DCP 1206: Probability
Lecture 18 — Conditional Distributions and Bivariate Normal

Ping-Chun Hsieh

November 20, 2019

LeCun Yann’s Cake Analogy

- <u>Suppose machine intelligence is like a cake:</u>

LeCun Yann

Reinforcement learning (cherry)

Supervised learning (icing)

Unsupervised learning (the bulk of cake)

- RL researchers were a bit offended…, so they fought back…

Another Cake In the Eyes of RL Researchers

- There could be many cherries on the same cake lol

https://medium.com/syncedreview/yann-lecun-cake-analogy-2-0-a361da560dae

This Lecture

1. Conditional Distributions

2. Bivariate Normal Random Variables

- Reading material: Chapter 8.3 and 10.5

1. Conditional Distributions

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

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

$$P(X=10 \mid Y=15) = \frac{P(X=10 \text{ and } Y=15)}{P(Y=15)} = \frac{0.1}{0.1+0.3+0.1} = 0.2$$

$$P(X=15 \mid Y=15) = \frac{0.3}{0.5} = 0.6$$

$$P(X=20 \mid Y=15) = \frac{0.1}{0.5} = 0.2$$

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

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

joint PMF

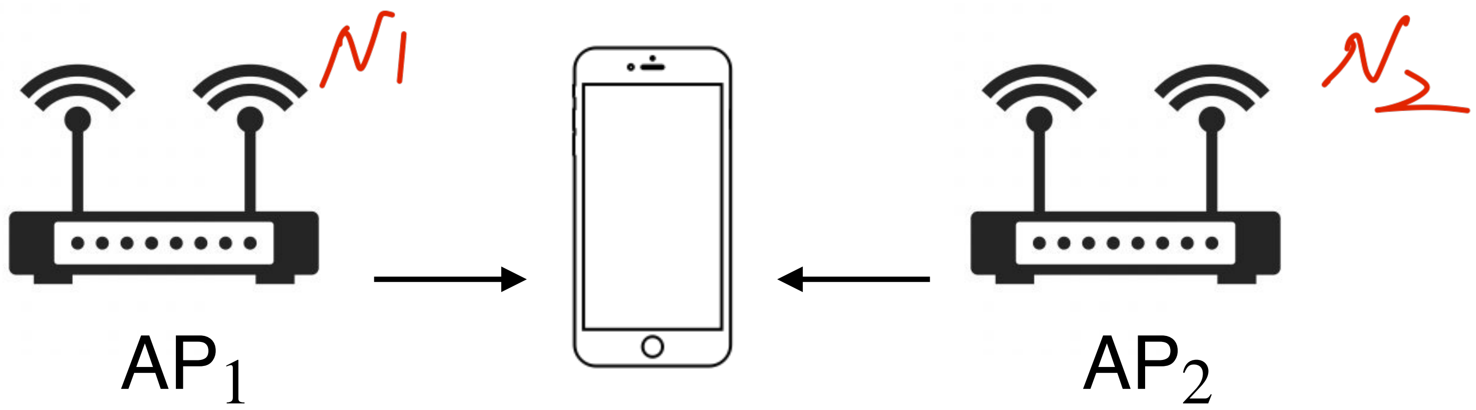
marginal PMF

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

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

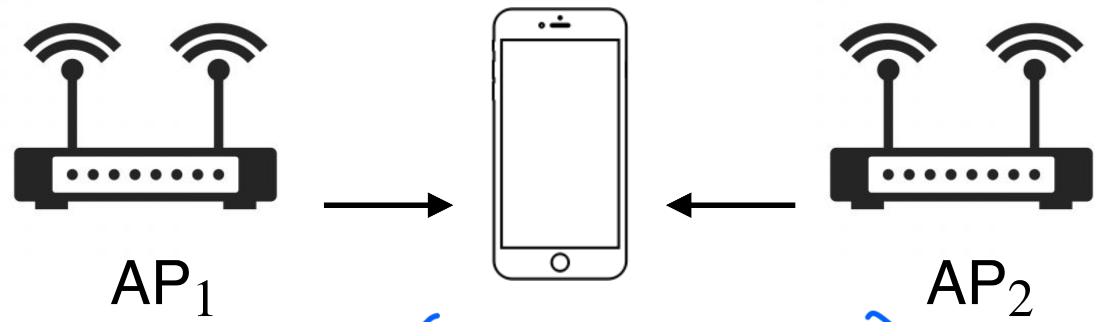
- Question: $\sum_x p_{X|Y}(x \mid y) =$

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 $p_{N_1|M}(n\,|\,m)$ = ?

Example: Conditioning and Sum of Poisson



$$P(N_1=k)=\begin{cases}\dfrac{e^{-\lambda_1 T}\cdot(\lambda_1 T)^k}{k!}, & k\ge 0\\ 0, & \text{else}\end{cases}$$

- Conditional PMF $p_{N_1|M}(n\,|\,m)$ $(m \ge n \ge 0)$

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

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

$$=\underbrace{\frac{m!}{n!(m-n)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^n\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{m-n}}$$

$C^m_n = \dfrac{m!}{n!(m-n)!}$

$\Downarrow$

PMF of Binomial$\left(m, \dfrac{\lambda_1}{\lambda_1+\lambda_2}\right)$

Conditional Expectation: Discrete Case (Formally)

- **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

$$E[X \mid Y = y] = \sum_x \underbrace{x}_{\text{function value}} \cdot \underbrace{P(X = x \mid Y = y)}_{\text{conditional PMF}} = \sum_x x \cdot \underbrace{P_{X|Y}(x|y)}$$

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

cf.:

$$E[X] = \sum_x x \cdot \underbrace{P(X=x)}$$

$$E[Y \mid X=x] = \sum_y y \cdot P_{Y|X}(y|x)$$

Law of Iterated Expectation (Formally)

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

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

$$= \sum_y (g(y))(p_Y(y))$$

$$= \sum_y E[X|Y=y] \cdot p_Y(y)$$

$$= \sum_y \left(\sum_x x \cdot p_{X|Y}(x|y) \right) \cdot p_Y(y)$$

$$= \sum_{x,y} x \cdot \underbrace{p_{X|Y}(x|y) \cdot p_Y(y)}_{= p_{XY}(x,y)}$$

$$= E[X]$$

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

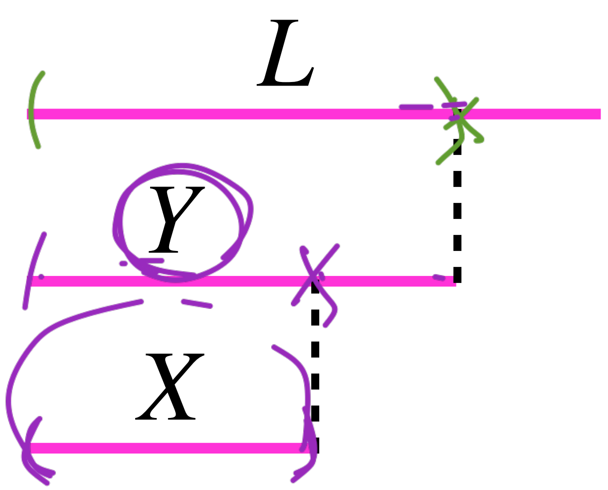
$$E\big[E[X \mid Y]\big] = E[X]$$

→ divide-and-conquer

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

Example: Breaking A Stick

- 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?



$$E[X] = E\big[\underbrace{E[X \mid Y]}_{= \frac{1}{2}\cdot Y}\big]$$

$$= E\left[\tfrac{1}{2} Y\right]$$

$$= \frac{1}{4} \cdot L$$

#

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

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

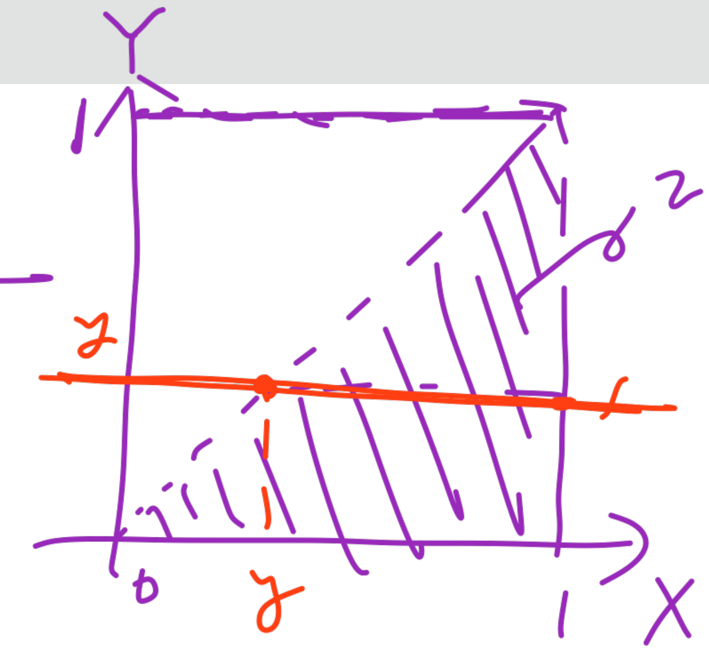
joint PDF

marginal PDF

- 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:

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

- $f_{X|Y}(x\,|\,y) = ?$ $\big(y \in (0,1),\ x \in (y,1) \big)$

$$= \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{2}{\int_y^1 f_{XY}(x,y)\,dx} = \frac{1}{1-y}$$

$$\int_y^1 2 \cdot dx = 2 \cdot (1-y)$$

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{1-y}, & \text{if } y \in (0,1),\ x \in (y,1) \\ 0, & \text{else} \end{cases}$$

Conditional Expectation: Continuous Case (Formally)

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

$$\underbrace{E[X \mid Y = y]} = \int_{-\infty}^{+\infty} x \cdot f_{X|Y}(x|y)\, dx$$

conditional PDF

function value

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

$$\sum x \cdot P(X = x \mid Y = y)$$

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

Example: Find Conditional Expectation

- Example:
$$f(x,y) = \begin{cases} 2 & , \text{ if } 0 < y < x < 1 \\ 0 & , \text{ otherwise} \end{cases}$$
- $E[X \mid Y = 0.5] = ?$

$$= \int_{-\infty}^{+\infty} (x)(f_{X|Y}(x \mid y=0.5))\,dx$$

$$= \int_{(0.5)}^{(1)} x \cdot (2)\,dx$$

$$= x^2 \Big|_{0.5}^{1} = \frac{3}{4}$$

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{1-y}, & \text{if } y \in (0,1) \\ & x \in (y,1) \\ 0, & \text{else} \end{cases}$$

$$\frac{1}{1-0.5} = 2$$

multivariate ...

2. Bivariate Normal Random Variables

Joint PDF of 2 Independent Normal R.V.s

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

- 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. → joint PDF = product of marginal PDF
  - What is the joint PDF?

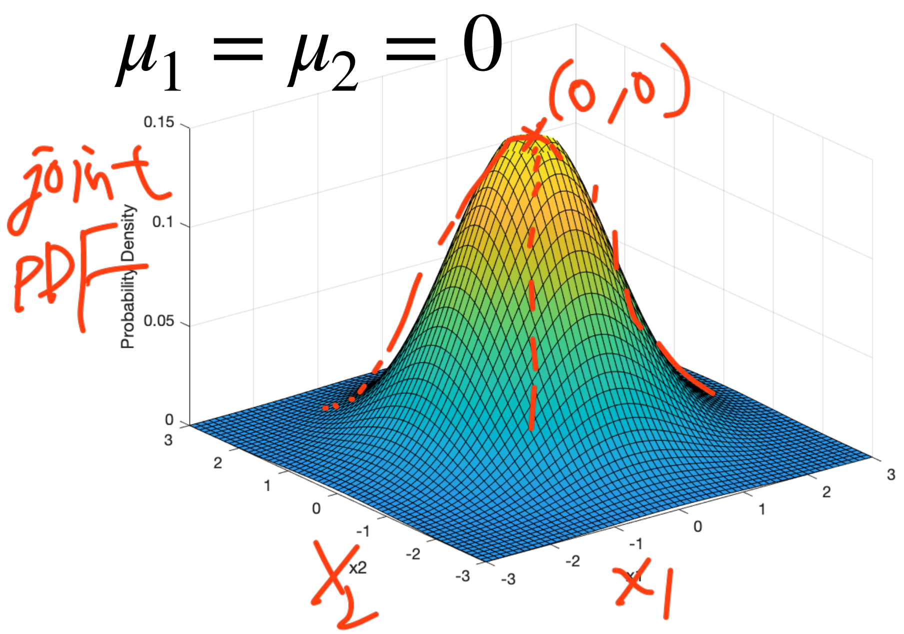
- **Joint PDF of 2 Independent Normal**:

$(\sqrt{2\pi}\sigma_1)\cdot(\sqrt{2\pi}\sigma_2)$

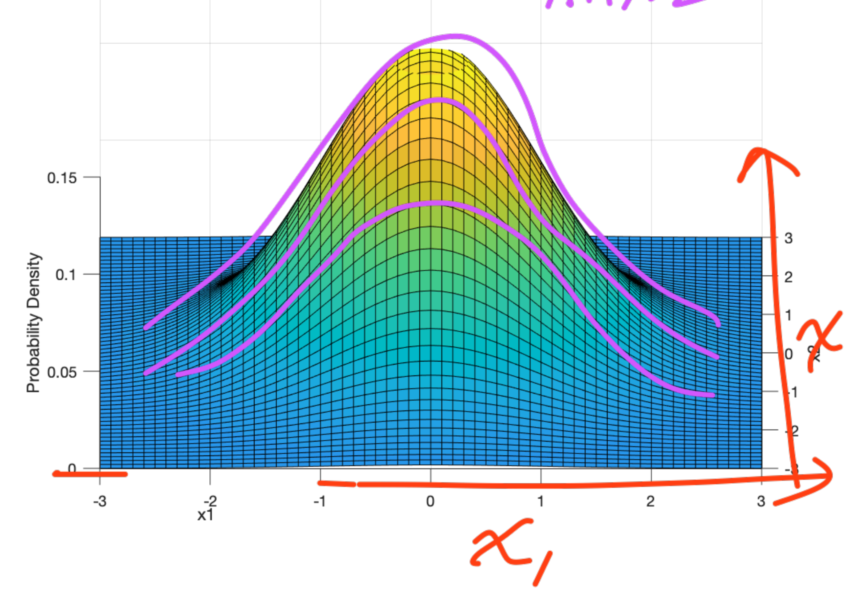
$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right]$$

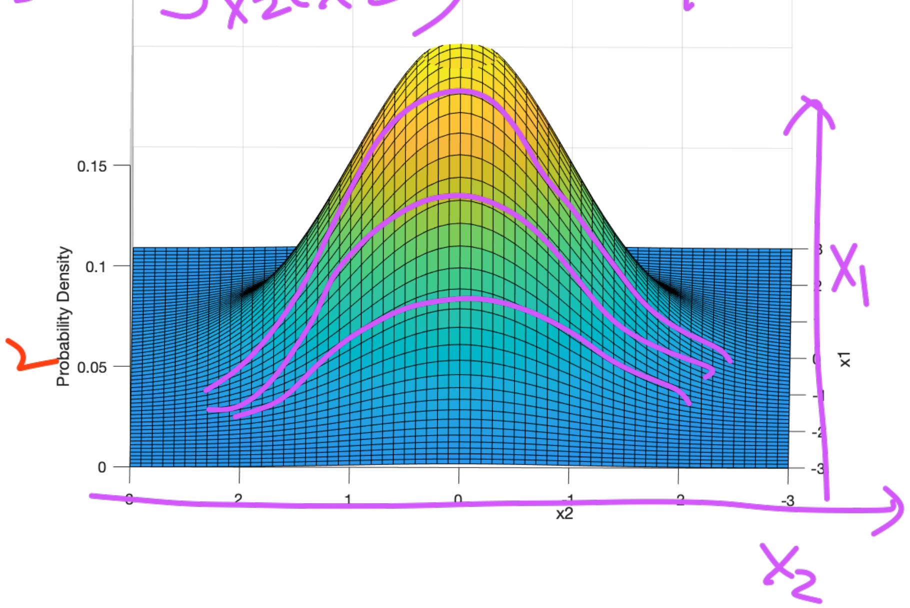
$\sigma_1 = \sigma_2 = 1$

$\mu_1 = \mu_2 = 0$



$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1X_2}(x_1,x_2)}{f_{X_2}(x_2)} = f_{X_1}(x_1)$





2 Independent Normal: Matrix Form

Review:
$$\Sigma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
$$\Sigma^{-1} = \frac{1}{\Delta}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$
主換次變 Δ

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

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

covariance matrix

$$\Sigma = \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix}$$

$$x =_2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \mu =_2 \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

mean vector

$$\det(\Sigma) = \sigma_1^2\sigma_2^2 - 0 \times 0 = \sigma_1^2\sigma_2^2$$

$$\underbrace{\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}}_{1\times 2}\underbrace{\begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix}}_{2\times 2}\underbrace{\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}}_{2\times 1}$$

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

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

How to extend the joint PDF for the case where X_1, X_2 are <u>NOT</u> independent?

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

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

where

covariance matrix

$$\Sigma = \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\cdot\sigma_1\cdot\sigma_2 \\ \rho\sigma_1\cdot\sigma_2 & \sigma_2^2 \end{bmatrix}$$

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

mean vector

$$\rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_1\sigma_2}$$

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

mean vector

covariance matrix

Bivariate Normal R.V.s: Alternative Expression

- **Joint PDF of Bivariate Normal:**

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

$$\Sigma = \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

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

$$\det(\Sigma) = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = (1-\rho^2)\sigma_1^2\sigma_2^2$$

$$\Sigma^{-1} = \frac{1}{\det(\Sigma)}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

$$= \frac{1}{1-\rho^2}\begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}$$

- **Alternative expression:**

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

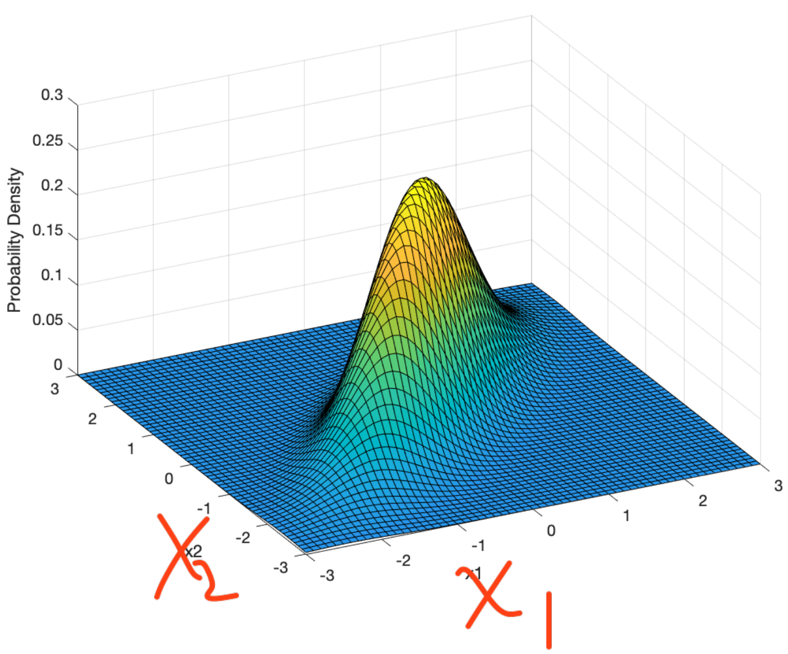
Plotting the Joint PDF Bivariate Normal

- **Joint PDF of Bivariate Normal**:

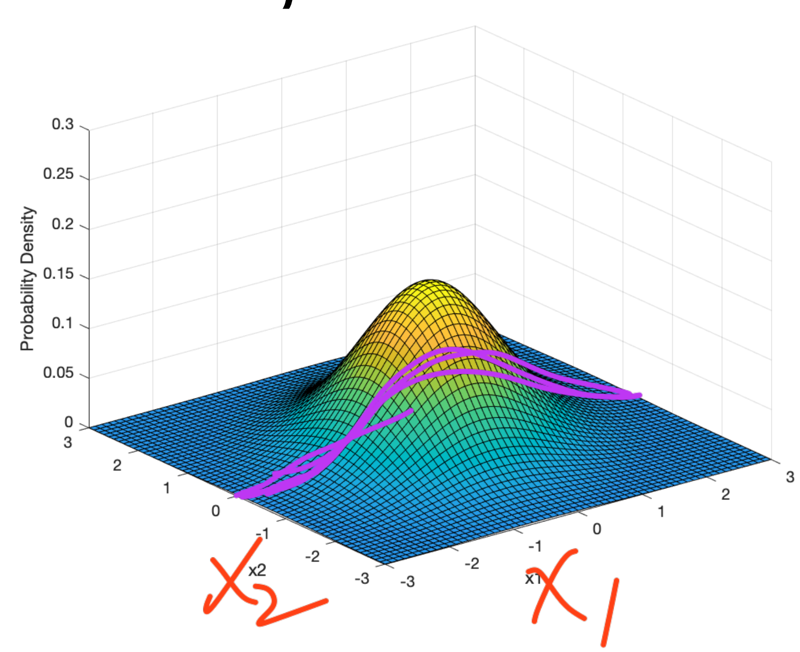
$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right)}{2(1-\rho^2)} \right]$$

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

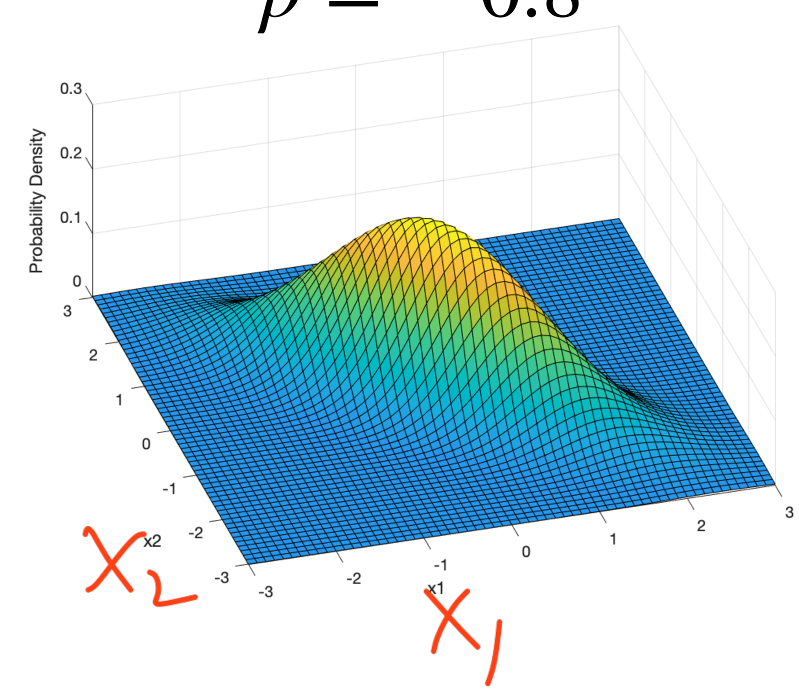
$\rho = 0.8$



$\rho = 0.2$



$\rho = -0.8$



X_1, X_2 Normal $\nRightarrow$ X_1, X_2 Bivariate Normal

$Y \sim N(0,1)$, $Z \sim N(0,1)$

100, 0, −100

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

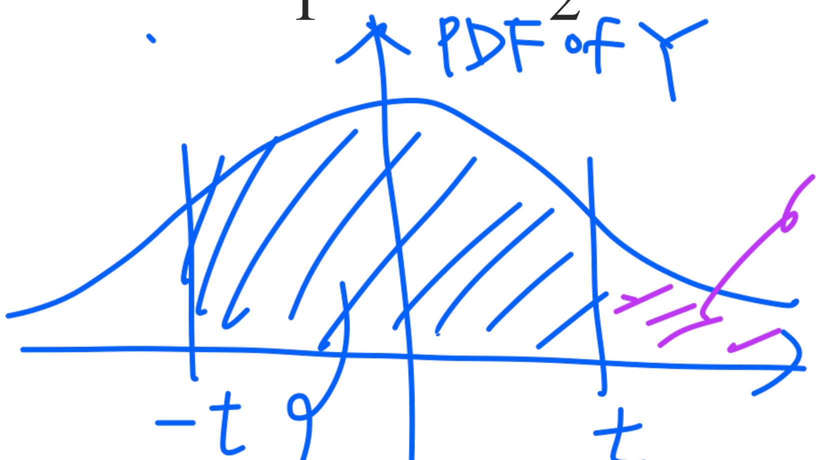
$$\text{sign}(Z) = \begin{cases} +1, & Z > 0 \\ 0, & Z = 0 \\ -1, & Z < 0 \end{cases}$$

$Y = 100$

$t > 0$ $\qquad$ $t \le 0$

$$P(X_1 \le t) = P(|Y| \cdot \text{sign}(Z) \le t)$$
$$= P(\text{sign}(Z) \le 0) + P(\text{sign}(Z) = 1 \text{ and } |Y| \le t)$$
$$= \frac{1}{2} + \frac{1}{2} \times P(|Y| \le t)$$
$$= \frac{1}{2} + \frac{1}{2} \cdot (2 \cdot \Phi(t) - 1) = \Phi(t)$$

- **Question**:
  - Are X_1 and X_2 normal? O
  - Are X_1 and X_2 bivariate normal? X



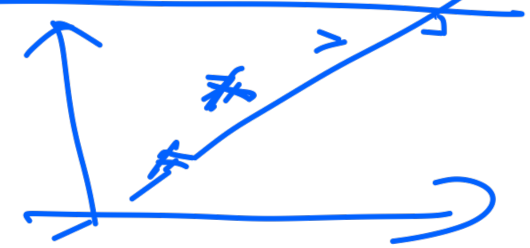
$1 - \Phi(t)$

$$1 - 2 \cdot (1 - \Phi(t)) = 2 \cdot \Phi(t) - 1$$

Applications of Bivariate / Multivariate Normal

parametric approach

- **Machine learning**: Regression / classification via Gaussian process
  - Non-parametric approach
  - https://www.youtube.com/watch?v=MfHKW5z-OOA (Nando de Freitas)



$f(x) = ax + b$
$= NN(x)$

state action noise

- **Control systems**: Linear dynamical systems
  - $x_{t+1} = Ax_t + Bu_t + w_k,\ w_k \sim \mathcal{N}(0,\Sigma)$
  - https://www.youtube.com/watch?v=bf1264iFr-w (Stephen Boyd)

- **Information theory and wireless communication**:
  - AWGN (additive white Gaussian noise) channels

Let’s study important properties of bivariate normal!

Properties of Bivariate Normal R.V.

- **If X_1, X_2 are bivariate normal, 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: $X_2 \,|\, X_1 = x_1 \sim \mathcal{N}\left(\mu_2 + \frac{\rho\sigma_2(x_1 - \mu_1)}{\sigma_1}, (1 - \rho^2)\sigma_2^2\right)$

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

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

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

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

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

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

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

$$f_{X_2}(x_2) =$$

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

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

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

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

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

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

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

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

(left blank intentionally for the proof)

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 $\rho = 0$**:

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

Extension: Multivariate Normal R.V.

- **Multivariate Normal Random Variables:** $X_1, \cdots, X_n$ are said to be multivariate normal random variables if the joint PDF of $X_1, \cdots, X_n$ is

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

where

$$\Sigma = \begin{bmatrix} \mathrm{Cov}(X_1, X_1)\cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1)\cdots & \mathrm{Cov}(X_2, X_n) \\ \cdots & \\ \mathrm{Cov}(X_n, X_1)\cdots & \mathrm{Cov}(X_n, X_n) \end{bmatrix}$$

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

There is still one remaining question:

Is it possible to construct
“bivariate normal” from “normal”?

Construction of Bivariate Normal R.V.

- Idea: Let Z, W be 2 independent standard normal r.v.s and define

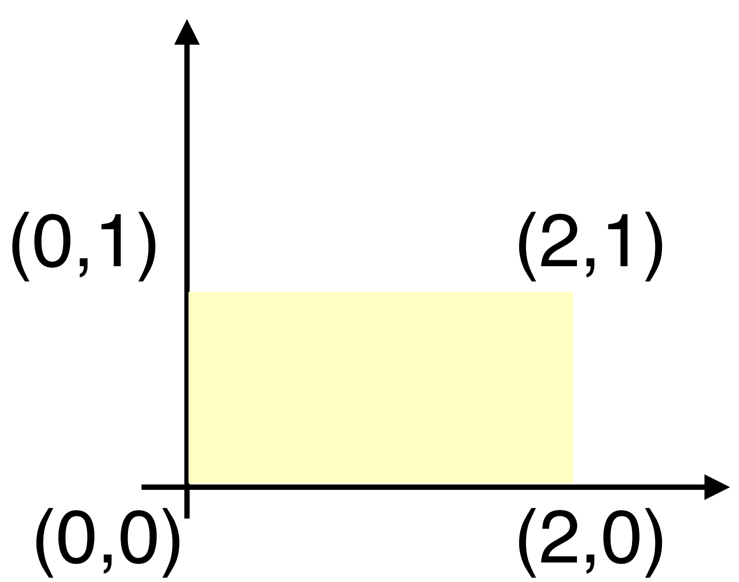
$$X_1 = \sigma_1 Z + \mu_1$$
$$X_2 = \sigma_2\left(\rho Z + \sqrt{1-\rho^2}W\right) + \mu_2$$

- Result: X_1, X_2 are bivariate normal with joint PDF

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

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**:



$$A = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$$

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

Joint PDF of X_1 and X_2

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

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

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

Quick Summary: Bivariate Normal

joint PDF (matrix form)

Marginal PDF

Conditional PDF

X_1, X_2 bivariate normal

uncorrelated vs independence?

Construction from 2 independent standard normal r.v.s

1-Minute Summary

1. Conditional Distributions

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

2. Bivariate Normal Random Variables

- Joint PDF
- X_1, X_2 normal $\not\Rightarrow X_1, X_2$ bivariate normal
- 4 properties: marginal PDF / conditional PDF / ρ / uncorrelated
- Construction from 2 independent standard normal r.v.s