1179: Probability
Lecture 21 — Bivariate Normal

Ping-Chun Hsieh (謝秉均)

November 26, 2021

Announcements

- No class next Wednesday (12/1, Sports Day)

Quick Review

X, Y

- Conditional PMF and PDF?

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

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

- LOTUS for two random variables

X, Y

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

- LIE (law of iterated expectation)

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

- X, Y independent $\Leftrightarrow$ $E[XY] = E[X]E[Y]$?

✓

$$X = \begin{cases} 0 \\ -1 \\ 1 \end{cases}$$

$Y = |X|$

This Lecture

1. Bivariate Normal Random Variables

- Reading material: Chapter 10.5

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

$a_i = t \cdot b_i$

- Recall: Cauchy inequality in high school

$a_1, \cdots, a_n \in \mathbb{R}$
$b_1, \cdots, b_n \in \mathbb{R}$

$$\underbrace{\left((a_1)^2 + \cdots + (a_n)^2\right)}_{\text{平方和}} \underbrace{\left((b_1)^2 + \cdots + (b_n)^2\right)}_{\text{平方和}} \geq \underbrace{\left((a_1)(b_1) + \cdots + (a_n)(b_n)\right)^2}_{\text{乘積和}}$$

- **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

$tX + Y = 0$ ✗

$Y = 0$ w.p. 1

$E[Y^2] = 0$, $E[XY] = 0$, $E[X^2] \neq 0$

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

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

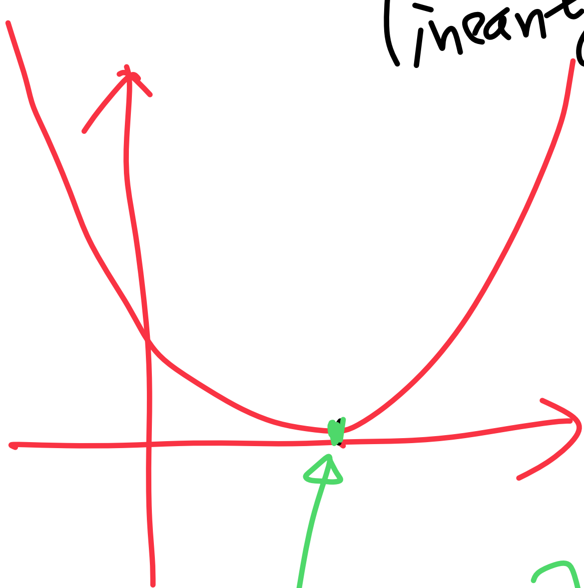
t: arbitrary real number

- **Proof:**

$$0 \leq E[(tX+Y)^2] = E[t^2X^2 + 2tXY + Y^2]$$

linearity

$$= \boxed{\underbrace{(E[X^2])}_{\geq 0} t^2 + (E[2XY])t + \underbrace{(E[Y^2])}_{\geq 0}}$$



$$t = \frac{E[XY]}{E[X^2]}$$

$\Rightarrow$ We must have

$$(E[2XY])^2 - 4 \cdot E[X^2] \cdot E[Y^2] \leq 0$$

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

$X \sim \mathcal{N}(0,1)$

$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x^2}{2}\right)}, \quad \forall x \in \mathbb{R}$

Bivariate Normal Random Variables

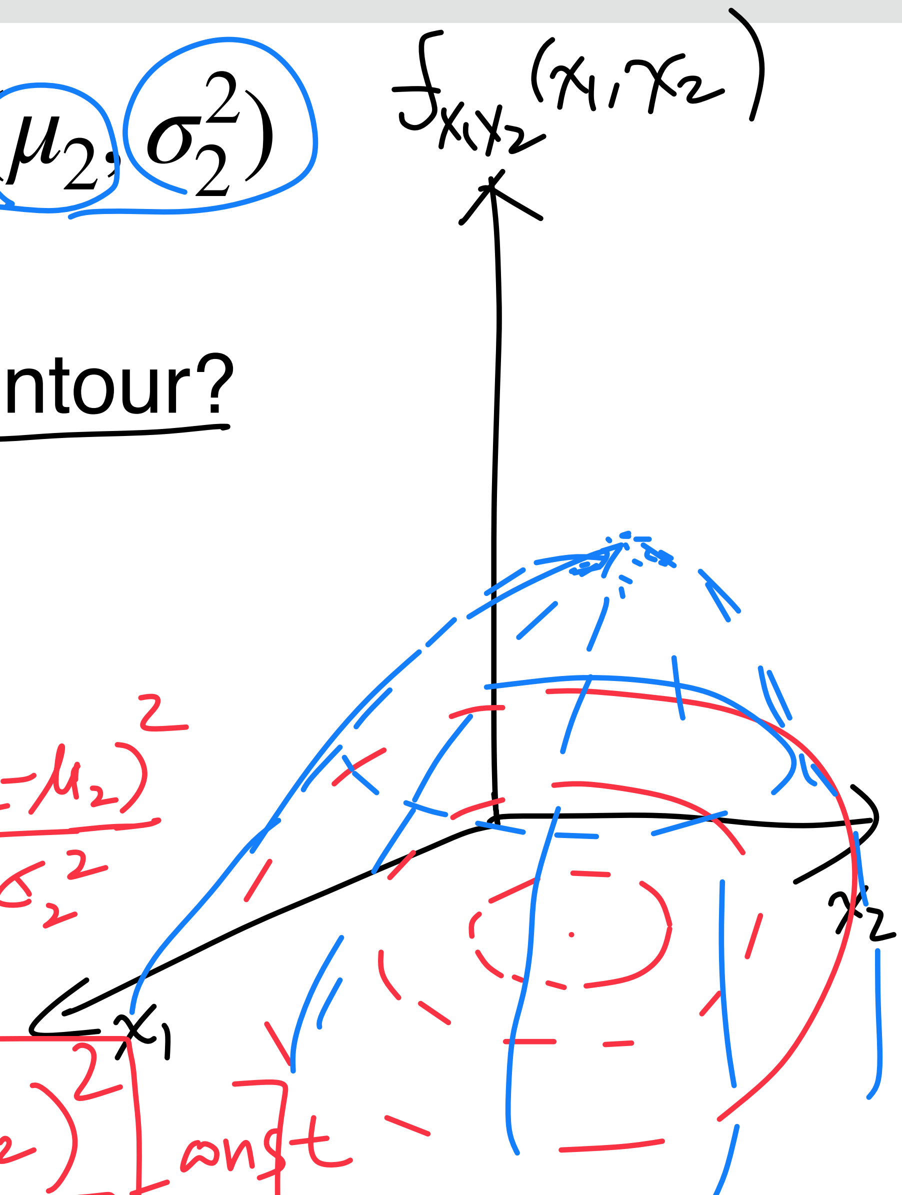
Example: 2 Independent Normal Random Variables

- **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?

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

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_1-\mu_1)^2}{2\cdot\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x_2-\mu_2)^2}{2\cdot\sigma_2^2}}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\left[\boxed{\frac{(x_1-\mu_1)^2}{2\sigma_1^2} + \frac{(x_2-\mu_2)^2}{2\sigma_2^2}} = \text{const}\right]}$$



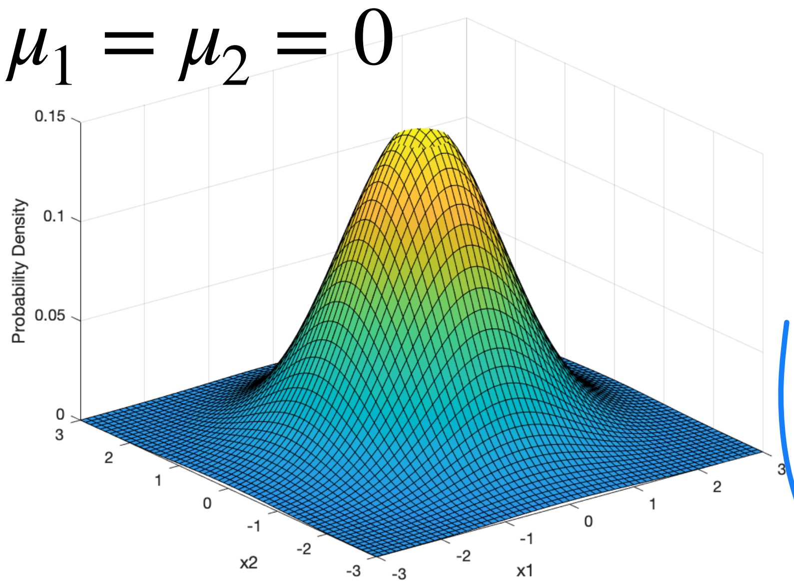
Joint PDF of 2 Independent Normal R.V.s (Formally)

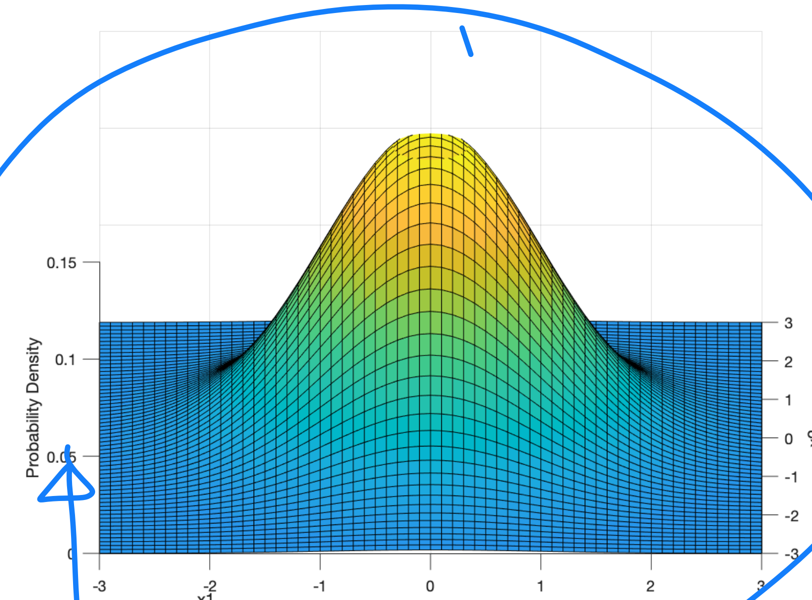
- Given: $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
- Suppose X_1, X_2 are independent.

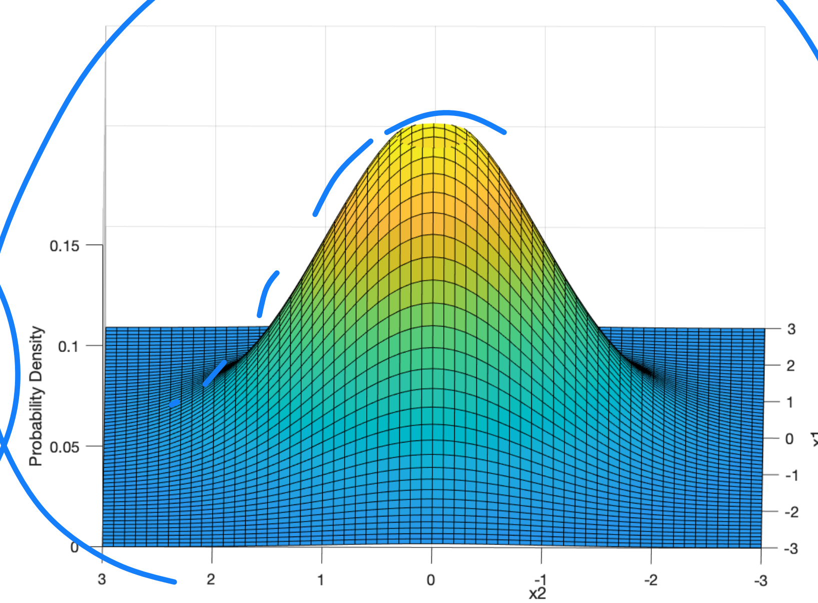
- **Joint PDF of 2 Independent Normal:**

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

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







2 Independent Normal: Matrix Form

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

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

$$\Sigma=\begin{bmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{bmatrix} \rightarrow \det(\Sigma)=\sigma_1^2\cdot\sigma_2^2$$

"covariance matrix"

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

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

"mean vector"

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

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

<u>One natural question:</u>

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

Construction of Bivariate Normal R.V.

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

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

$$X_1 = \sigma_1 Z + \mu_1 \longrightarrow X_1 \sim N(\mu_1, \sigma_1^2)$$

$$X_2 = \sigma_2\left(\rho Z + \sqrt{1-\rho^2}W\right) + \mu_2 \longrightarrow X_2 \sim N(\mu_2, \sigma_2^2)$$

E

$0.5Z + \sqrt{0.75}\,W$

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

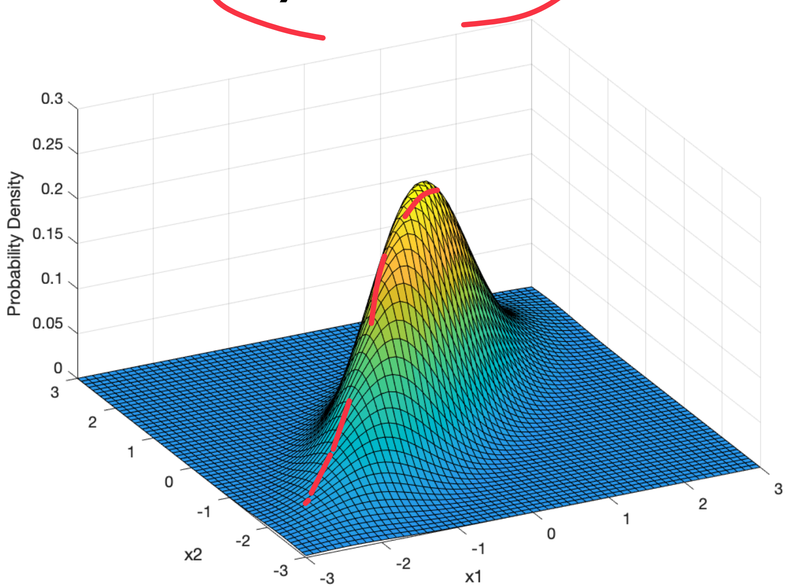
Plotting the Joint PDF Bivariate Normal

- **Joint PDF of Bivariate Normal**:

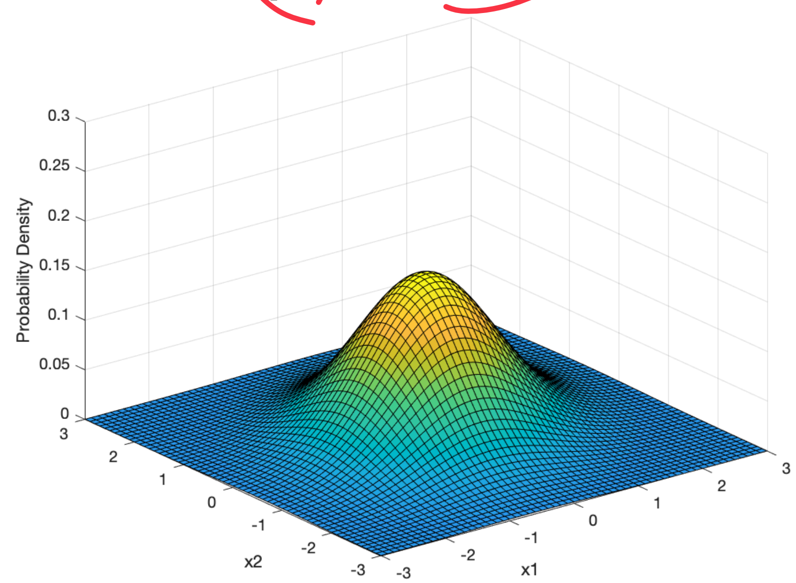
$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right)}{2(1-\rho^2)} \right]$$

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

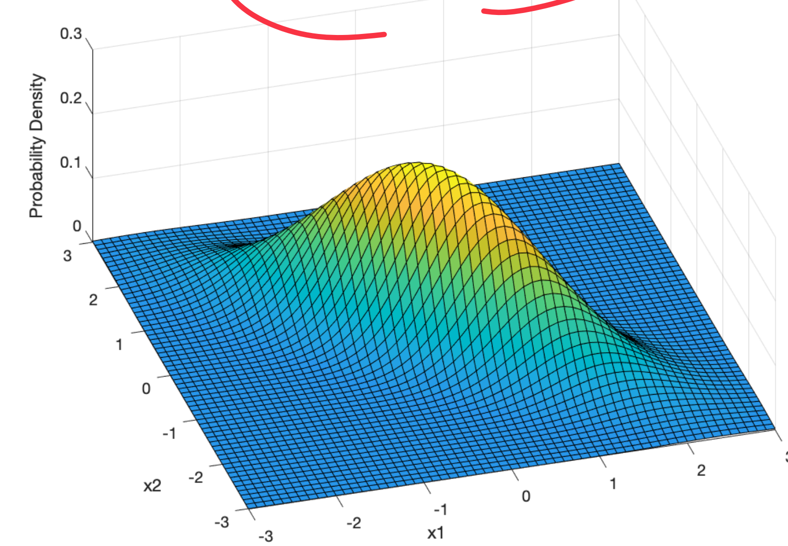
$\rho = 0.8$



$\rho = 0.2$



$\rho = -0.8$



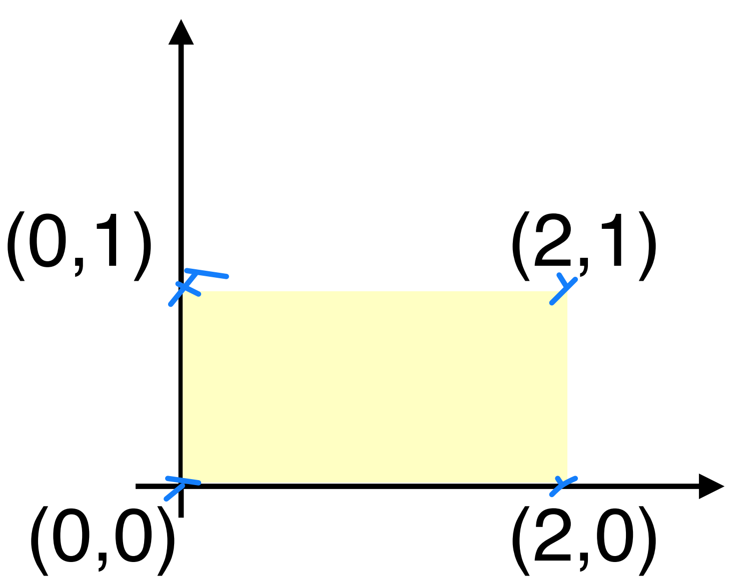
Linear Transformation of 2 Random Variables

- **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

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

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

- **Intuition**:



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

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)

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

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

Applications of Bivariate / Multivariate Normal

- **Machine learning —** e.g. Regression / classification / black-box optimization via Gaussian process
  - https://www.youtube.com/watch?v=MfHKW5z-OOA (Nando de Freitas)
- **Deep learning —** e.g. Variational autoencoder
  - https://www.youtube.com/watch?v=uaaqyVS9-rM (Ali Ghodsi)
- **Control systems —** e.g. 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)

There are still a few remaining questions:

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

Sum of independent random variables

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

Covariance

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

Conditional PDF and beyond

(Q1) Sum of Independent Random Variables and Moment Generating Functions (MGF)

$Z = X + Y$ and X, Y Independent — Discrete Case

- Question: X, Y are two <u>independent</u> discrete random variables.
  - Define $Z = X + Y$
  - What's the PMF of Z?

Convolution Theorem: Let X, Y be two independent discrete random variables with PMF $p_X(x)$ and $p_Y(y)$. Define $Z = X + Y$. Then, the PMF of Z is

$$p_Z(z) = P(Z = z) = \sum_x p_X(x) p_Y(z - x)$$

- Recall: $X \sim \text{Poisson}(\lambda_1, T)$ and $Y \sim \text{Poisson}(\lambda_2, T)$
  - What's the PMF of Z?

$Z = X + Y$ and X, Y Independent — Continuous Case

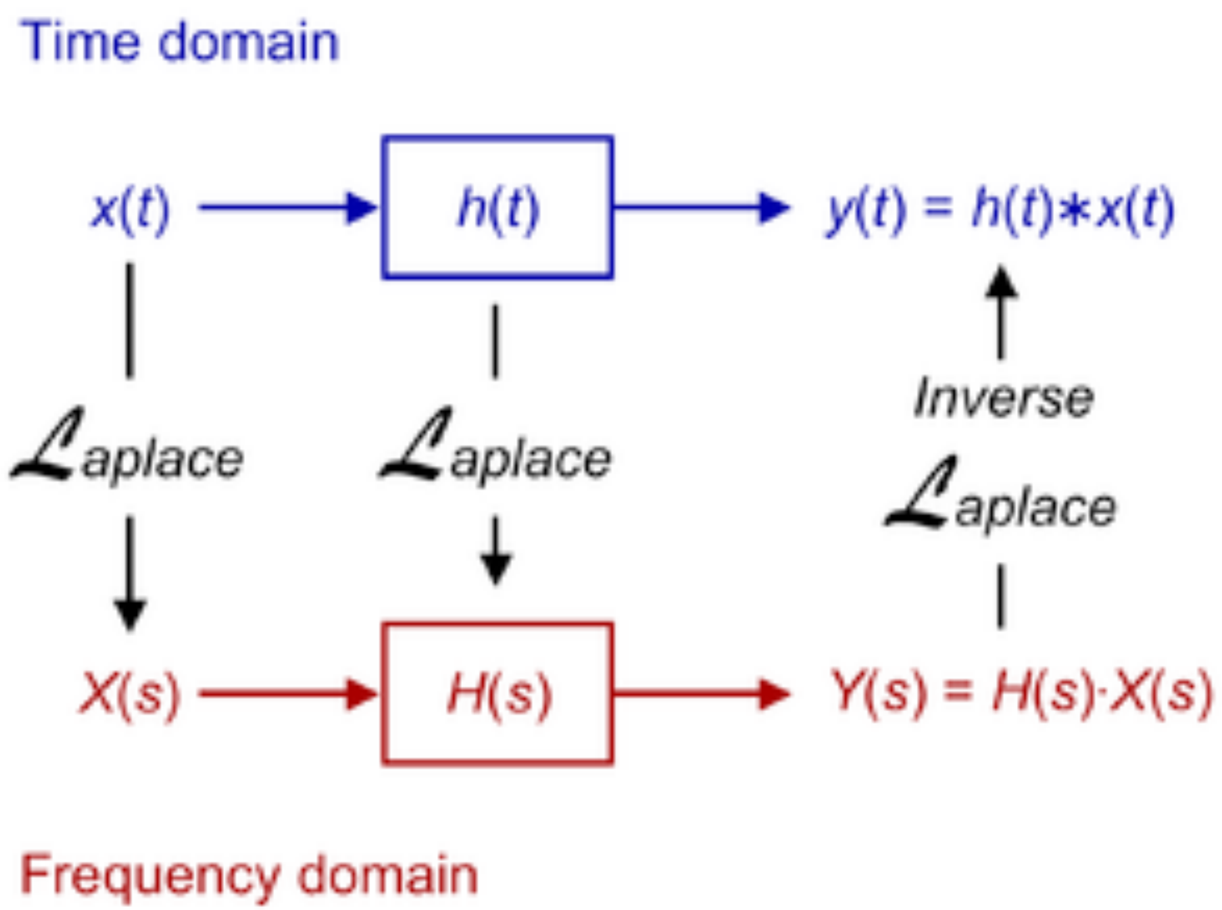
- For continuous random variables:

Convolution Theorem: Let X, Y be two continuous independent random variables with PDF f_1 and f_2. Define $Z = X + Y$. Then, the PDF of Z is

$$f_Z(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z - x)\, dx$$

Any Issue With Convolution Theorem?

- Issue: Sometimes it is quite tedious to do convolution
- Question: Any other approach?
- Idea: Borrow ideas from signal processing — Laplace transform



- In Probability, this is called "Moment Generating Function"

Moment Generating Function (Formally)

- **Moment Generating Function (MGF)**: For a random variable X, define
$$M_X(t) = E[e^{tX}],\ \ t \in \mathbb{R}$$
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
$$M_X(t) =$$
- Remark: If X is continuous with PDF $f_X(x)$, then
$$M_X(t) =$$