1179: Probability
Lecture 20 — Conditionals, Expected Value, and Bivariate Normal

Ping-Chun Hsieh (謝秉均)

November 24, 2021

Announcements

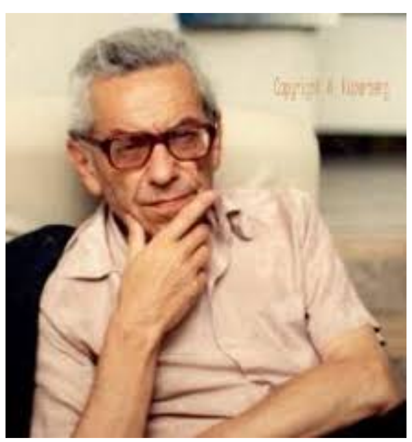
- Midterm exam booklets will be returned to you on 11/25 (Thu.)
  - 6:30pm - 8pm @ EC345

- No class next Wednesday (12/1)

Erdős Number

- Six degree of separation?
- In math, Erdős Number embodies a similar principle

Richard Bellman
(Erdős number = 1)



Paul Erdős

Noga Alon
(Erdős number = 1)

Nick Duffield
(Erdős number = 2)

(the most prolific mathematician: 1500+ papers)

The Erdős Number Project: http://www.oakland.edu/enp/

Quick Overview

- Given 2 random variables X, Y: what have we learned so far?
  1. Joint CDF
  2. Marginal CDF
  3. Joint PMF / PDF
  4. Marginal PMF / PDF
  5. Independence
  6. Conditional distribution → $P(Y=2 \mid X=1)$
  7. Expected value involving both X, Y $E[X+Y]$ $E[XY]$ $E[\log X + e^{Y}]$
  8. Bivariate normal
  9. Distribution of $X + Y$
  10. Covariance and correlation

This Lecture

1. Conditional Distributions

2. Expected Value Regarding 2 Random Variables

3. Bivariate Normal Random Variables

- Reading material: Chapter 8.3~8.4 and 10.5

Example: Using Joint PMF to Find Conditional PMF

- Example: Bus #2 (NCTU - Mackay - Train Station)
  - $X =$ traveling time from NCTU to Mackay
  - $Y =$ traveling time from Mackay to Train Station
  - $P(X = 10 \,|\, \underline{Y = 15})$ = ?

Joint PMF	X=10	X=15	X=20
Y=10	0.1	0.1	0.05
Y=15	0.1	0.3	0.1
Y=20	0.05	0.1	0.1

$P(X=10 \mid Y=15)$

$= \dfrac{P(X=10 \text{ and } Y=15)}{P(Y=15)} = \dfrac{0.1}{0.5}$

$= 0.2$

$P(X=15 \mid Y=15)$

$P(X=20 \mid Y=15)$

Conditional PMF (Formally)

- **Conditional PMF**: Let X, Y be two discrete random variables with joint PMF $p_{XY}(x, y)$. When $P(Y = y) > 0$, the conditional PMF of X given $Y = y$ is

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

- **Question:** Conditional PMF of Y given $X = x$?

$$P_{Y|X}(y|x) = \frac{P_{XY}(x,y)}{P_X(x)}$$

- **Question:** $\sum_{\text{all } x} p_{X|Y}(x \mid y) = \sum_{\text{all } x} \frac{P_{XY}(x,y)}{P_Y(y)} = 1$

Conditional CDF of Discrete Random Variables

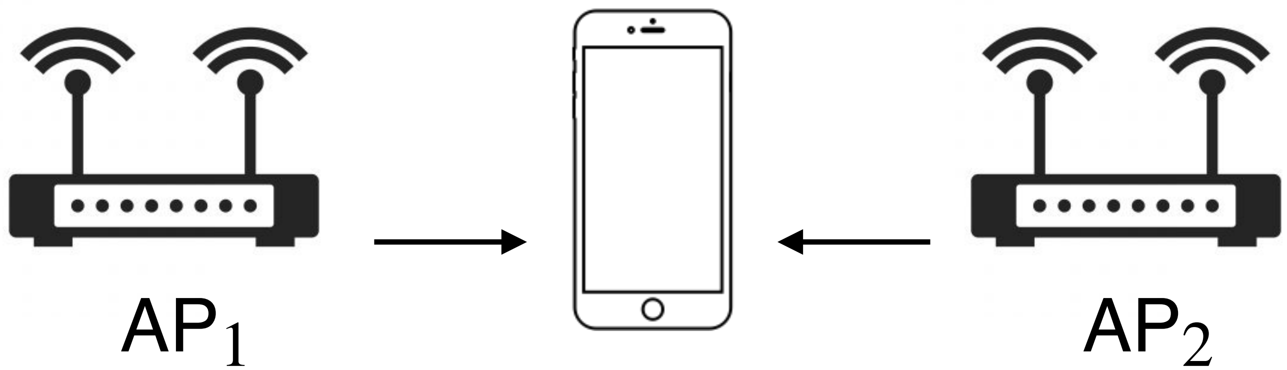
- **Conditional CDF**: Let X, Y be two discrete random variables with joint PMF $p_{XY}(x, y)$ and marginal PMFs $p_X(x), p_Y(y)$. When $P(Y = y) > 0$, the conditional CDF of X given $Y = y$ is

$$\boxed{F_{X|Y}(x\,|\,y)} := P(X \leq x | Y = y) = \sum_{t \leq x} p_{X|Y}(t|y) = \sum_{t \leq x} \frac{p_{XY}(t, y)}{p_Y(y)}$$

P(X≤t)

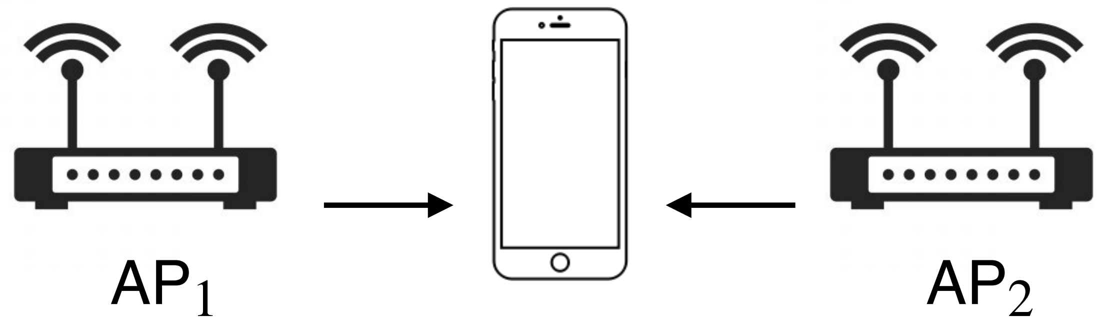
conditional PMF

Example: Conditioning and Sum of Poisson



- Let N_1 and N_2 be the # of bits transmitted by AP$_1$ and AP$_2$ in a time interval T, respectively
  - N_1 and N_2 are Poisson with rates λ_1 and λ_2, respectively.
  - Moreover, N_1 and N_2 are independent
  - Define $M = N_1 + N_2$
  - Question: Conditional PMF $p_{N_1|M}(n \mid m)$ = ?

Example: Conditioning and Sum of Poisson



- Conditional PMF $p_{N_1|M}(n\,|\,m)$

$M = N_1 + N_2$

$N_1 \sim \mathrm{Poi}(\lambda_1, T)$

$N_2 \sim \mathrm{Poi}(\lambda_2, T)$

N_1, N_2 independent

$M \sim \mathrm{Poi}(\lambda_1+\lambda_2, T)$

$$p_{N_1|M}(n|m) = \frac{P(N_1 = n \text{ and } M = m)}{P(M=m)}$$

$(m \geq 0,\ n \geq 0,\ m \geq n)$

$P(N_1 = n \text{ and } N_2 = m-n)$

$$= \frac{P(N_1=n)\cdot P(N_2=m-n)}{P(M=m)}$$

$$= \frac{\dfrac{\cancel{e^{-\lambda_1 T}}\cdot(\lambda_1 \cancel{T})^n}{n!}\cdot\dfrac{\cancel{e^{-\lambda_2 T}}\cdot(\lambda_2 \cancel{T})^{m-n}}{(m-n)!}}{\dfrac{\cancel{e^{-(\lambda_1+\lambda_2)T}}\cdot((\lambda_1+\lambda_2)\cancel{T})^m}{m!}} = \frac{m!}{n!(m-n)!}\cdot\frac{\lambda_1^n\cdot\lambda_2^{m-n}}{(\lambda_1+\lambda_2)^m}$$

$$C^m_n \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^n \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{m-n}$$

$\mathrm{Binomial}\left(m, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$

Conditional Expectation: Discrete Case

- **Conditional Expectation**: Let X, Y be two discrete random variables with joint PMF $p_{XY}(x, y)$. When $P(Y = y) > 0$, the conditional expected value of X given $Y = y$ is

$$E[X \mid Y = y] = \sum_{\text{all } x} x \cdot P(X = x \mid Y = y) = \sum_{\text{all } x} x \cdot \frac{P_{XY}(x,y)}{P_Y(y)}$$

- Question: Conditional expectation of Y given $X = x$?

$$E[Y \mid X=x] = \sum_{\text{all } y} y \cdot P(Y=y \mid X=x)$$

Useful Property: Law of Iterated Expectation

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

- **Question:** Define $g(y) = E[X \mid Y = y]$
  - What kind of object is $g(Y)$? random variable
  - $E[g(Y)]$ = ?

$E[E[X|Y=y]]_{y\sim Y}$, where $E[X|Y=y] = g(y)$

$= E[E[X|Y]]$, where $E[X|Y] = g(Y)$

$= \sum_{\text{all } y} g(y) \cdot p_Y(y)$ ... outer expectation

$= \sum_{\text{all } y} \left(\sum_{\text{all } x} x \cdot p_{X|Y}(x|y) \right) p_Y(y)$

$= \sum_{\text{all } x} \sum_{\text{all } y} x \boxed{p_{X|Y}(x|y) \cdot p_Y(y)}$, where $p_{X|Y}(x|y) \cdot p_Y(y) = p_{XY}(x,y)$

$= \sum_{\text{all } x} x \cdot p_X(x)$

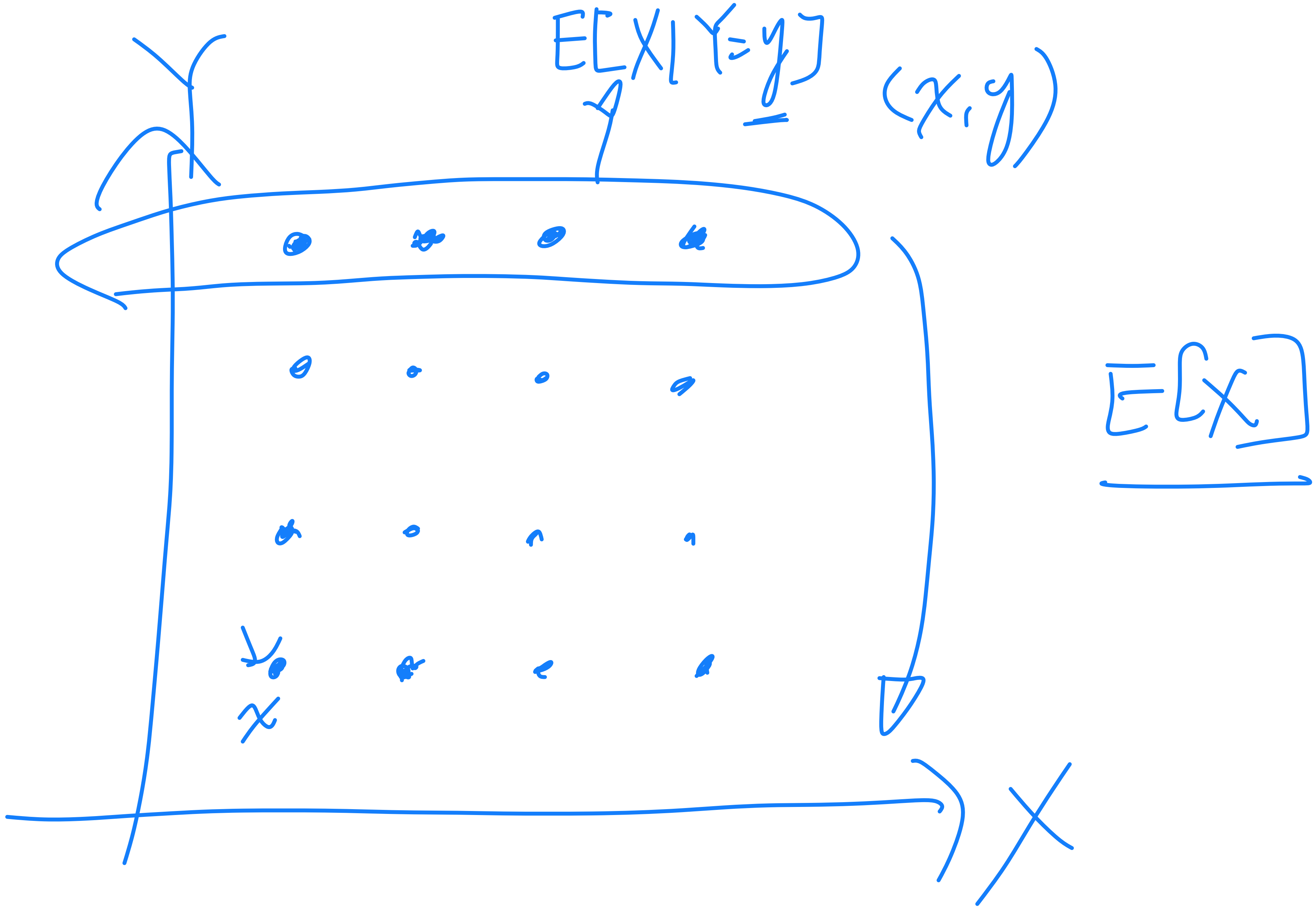
$= E[X]$

- **Law of Iterated Expectation (LIE)**: Let X, Y be two discrete random variables with joint PMF $p_{XY}(x, y)$. Then,

"divide-and-conquer"

$$E\big[E[X \mid Y]\big] = E[X]$$

- **Remark:** This still holds for continuous cases

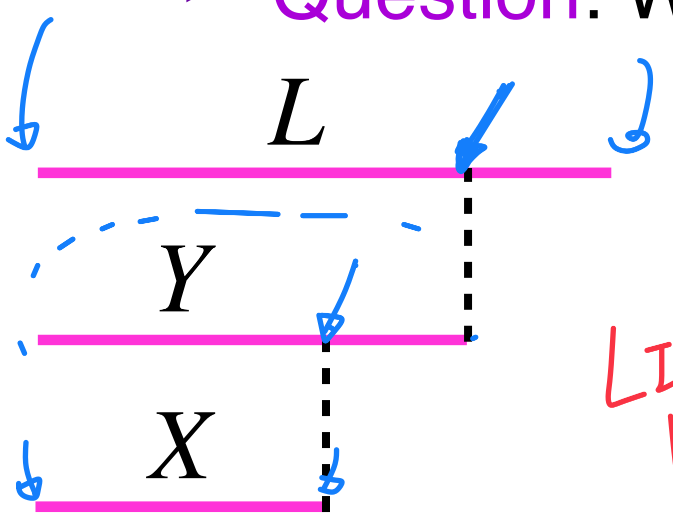


$E[X \mid Y = y]$
(x, y)
$E[X]$
Y
X
y
x

Example: Breaking A Stick

$g(y) = E[X \mid Y = y]$

- Example: We are breaking a stick of length L at a point which is chosen uniformly over its length and keep the piece that contains the left end. Next, we repeat the process with the piece we have.
  - Question: What is the expected length of the remaining stick? $E[X]$?



$E[X] \to$ ① $f_{XY}(x,y) \to f_X(x)$

$E[X] \to$ ② $E[E[X \mid Y]]$

$$E[X] \overset{\text{LIE}}{=} E[\underbrace{E[X \mid Y]}_{g(Y)}] = E\left[\frac{1}{2} \cdot Y\right]$$

$$= \frac{1}{2} \cdot \frac{1}{2} L$$

$$= \frac{L}{4}.$$

$Y \sim Unif(0, L)$

Conditional PDF (Formally)

- **Conditional PDF**: Let X, Y be two continuous random variables with joint PDF $f_{XY}(x, y)$. When $f_Y(y) > 0$, the conditional PDF of X given $Y = y$ is

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

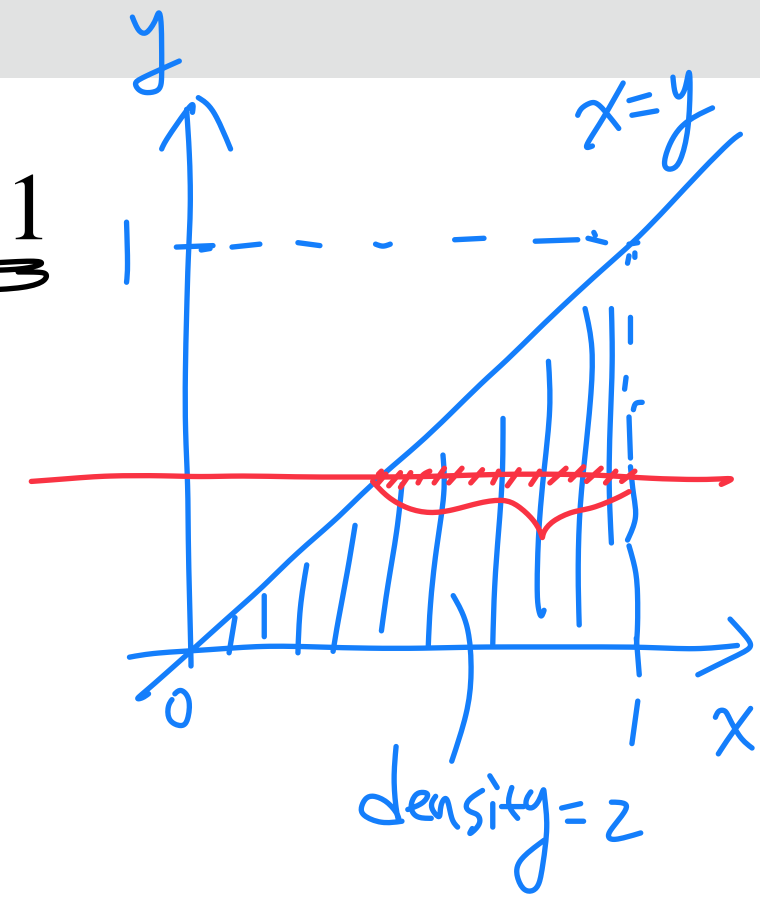
- Question: Conditional PDF of Y given $X = x$?

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

Example: Find Conditional PDF From Joint PDF

- Example:

$$f(x, y) = \begin{cases} 2 & , \text{ if } 0 < y < x < 1 \\ 0 & , \text{ otherwise} \end{cases}$$



- $f_{X|Y}(x \mid y) = ?$

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \qquad (0<y<1)$$

marginal PDF

$$= \begin{cases} 0 & , \quad x \le y \text{ or } x \ge 1 \\ \dfrac{2}{2(1-y)} & , \quad y < x < 1 \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\, dx = \int_{y}^{1} 2 \cdot dx = 2x\Big|_y^1 = 2(1-y)$$

Expected Value Regarding 2 Random Variables

Recall: LOTUS for 1 Discrete Random Variable

Expected Value of a Function of Discrete R.V.:

1. Let X be a discrete random variable with
   - the set of possible values S
   - PMF of X is $p_X(x)$
2. Let $g(\cdot)$ be a real-valued function

The expectation of $g(X)$ is

$$E[g(X)] = \sum_{x \in S} g(x) \cdot p_X(x)$$

LOTUS for 2 Discrete Random Variables

Expected Value of a Function of 2 Discrete RVs:

1. Let X, Y be 2 discrete random variables with sets of possible values S_X, S_Y and joint PMF $p(x, y)$

2. Let $g(\,\cdot\,,\,\cdot\,)$ be a function from $\mathbb{R}^2 \to \mathbb{R}$

The expected value of $g(X, Y)$ is

LOTUS for 2 random variables

$$\underbrace{E[g(X, Y)]}_{\substack{\| \\ Z}} = \sum_{\text{all } z} z \cdot p_Z(z) = \sum_{\substack{x \in S_X \\ y \in S_Y}} g(x, y) \cdot \underbrace{p(x, y)}_{\text{joint PMF}},$$

Example: Using Joint PMF to Find Expected Value

$g(x,y) = x+y$

- Example: Bus #2 (NCTU - Mackay - Train Station)
  - $X =$ traveling time from NCTU to Mackay
  - $Y =$ traveling time from Mackay to Train Station
  - $E[X+Y]$ = ?

Joint PMF	X=10	X=15	X=20
Y=10	0.1	0.1	0.05
Y=15	0.1	0.3	0.1
Y=20	0.05	0.1	0.1

$g(X,Y)$

$E[\underline{X+Y}]$

LOTUS for 2 r.v.s

$$= \sum_{\substack{x=10,15,20 \\ y=10,15,20}} (x+y) \cdot P_{XY}(x,y)$$

$$= (20\times 0.1 + 25\times 0.1 + 30\times 0.05)$$
$$+ (25\times 0.1 + 30\times 0.3 + 35\times 0.1)$$
$$+ (30\times 0.05 + 35\times 0.1 + 40\times 0.1)$$

Useful Property (I)

Recall:
$E[\alpha X + \beta] = \alpha \cdot E[X] + \beta$

- **Linearity Property**:

  $E[\alpha \cdot g_1(X, Y) + \beta \cdot g_2(X, Y)] = \alpha E[g_1(X, Y)] + \beta E[g_2(X, Y)]$

- Remark: X, Y are NOT required to be independent
- Remark: This results holds for both discrete and continuous cases
- Proof: LOTUS

Useful Property (II)

$g, h =$ two functions

$g(x) = x$

$h(y) = y$

- **Property under Independence**: Suppose X, Y are independent random variables. Then, we have

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

- Remark: This result holds for both discrete and continuous cases
- Proof: (Assume discrete cases).

$$E[g(X) \cdot h(Y)] = \sum_{\text{all } x} \sum_{\text{all } y} \left(g(x) h(y)\right) \cdot \underbrace{P_{XY}(x,y)}_{= P_X(x) \cdot P_Y(y)}$$

$$= \sum_{\text{all } x} g(x) P_X(x) \underbrace{\left(\sum_{\text{all } y} h(y) \, P_Y(y) \right)}_{E[h(Y)]}$$

$$= \sum_{\text{all } x} g(x) \cdot P_X(x) E[h(Y)] = E[g(X)] \cdot E[h(Y)]$$

$E[XY] = E[X]E[Y]$ If X, Y Are Independent

$g(x) = x$
$h(y) = y$

Corollary: Let X, Y be 2 independent random variables. Then,

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

- Question: How about the reverse argument?

X, Y independent $\Longrightarrow$ $E[XY] = E[X] \cdot E[Y]$

$\Longleftarrow$??

$E[XY] = E[X]E[Y] \nRightarrow X, Y$ Independent

- Example: Let X be a discrete random variable with $P(X = 1) = P(X = -1) = 0.5$.
  - Define $Y = |X|$
  - $E[X] = ?$ $E[Y] = ?$
  - $E[XY] = ?$
  - Are X, Y independent?

$$X = \begin{cases} 1, & \text{w.p. } \frac{1}{3} \\ -1, & \text{w.p. } \frac{1}{3} \\ 0, & \text{w.p. } \frac{1}{3} \end{cases}$$

$$E[X] = 0$$

$$Y = |X|$$

$$E[Y] = \frac{2}{3}$$

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

$$E[XY] = 0$$

$$XY = \begin{cases} 1, & \text{w.p. } \frac{1}{3} \\ -1, & \text{w.p. } \frac{1}{3} \\ 0, & \text{w.p. } \frac{1}{3} \end{cases}$$

X, Y independent?? NOT!

More on $E[XY]$: Cauchy-Schwarz Inequality

- Recall: Cauchy Inequality in high school

- **Cauchy-Schwarz Inequality**: Let X, Y be two random variables. Then, we have

$$E[X^2] \cdot E[Y^2] \geq (E[XY])^2$$

- Question: Under what condition do we have "="?

Proof of Cauchy-Schwarz Inequality

$$E[X^2] \cdot E[Y^2] \geq (E[XY])^2$$

- Hint: Start from that $E[(tX + Y)^2] \geq 0$
- Proof:

Bivariate Normal Random Variables

Example: 2 Independent Normal Random Variables

- Example: $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
  - Suppose X_1, X_2 are independent.
  - What is the joint PDF? How to plot the contour?

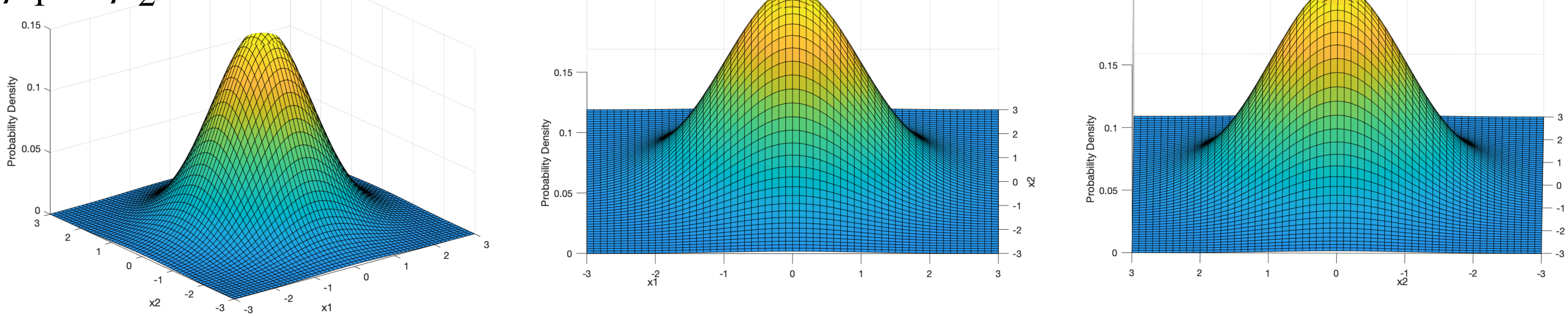
Joint PDF of 2 Independent Normal R.V.s (Formally)

- Given: $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
- Suppose X_1, X_2 are independent.

- **Joint PDF of 2 Independent Normal**:

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

$\sigma_1 = \sigma_2 = 1$
$\mu_1 = \mu_2 = 0$



2 Independent Normal: Matrix Form

- **Joint PDF of 2 Independent Normal**:

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

$$\Sigma = \begin{bmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{bmatrix}$$

$$x = \begin{bmatrix}x_1\\ x_2\end{bmatrix}, \mu = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$$

- **Joint PDF of 2 Independent Normal**:

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{|\det(\Sigma)|}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

One natural question:

Is it possible to construct a "jointly normal r.v." from "2 non-independent normal r.v.s"?

Construction of Bivariate Normal R.V.

- Idea: Let Z, W be 2 independent standard normal r.v.s and $\rho \in [-1,1]$. Define two random variables

$$X_1 = \sigma_1 Z + \mu_1$$
$$X_2 = \sigma_2\left(\rho Z + \sqrt{1-\rho^2}W\right) + \mu_2$$

- Question: Is it possible to find the joint PDF of X_1, X_2?

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

Bivariate Normal R.V.s (Formally)

- **Bivariate Normal**: X_1 and X_2 are said to be bivariate normal random variables if the joint PDF of X_1, X_2 is

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

The joint PDF can be written in matrix form as

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{|\det(\Sigma)|}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) \right]$$

where

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

- Notation for bivariate normal: $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$

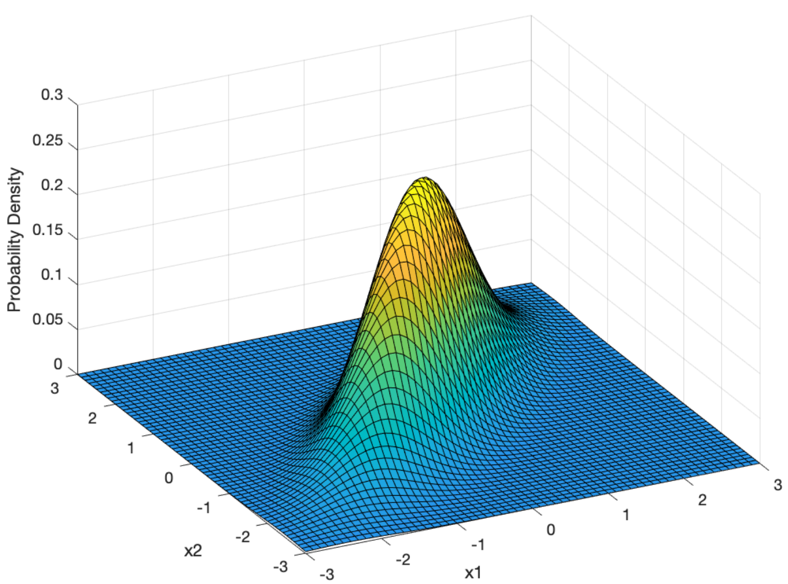
Plotting the Joint PDF Bivariate Normal

- **Joint PDF of Bivariate Normal**:

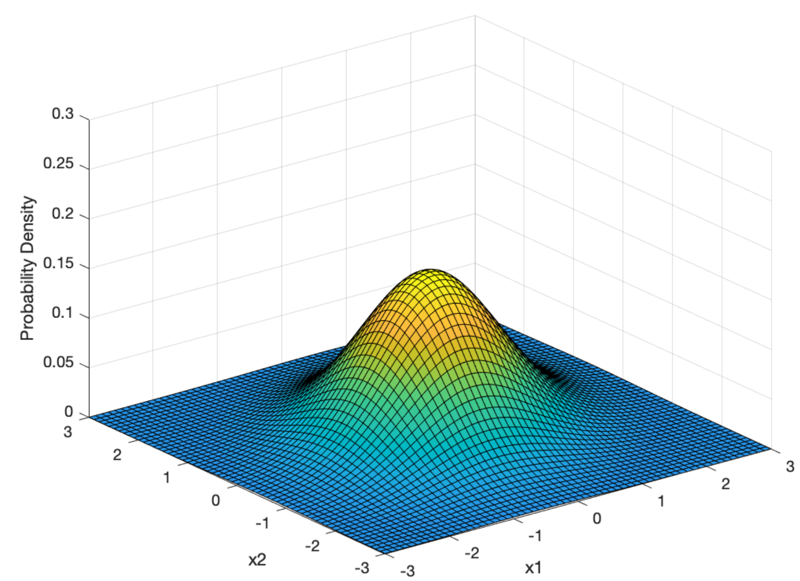
$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right)}{2(1-\rho^2)} \right]$$

- **Example:** $\sigma_1 = \sigma_2 = 1,\ \mu_1 = \mu_2 = 0$

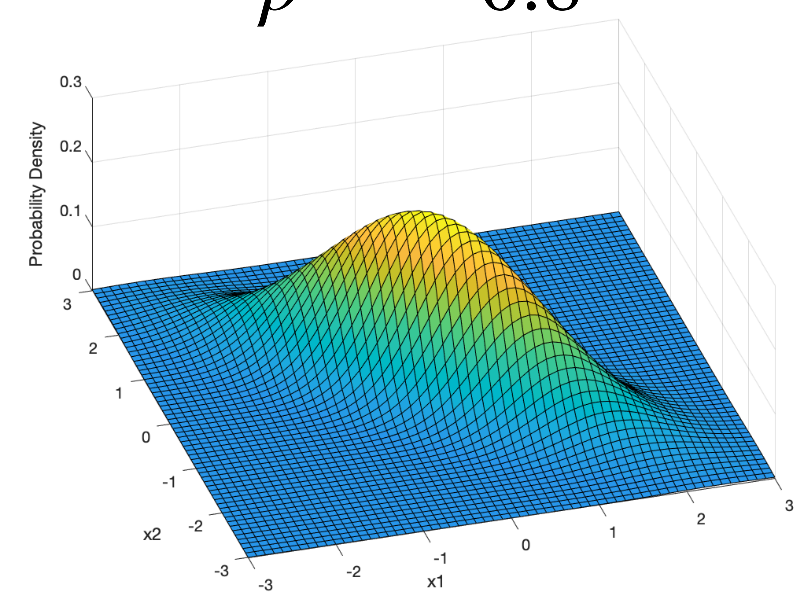
$\rho = 0.8$



$\rho = 0.2$



$\rho = -0.8$

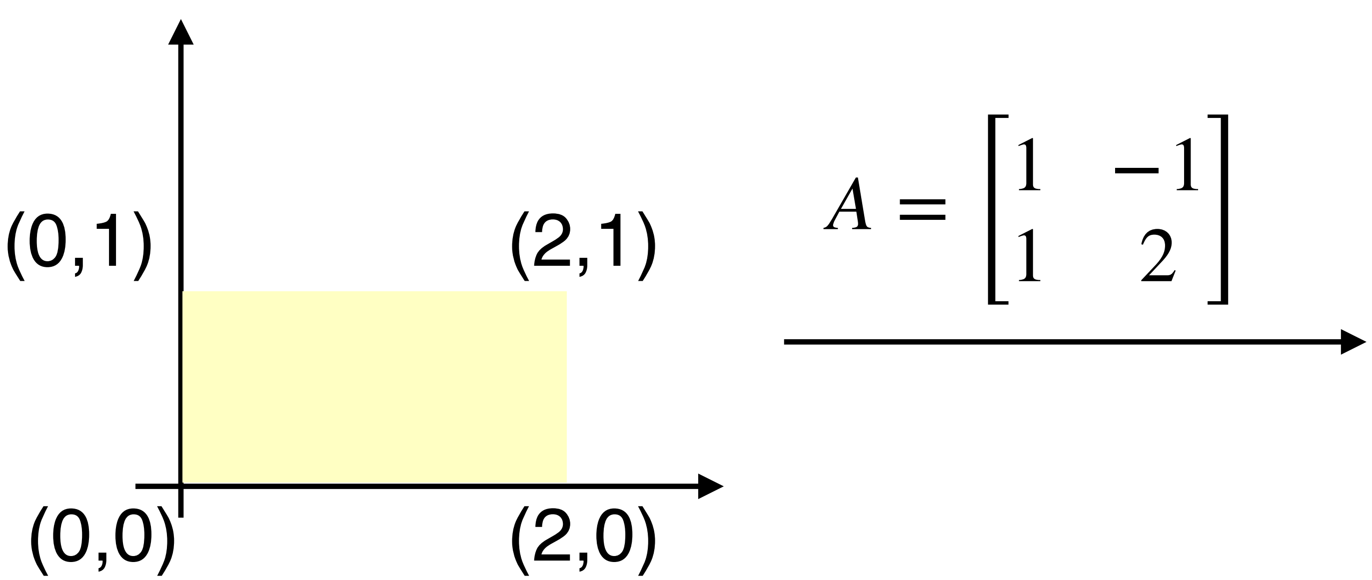


Linear Transformation of 2 Random Variables

- **Theorem**: Let U_1, U_2, V_1, V_2 be random variables that satisfy $V_1 = aU_1 + bU_2$ and $V_2 = cU_1 + dU_2$. Define the matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then, we have

$$f_{V_1V_2}(v_1, v_2) = \frac{1}{|\det(A)|} f_{U_1U_2}(A^{-1}[v_1, v_2]^T)$$

- **Intuition**:



For more details, please check: **https://www.stat.berkeley.edu/~aditya/resources/AllLectures2018Fall201A.pdf**

Bivariate Normal and Linear Transformation

- For simplicity, assume $\mu_1 = \mu_2 = 0$ (can be handled via translation)

$$X_1 = \sigma_1 Z$$
$$X_2 = \sigma_2\left(\rho Z + \sqrt{1-\rho^2}W\right)$$

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{|\det(A)|} f_{ZW}(A^{-1}[x_1, x_2]^T)$$

1-Minute Summary

1. Conditional Distributions

- Conditional PMF / PDF
- Law of iterated expectation (LIE): $E[E[X \mid Y]] = E[X]$

2. Expected Value Regarding 2 Random Variables

- LOTUS for 2 random variables
- $E[XY] = E[X]E[Y]$ under independence
- Cauchy-Schwarz: $E[X^2] \cdot E[Y^2] \geq (E[XY])^2$

3. Bivariate Normal Random Variables

- Linear transformation of 2 random variables