are X, Y independent?

$$P_{XY}(x,y) = \begin{cases} \frac{1}{4}, & \text{if } (x,y) = (1,0) \\ \frac{1}{4}, & (0,1) \\ \frac{1}{4}, & (-1,0) \\ \frac{1}{4}, & (0,-1) \end{cases}$$

$$\text{Cov}(X,Y) = \underbrace{E[XY]}_{=0} - \underbrace{E[X]}_{=0} \cdot \underbrace{E[Y]}_{=0} = 0$$

Find some events A, B, show $P(X \in A, Y \in B) \neq P(X \in A) \cdot P(Y \in B)$

$A = \{1\}$

$B = \{1\}$

$P(X \in A, Y \in B) = 0$

$P(X \in A) = \frac{1}{4}$

$P(Y \in B) = \frac{1}{4}$

A Property of Covariance

- Property:

$$\left(\text{Cov}(X, Y)\right)^2 \leq \text{Var}[X] \cdot \text{Var}[Y]$$

- Question: How to show this?

Any Issue With Covariance?

- Example: Bus #2 (NCTU - Mackay - Train Station)
  - From NCTU to Mackay: X minutes
  - From Mackay to Train Station: Y minutes
  - Question: Cov(X, Y) = ?
  - Question: What if time is measured in "seconds"? Any change in the covariance?

Covariance is Sensitive to the Units

- Property: $\underline{\text{Cov}(aX, aY) = a^2 \cdot \text{Cov}(X, Y)}$
  - a: scaling factor due to change of unit

- Question: Any suggested solution?

Correlation Coefficient

- **Correlation Coefficient**: Let X, Y be two random variables with finite variance $\sigma_X^2 > 0$ and $\sigma_Y^2 > 0$. Then, the correlation coefficient of X and Y is defined as

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- Question: Do we have $\rho(X, Y) = \rho(aX, aY)$, for any $a \neq 0$?

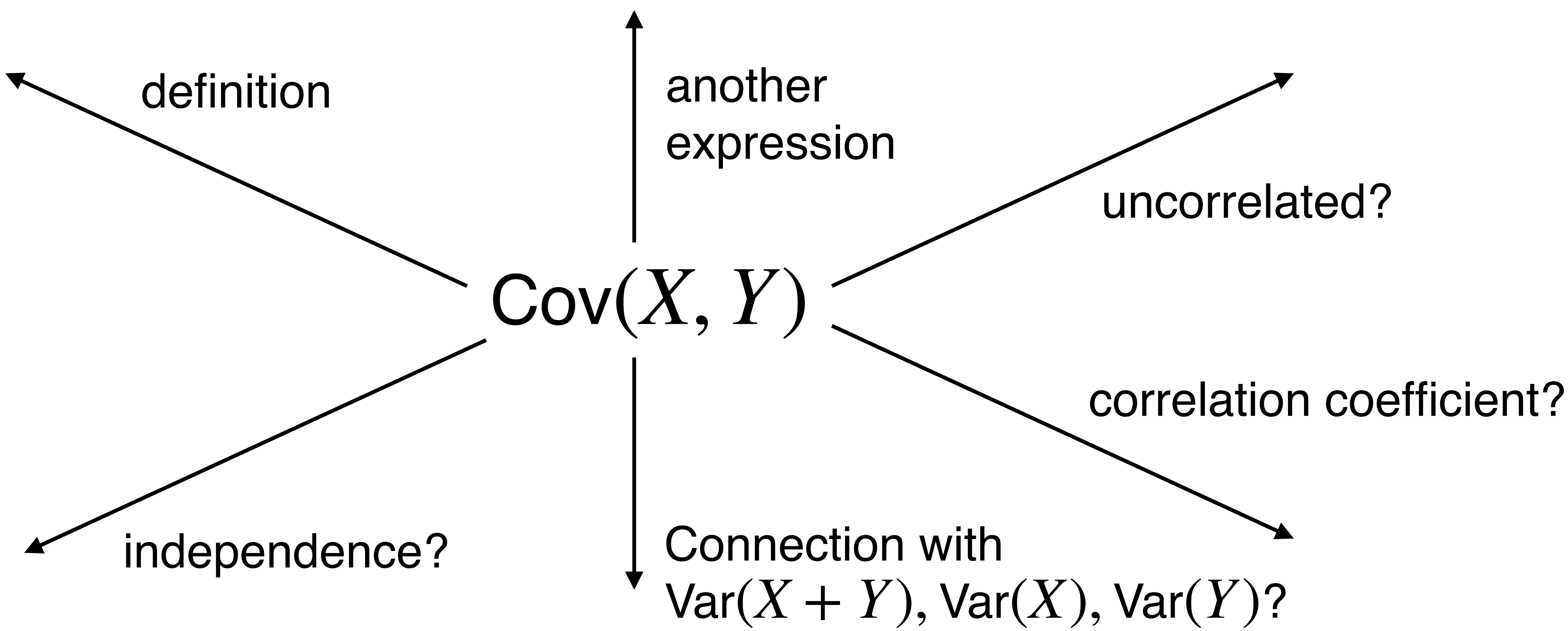
A Property of Correlation Coefficient

- **Property**:

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

- Question: How to prove this?

A Brief Summary of Covariance



(Q3) Nice Properties of Bivariate Normal

Properties of Bivariate Normal R.V.

- **Suppose the joint PDF of X_1, X_2 is bivariate normal as**

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right)}{2(1-\rho^2)} \right]$$

Then we have:

1. Marginal: $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
2. Conditional: $X_2 \mid X_1 = x_1 \sim \mathcal{N}\left(\mu_2 + \frac{\rho\sigma_2(x_1-\mu_1)}{\sigma_1}, (1-\rho^2)\sigma_2^2\right)$
3. Correlation coefficient: $\rho(X_1, X_2) = \rho$
4. If X_1, X_2 are uncorrelated ($\rho = 0$), then X_1, X_2 are independent

1. Marginal: $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[- \frac{\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right)}{2(1-\rho^2)} \right]$$

$$\frac{\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right)}{2(1-\rho^2)} = \frac{(x_1-\mu_1)^2}{2\sigma_1^2} + \frac{1}{2\sigma_2^2}\left(\frac{(x_2-\mu_2) - \rho(x_1-\mu_1)}{\sqrt{1-\rho^2}} \right)^2$$

Take X_1 for example (X_2 would be similar)

$$f_{X_1}(x_1) =$$

2. Conditional: $X_2 \mid X_1 = x_1 \sim \mathcal{N}\left(\mu_2 + \frac{\rho\sigma_2(x_1 - \mu_1)}{\sigma_1}, (1-\rho^2)\sigma_2^2\right)$

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[- \frac{\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}{2(1-\rho^2)}\right]$$

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[- \frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right]$$

$$f_{X_2|X_1}(x_2 \mid x_1) =$$

3. Correlation Coefficient: $\rho(X_1, X_2) = \rho$

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}{2(1-\rho^2)} \right]$$

$$\text{Cov}(X_1, X_2) = E[(X_1-\mu_1)(X_2-\mu_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1-\mu_1)(x_2-\mu_2) f_{X_1X_2}(x_1, x_2) dx_1 dx_2$$

Hint: $f_{X_2|X_1} = \dfrac{f_{X_1X_2}}{f_{X_1}} \Rightarrow f_{X_1X_2} = f_{X_2|X_1} f_{X_1}$

$\text{Cov}(X_1, X_2) =$

4. Uncorrelated ($\rho = 0$) Implies Independence

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right)}{2(1-\rho^2)} \right]$$

- **If $\rho = 0$**:

$f_{X_1X_2}(x_1, x_2) =$

Final Remark: X_1, X_2 Normal $\nRightarrow$ X_1, X_2 Bivariate Normal

- Example: Let Y and Z be two independent standard normal r.v.s
  - $X_1 = |Y| \cdot \text{sign}(Z)$
  - $X_2 = Y$

- Question:
  - Are X_1 and X_2 normal?
  - Are X_1 and X_2 bivariate normal?