DCP 1206: Probability
Lecture 11 — Special Continuous Random Variables (II)

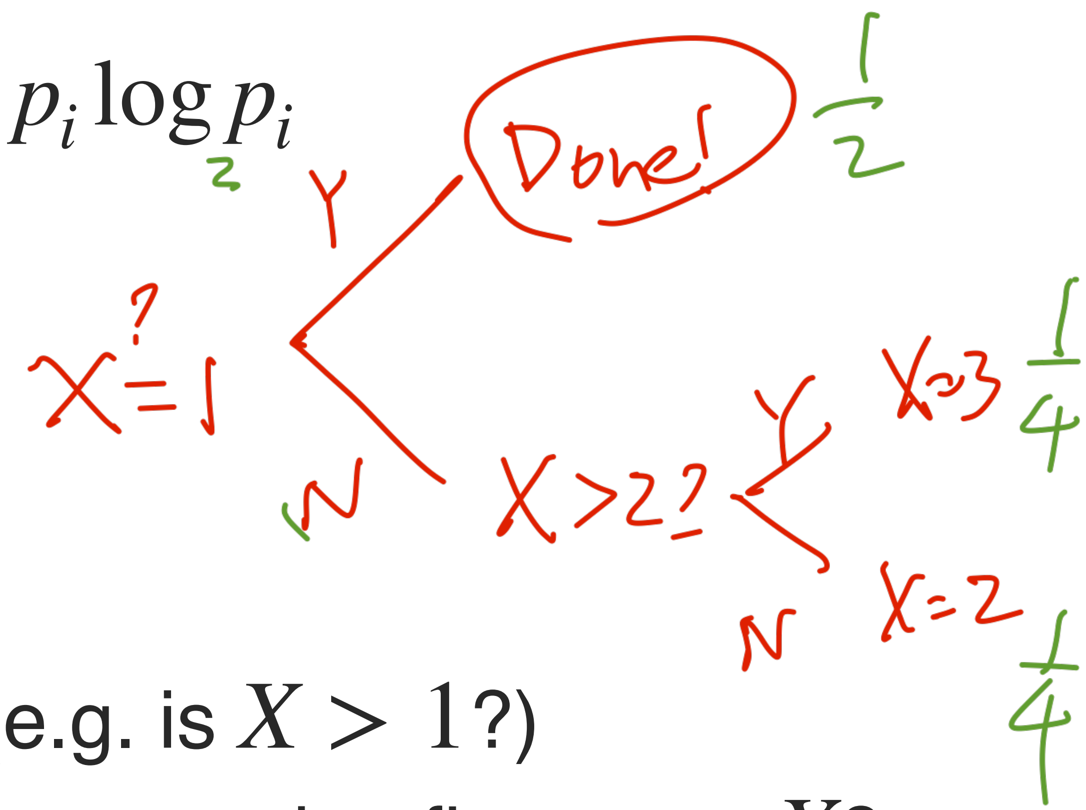
Ping-Chun Hsieh

October 23, 2019

About HW2: Entropy

$1 \times \frac{1}{2} + 2 \times \frac{1}{2} = 1.5$

- X has n possible values $\{1, 2, \cdots, n\}$ and $P(X = i) = p_i$
  - **Entropy** is defined as $H(X) := -\sum_{i=1}^{n} p_i \log_2 p_i$



- Example:

$$P(X = i) = \begin{cases} 1/2 & , \ i = 1 \\ 1/4 & , \ i = 2 \\ 1/4 & , \ i = 3 \end{cases}$$

  - Suppose we ask Yes / No questions (e.g. is $X > 1$?)
  - On average, how many questions do we need to figure out X?
  - What is $H(X)$?

$H(X) = -P_1 \log P_1 - P_2 \log P_2 - P_3 \log_3 = 1.5$

This Lecture

1. Normal and Mixture of Normal

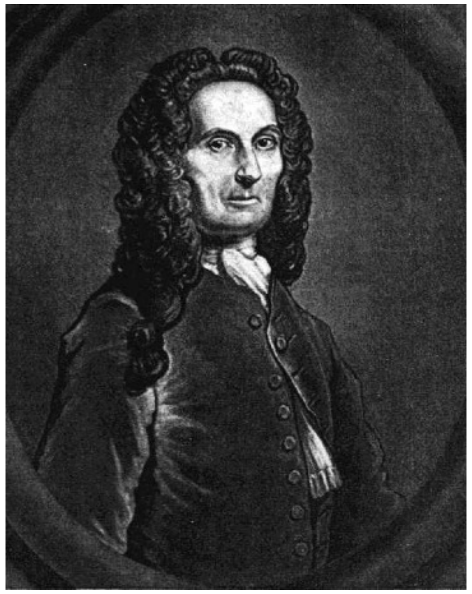
2. Exponential and Erlang Random Variables

- Reading material: Chapter 7.2-7.5

2. Standard Normal Random Variables

- Motivation: Consider $X \sim$ Binomial$(n, \frac{1}{2})$
  - Define $Y = \dfrac{X - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}}$. What is the CDF of Y vs n?

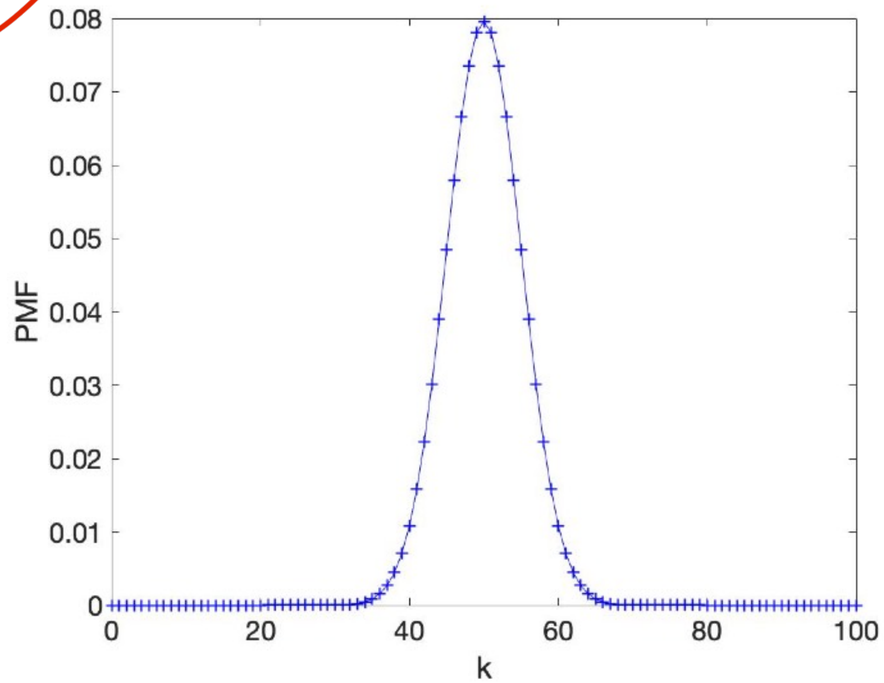
What's the intuition??



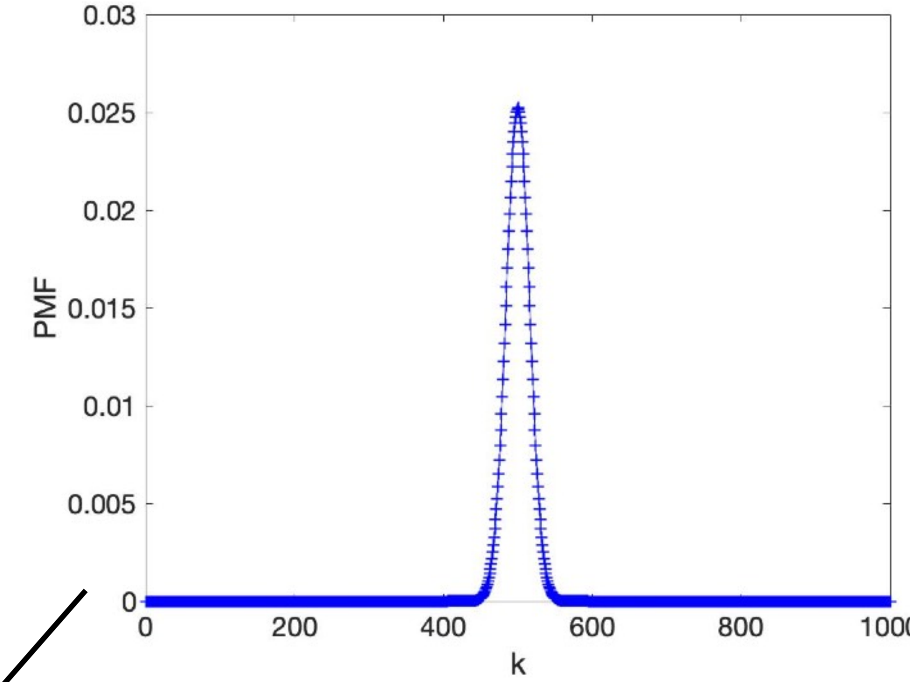
Abraham de Moivre

Plotting $X \sim$ Binomial$(n, 1/2)$

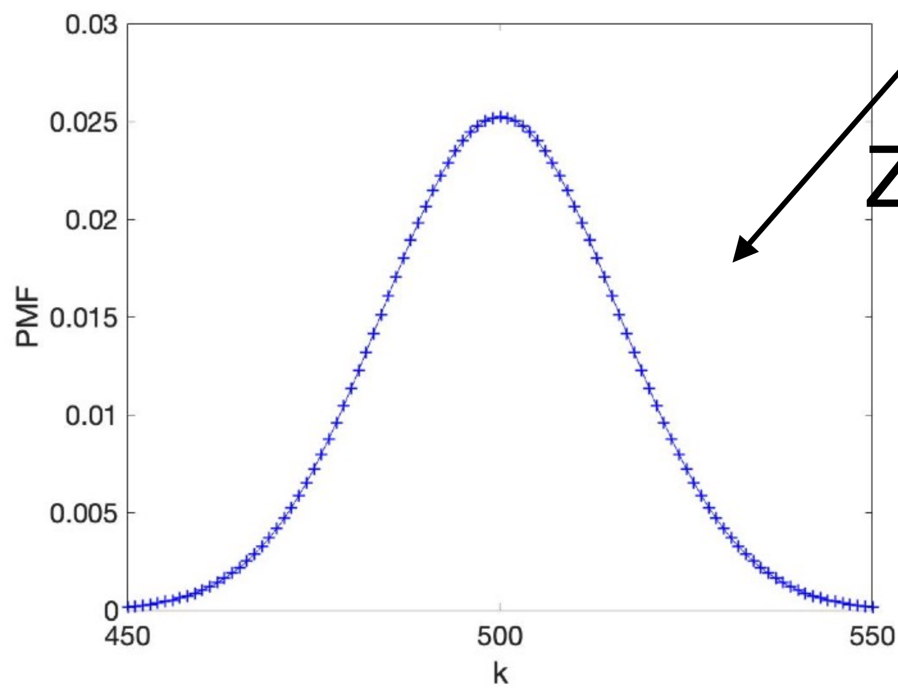
$n = 100$



$n = 1000$

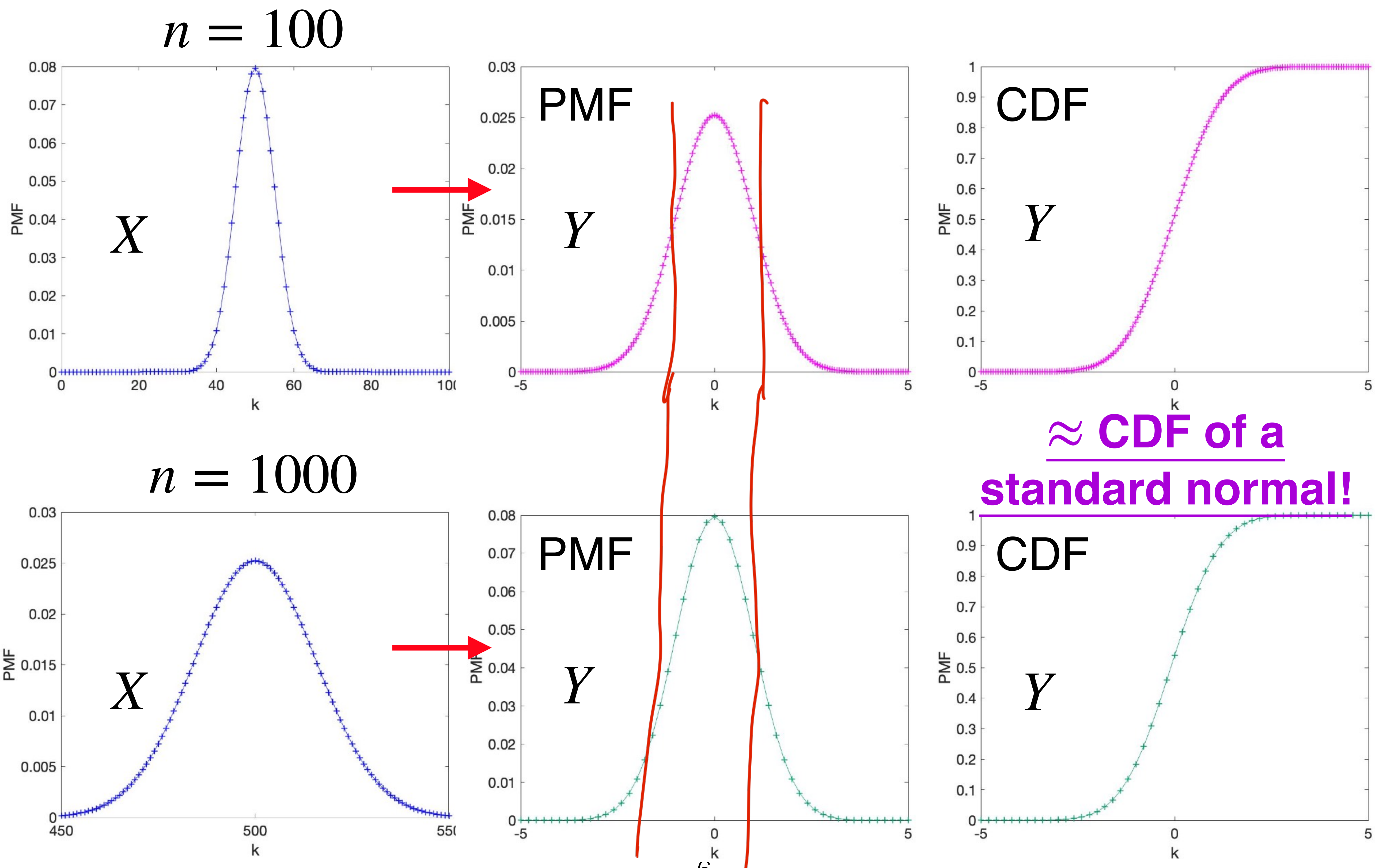


$n = 1000$



Zoom in to the range
$k \in [450, 550]$

Plotting $Y = (X - 0.5n)/(0.5\sqrt{n})$

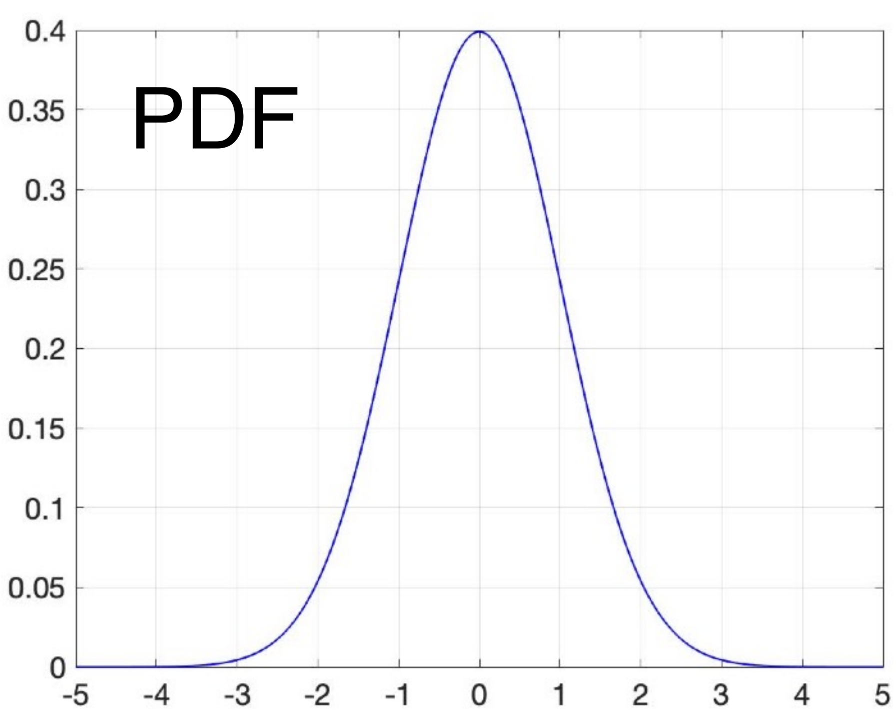


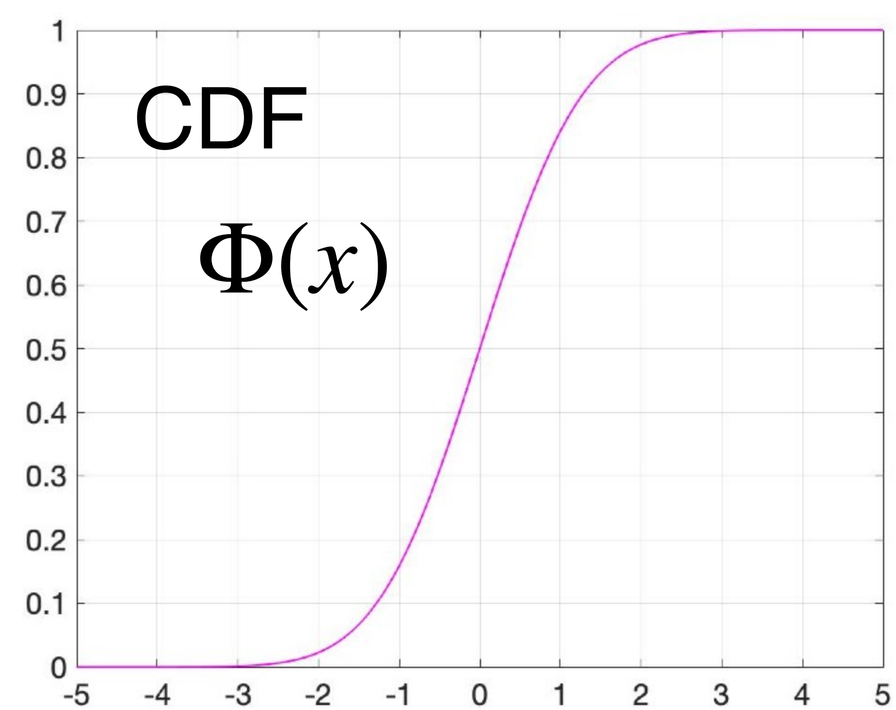
2. Standard Normal Random Variables (Formally)

Standard Normal Random Variables: A random variable X is called <u>standard normal</u> if its PDF is

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right), \text{ for all } x \in \mathbb{R}$$

- How to plot the PDF and CDF?





$$E[X] = 0$$

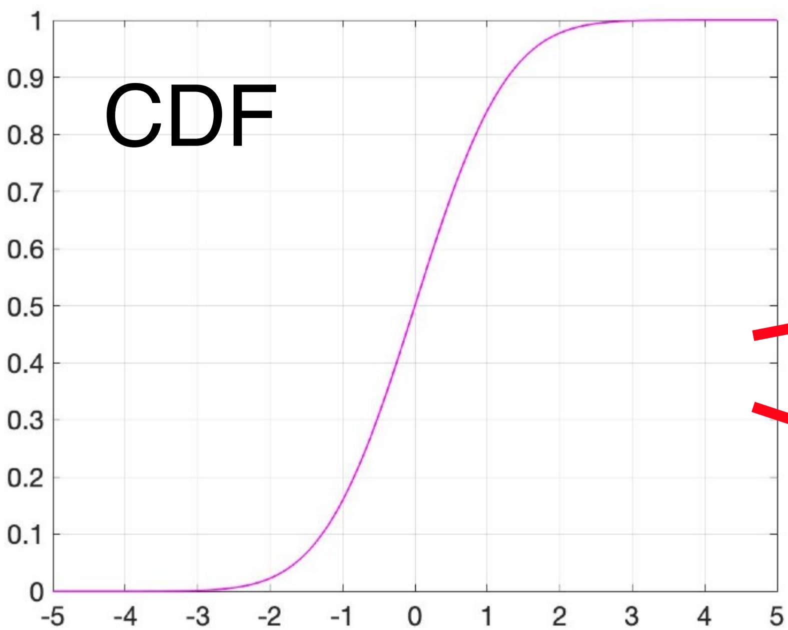
$$\text{Var}[X] = 1$$

From Standard Normal to Normal: CDF

- X is standard normal

$P(Y \le t) = P(X+5 \le t) = P(X \le t-5)$

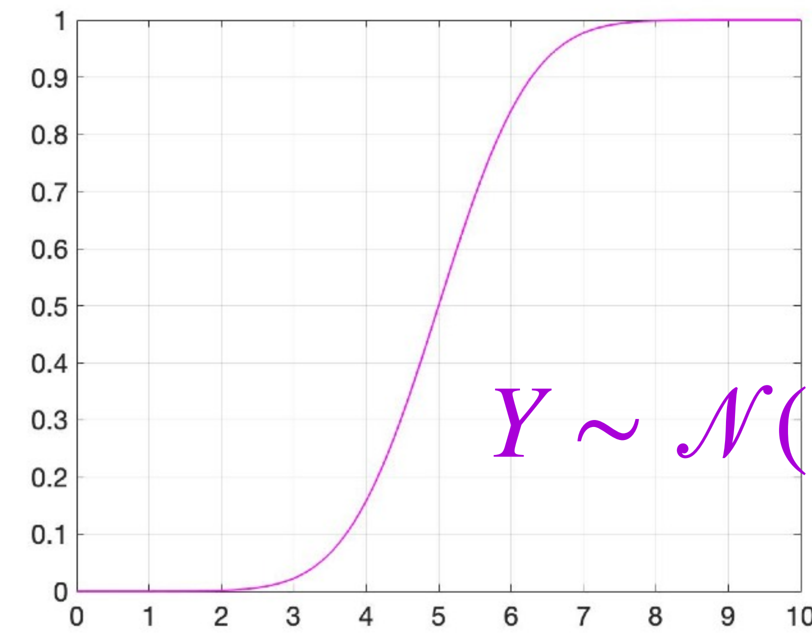
$= F_Y(t) = \Phi(t-5)$



$F_X(t) = \Phi(t)$

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

$Y = X + 5$

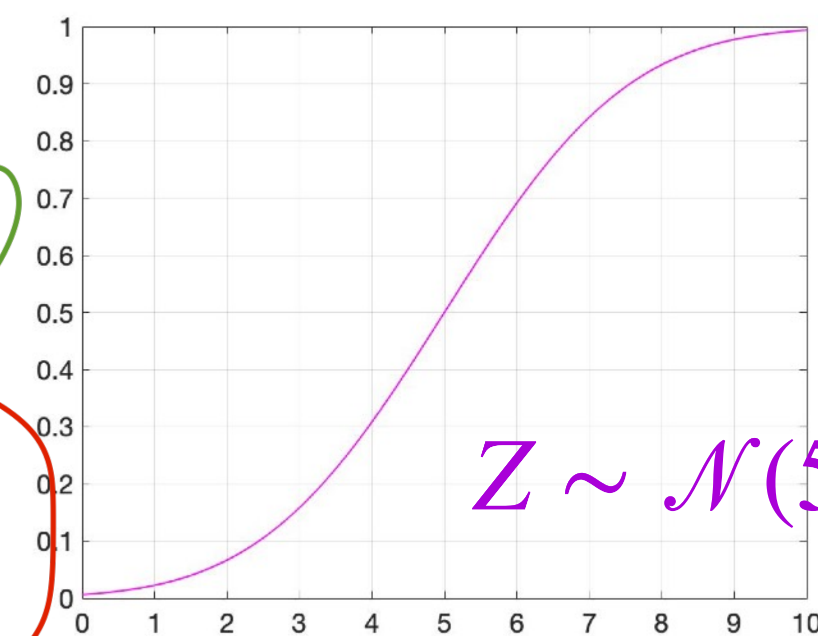


$Y \sim \mathcal{N}(5,1)$

$Z = 2X + 5$

$P(Z \le t) = P(2X+5 \le t)$

$F_Z(t) = \Phi\left(\frac{t-5}{2}\right)$

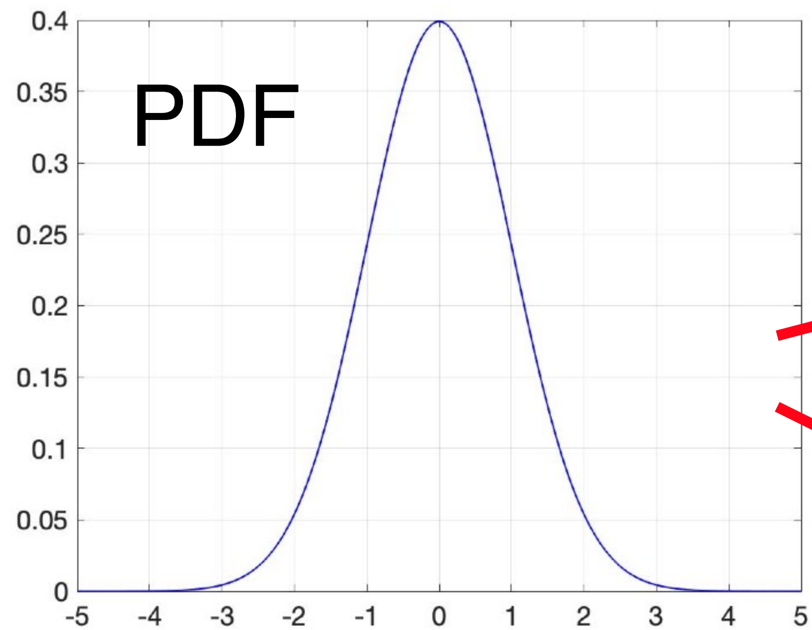


$Z \sim \mathcal{N}(5,2^2)$

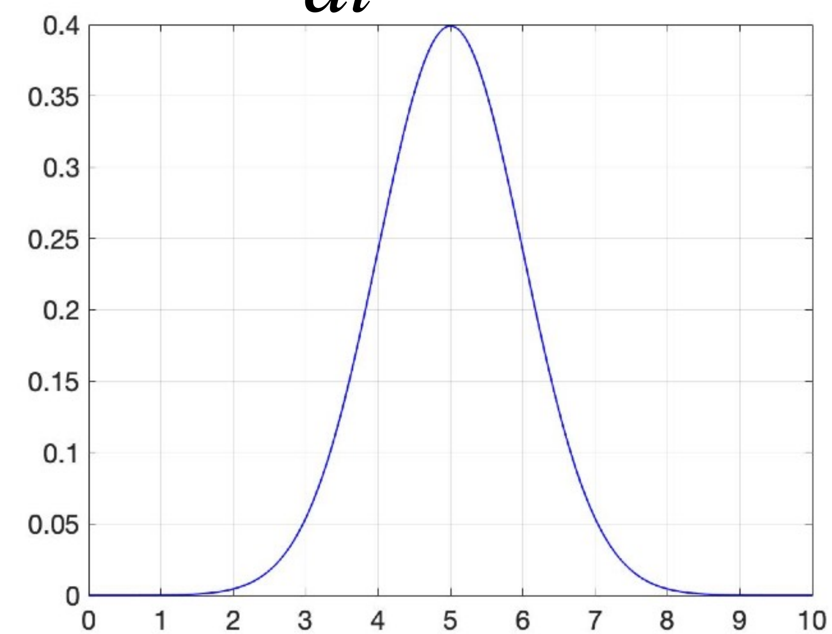
From Standard Normal to Normal: PDF

- X is standard normal

$$f_X(t) = \frac{d\Phi(t-5)}{dt} = \Phi'(t-5)$$

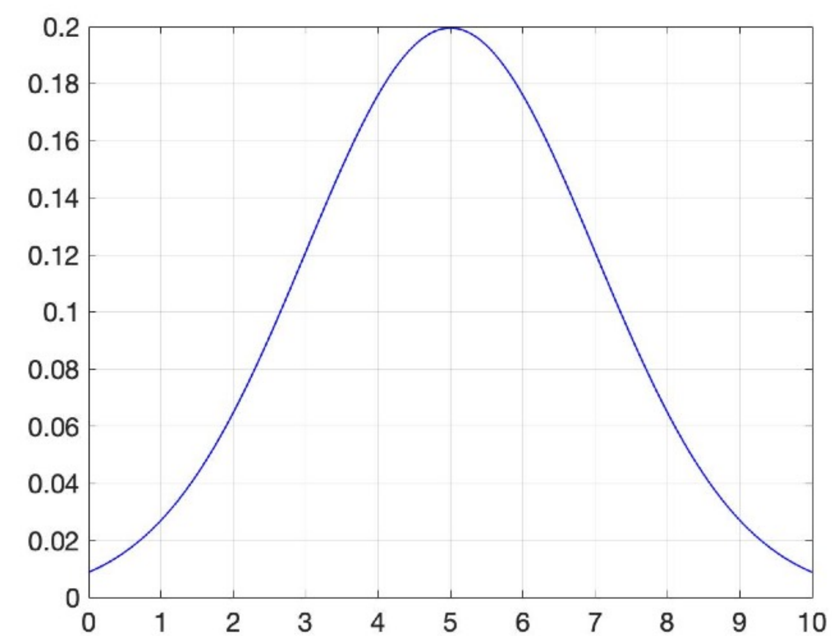


$$Y = X + 5$$



$$f_X(t) = \Phi'(t)$$

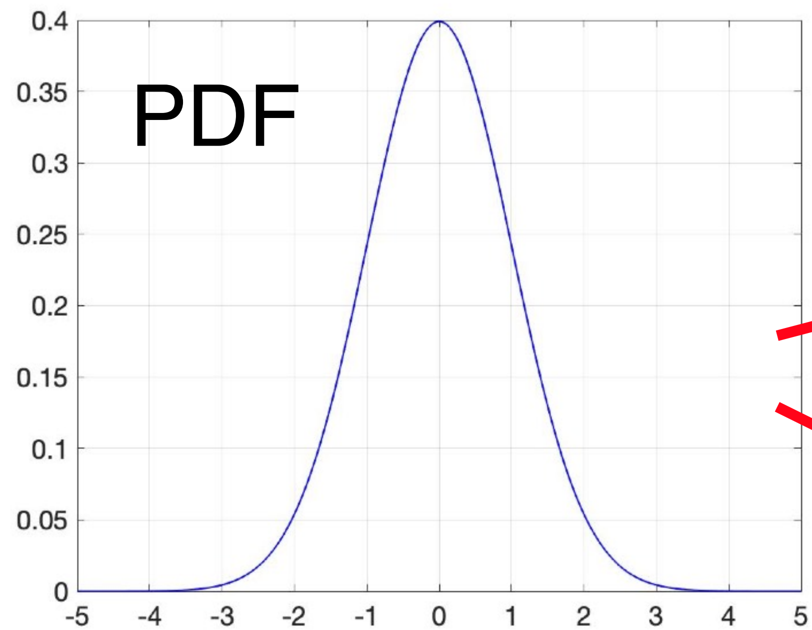
$$Z = 2X + 5$$



$$f_X(t) = \frac{d\Phi(\frac{t-5}{2})}{dt} = \frac{1}{2}\Phi'(\frac{t-5}{2})$$

From Standard Normal to Normal: PDF

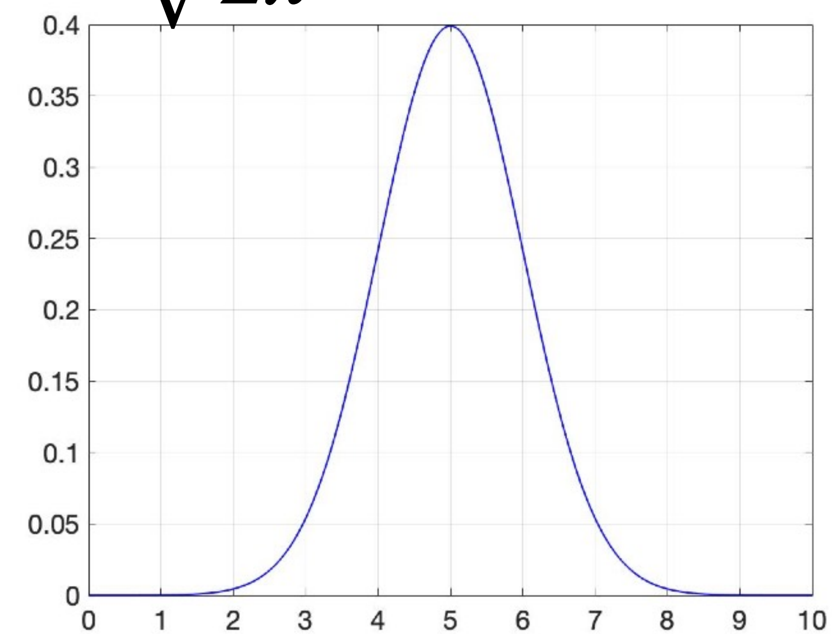
- X is standard normal



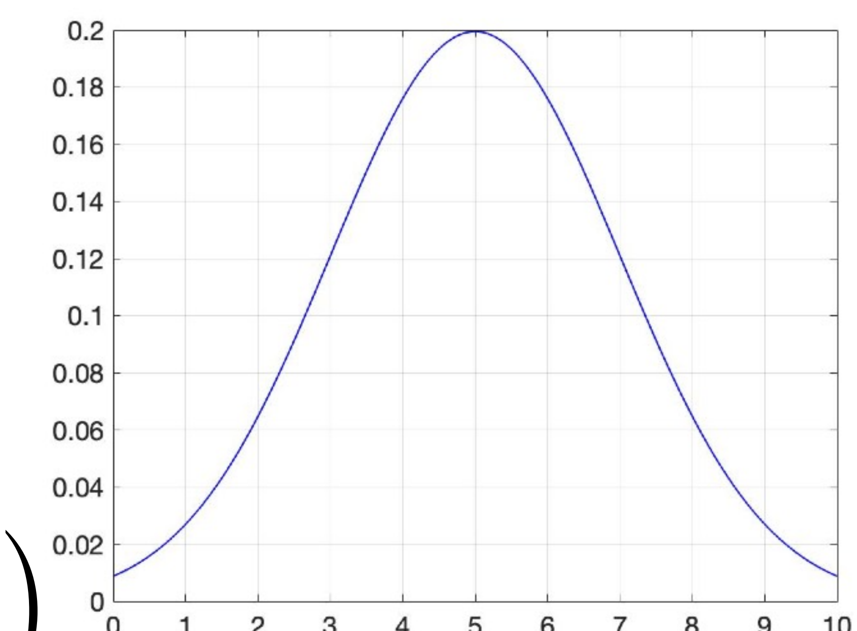
$$f_X(t) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right)$$

$$Y = X + 5$$

$$f_Y(t) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(t-5)^2}{2}\right)$$



$$Z = 2X + 5$$



μ

$$f_Z(t) = \frac{1}{2\sqrt{2\pi}} \exp\left(\frac{-(t-5)^2}{2 \cdot 2^2}\right)$$

σ σ^2

A General Recipe for Linear Transformation

- X is a continuous random variable
  - CDF: $F_X(t)$
  - PDF: $f_X(t) = \dfrac{dF_X(t)}{dt}$
- Consider $Y = aX + b,\ a, b \in \mathbb{R}, a \neq 0$
  - CDF $F_Y(t)$?
  - PDF $f_Y(t)$?
  - If $\underbrace{X \sim \mathcal{N}(0,1)}_{\text{standard normal}}$, then $F_Y(t)$ = ?

$$F_Y(t) = P(Y \le t) = P(aX + b \le t)$$

$$= P\left(X \le \frac{t-b}{a}\right)$$

$$= F_X\left(\frac{t-b}{a}\right)$$

$$f_Y(t) = \frac{dF_Y(t)}{dt} = \frac{d\left(F_X\left(\frac{t-b}{a}\right)\right)}{dt}$$

$$= \frac{1}{a} f_X\left(\frac{t-b}{a}\right)$$

$$\Phi\left(\frac{t-b}{a}\right)$$

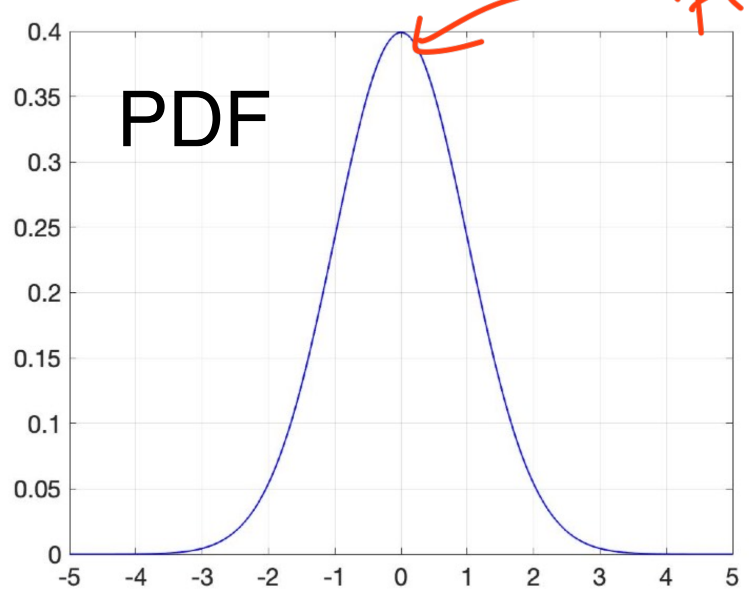
Example: Normal Distribution

- Example: Let $X \sim \mathcal{N}(-2,5)$
  - What is $P(|X| < 4)$?

Uni-Modal and Multi-Modal Distributions

- Normal distribution is useful for approximating uni-modal distributions
  - Uni-modal: only 1 local maximum

peak

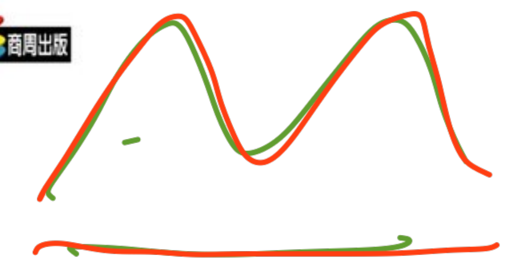


$$f_X(t) = \Phi'(t)$$

- How about multi-modal distributions?

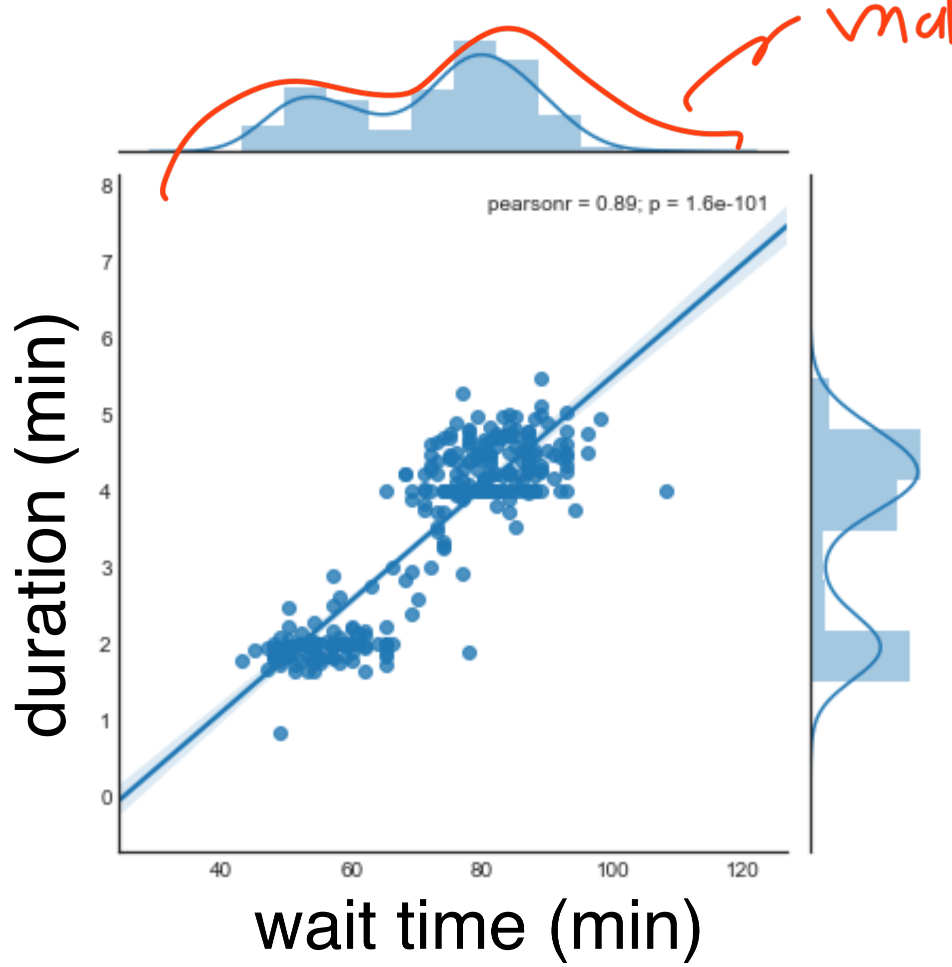
Examples: Multi-Modal Distributions

- Example: M-shaped income distribution



- Example: Old Faithful Geyser

multi-modal



http://www.stat.cmu.edu/~larry/all-of-statistics/=data/faithful.dat

3. Mixture of Normal (Formally)

Mixture of Normal: A random variable X is a mixture of normal with parameters $\{\mu_1, \sigma_1, \cdots, \mu_n, \sigma_n\}$ and weights $\{\alpha_1, \cdots, \alpha_n\}$ $(\alpha_1 + \cdots + \alpha_n = 1)$ if its PDF is

$$f_X(t) = \sum_{i=1}^{n} \alpha_i \cdot \Phi'\left(\frac{t - \mu_i}{\sigma_i}\right)$$

- How to plot the PDF?

$$\mu_1 = 5, \sigma_1 = 1$$
$$\mu_2 = 10, \sigma_2 = 1$$
$$\alpha_1 = 0.5, \alpha_2 = 0.5$$

2. Exponential Random Variables

Recall: Geometric Random Variables

- Suppose $X \sim \text{Geometric}(p)$
  - What is the PMF of X?
  - Memoryless property?

$$P(X=n) = (1-p)^{n-1} \cdot p$$

$$P(X=m+n \mid X>m) = P(X=n)$$

- Question: Is there a continuous counterpart of a geometric random variable?

3. Exponential Random Variables

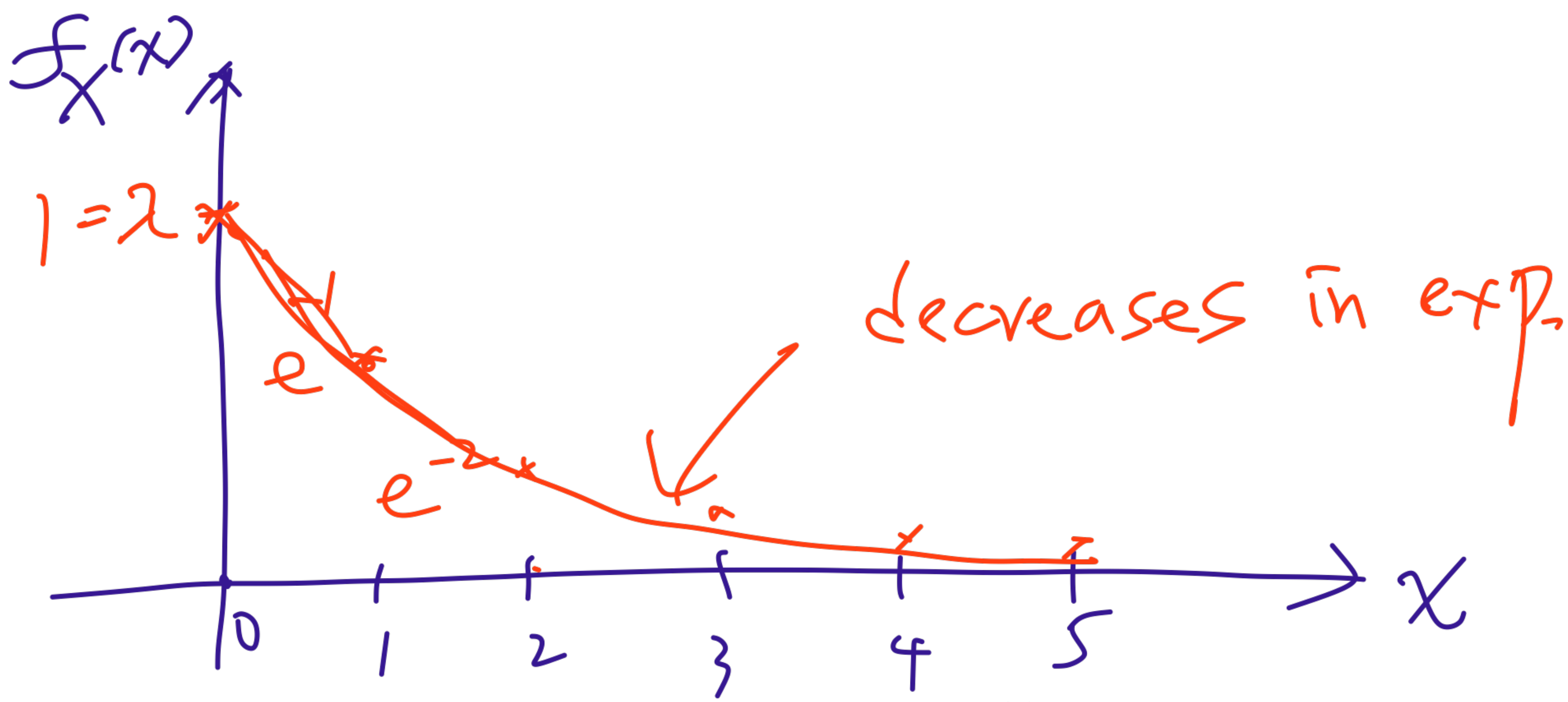
rate

Exponential Random Variables: A random variable X is exponential with parameters $\lambda > 0$ if its PDF is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{, if } x \geq 0 \\ 0 & \text{, otherwise} \end{cases}$$

- How to plot the PDF of Exp$(\lambda = 1)$?

$e^{-\lambda}$



3. Exponential Random Variables

PDF = $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{, if } x \geq 0 \\ 0 & \text{, otherwise} \end{cases}$

$P(X>t) = e^{-\lambda t}$

- What is the CDF of X?

$t>0$

$F_X(t) = P(X \leq t) = \int_0^t \lambda e^{-\lambda x} \, dx = \lambda \cdot \frac{1}{-\lambda} e^{-\lambda x} \Big|_0^t$

$= 1 - e^{-\lambda t}$

$F_X(t) = 0, \ \forall t < 0$

Memoryless Property

- Suppose $X \sim \text{Exp}(\lambda)$
  - What is $P(X > s + t \,|\, X > t)$?

$$= \frac{P(X > s+t \text{ and } X > t)}{P(X>t)} = \frac{\overbrace{P(X>s+t)}^{e^{-\lambda(s+t)}}}{\underbrace{P(X>t)}_{e^{-\lambda t}}}$$

$$= e^{-\lambda s} = P(X>s)$$

Example: Nokia 3310

- Example: Suppose the lifetime of a Nokia 3310 is an exponential random variable with mean = 10 years. $= \frac{1}{\lambda}$
  - Suppose a Nokia 3310 was bought 15 years ago.
  - P(it will last another 5-10 years)?

$(\lambda = \frac{1}{10})$

X = total lifetime

$$P(15+5 \le X \le 15+10 \mid X > 15)$$

$$= P(5 \le X \le 10)$$

$$= F_X(10) - F_X(5)$$

Exponential Distribution: A Good Model for Occurrence of Events

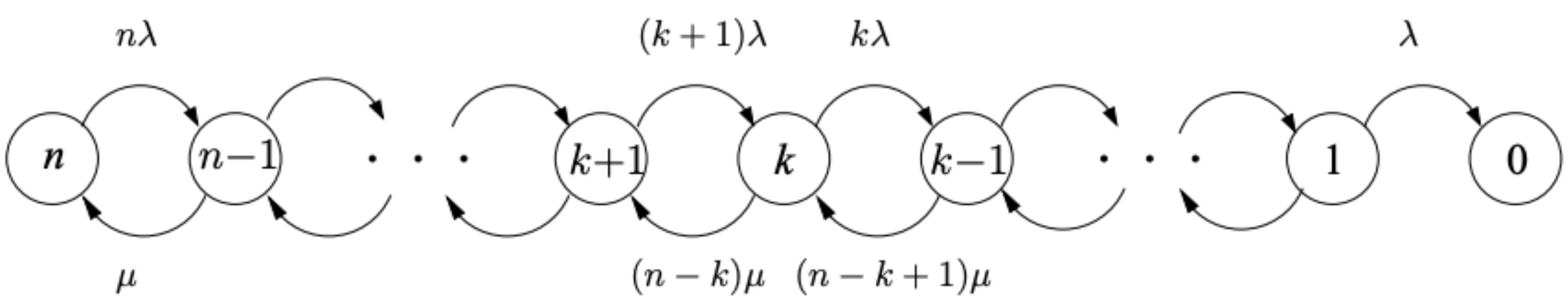
- Communication networks: Inter-arrival time between two data packets

- Survival analysis: User's lifetime (App, social network…)

- Reliability modeling: Amount of time until the hardware on AWS EC2 fails

Example: Reliability Modeling

- Replication in a distributed system:
  - At each time, the system maintains at most N replicas
  - Machines can join or leave the system
  - The time that a machine stays in the system $\sim$ Exp(λ)
  - The time required for creating a new replica $\sim$ Exp(μ)



http://www1.cs.columbia.edu/~danr/courses/6772/Fall06/papers/longrun.pdf

Mean and Variance

Integration by part

$$\int u\,dv = uv - \int v\,du$$

- Suppose $X \sim \text{Exp}(\lambda)$
  - How to get $E[X] = 1/\lambda$?
  - How to get $\text{Var}[X] = 1/\lambda^2$?

$$E[X] = \int_0^\infty \underbrace{P(X>t)}_{= e^{-\lambda t}}\,dt = \frac{1}{-\lambda} e^{-\lambda t}\Big|_0^\infty = \frac{1}{\lambda}$$

$$= \int_0^\infty \underline{x} \cdot \underbrace{\lambda e^{-\lambda x}}_{= \text{PDF}}\,dx$$

$$E[X^2] = \int_0^\infty \underbrace{x^2}_{= u}\, \underbrace{\lambda e^{-\lambda x}\,dx}_{= dv}$$

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

Example

- Example: Consider $X \sim \text{Exp}(\lambda = 2)$
  - What is $P(|X - \mu_X| \geq \sigma_X)$?

$\mu_X = \frac{1}{2}$

$\sigma_X^2 = \frac{1}{2^2} \Rightarrow \sigma_X = \frac{1}{2}$

$$P\left(\left|X - \tfrac{1}{2}\right| \geq \tfrac{1}{2}\right) = P(X \geq 1 \text{ or } X \leq 0)$$

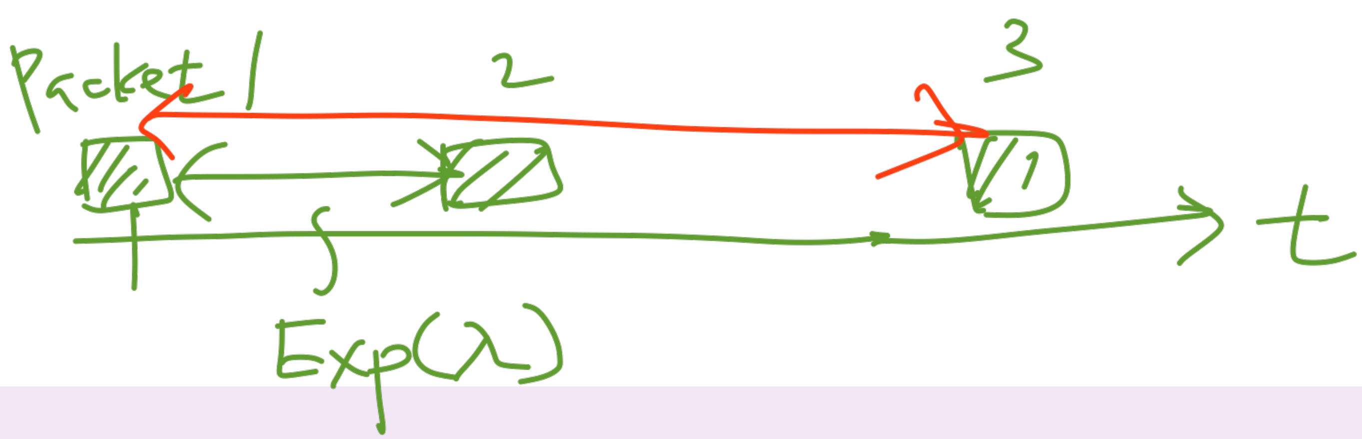
$$= \underbrace{P(X \geq 1)}_{\text{tail probability}} + \underbrace{P(X \leq 0)}_{=0}$$

tail probability $= e^{-\lambda t} = e^{-2}$

3. Erlang Random Variables

Sum of Exponential Random Variables

- Example: Suppose the inter-arrival time of packets is exponential with parameter λ
  - How long does it take to see 2 arriving packets?



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

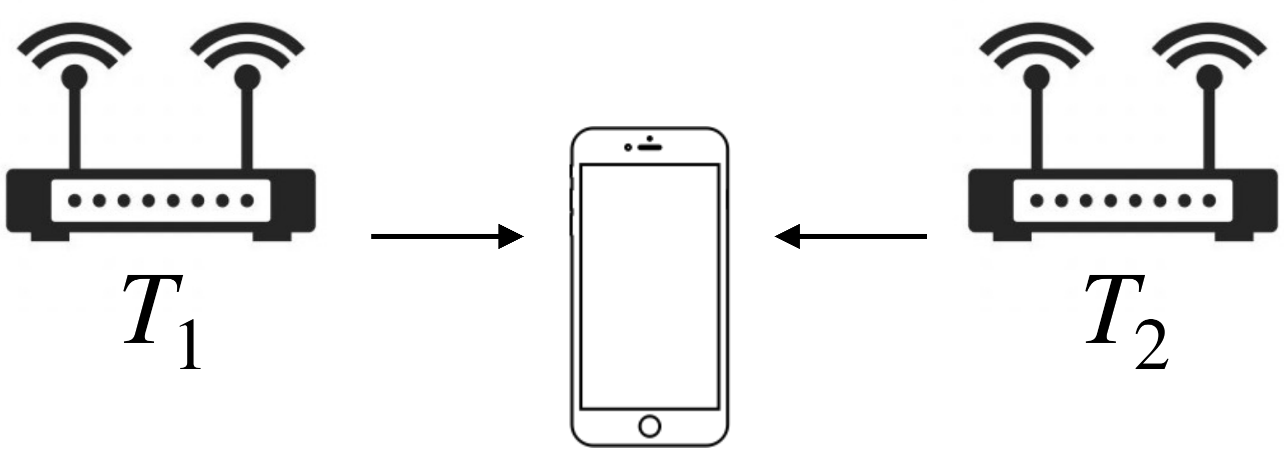
$$f_Z(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z - x) dx$$

Convolution Theorem vs HW2 Problem 2(c)

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

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

HW2 Problem 2(c):



$$P(k \text{ bits received}) = \sum_{m=0}^{k} P(T_1 \text{ sends } m \text{ bits}) \cdot P(T_2 \text{ sends } k - m \text{ bits})$$

Sum of Exponential Random Variables (Cont.)

- Example: Suppose the inter-arrival time of packets is exponential with parameter λ
  - How long does it take to see 2 arriving packets?

$X \sim Exp(\lambda)$

$Y \sim Exp(\lambda)$

$Z = X + Y$

PDF of Z = ?

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

$$f_1(x) = \begin{cases} \lambda e^{-\lambda x} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

$f_2(x)$ "

$$f_Z(z) = \int_0^z \underbrace{(\lambda e^{-\lambda x}) \cdot (\lambda e^{-\lambda(z-x)})}\, dx$$

$$= \lambda^2 \cdot \int_0^z e^{-\lambda} dx = z\lambda^2 e^{-\lambda}$$

4. Erlang Random Variables

- Question: Sum of n independent exponential r.v.'s?

1, 2, 3, ... rate

Erlang Random Variables: A random variable X is Erlang with parameters n, λ ($n \in \mathbb{N}, \lambda > 0$) if its PDF is

convolution theorem

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} \dfrac{(\lambda x)^{n-1}}{(n-1)!} & \text{, if } x \geq 0 \\ 0 & \text{, otherwise} \end{cases}$$

$x < 0$

- What is Erlang($n = 1, \lambda$)? $= Exp(\lambda)$

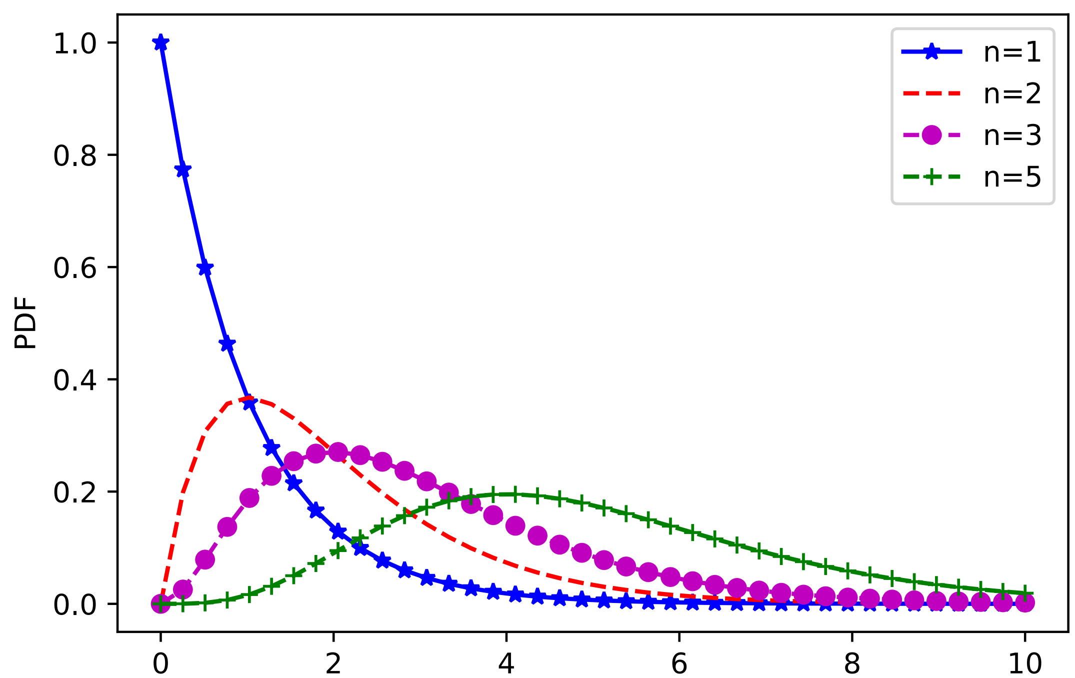
PDF of Erlang Random Variables

X is Erlang with parameters n, λ ($n \in \mathbb{N}, \lambda > 0$) if its PDF is

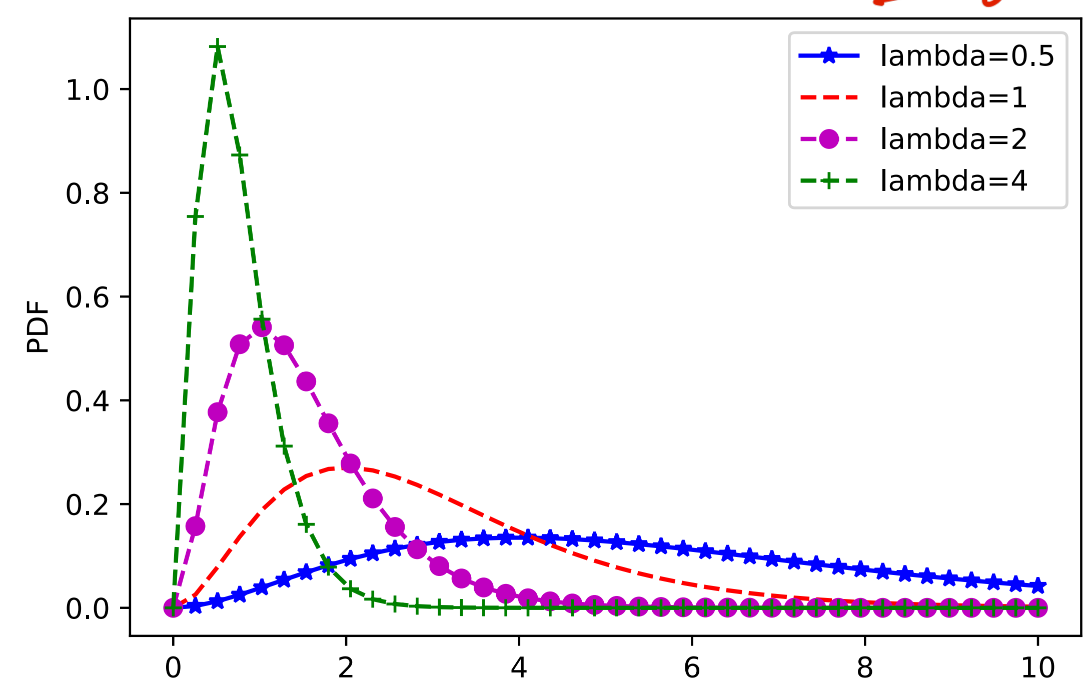
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} \dfrac{(\lambda x)^{n-1}}{(n-1)!} & \text{, if } x \geq 0 \\ 0 & \text{, otherwise} \end{cases}$$

python
scipy
erlangpdf

- How to plot the PDF?



Fix $\lambda = 1$



Fix $n = 3$

Mean and Variance

- Suppose $X \sim \text{Erlang}(n, \lambda)$
  - How to get $E[X] = n/\lambda$?
  - How to get $\text{Var}[X] = n/\lambda^2$?

Exp(λ)

$\frac{1}{\lambda}$

$\frac{1}{\lambda^2}$

4. Gamma and Beta Random Variables

5. Gamma Random Variables

Gamma Random Variables: A random variable X is Gamma with parameters r, λ $(r > 0, \lambda > 0)$ if its PDF is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} \dfrac{(\lambda x)^{n-1}}{\Gamma(r)} & \text{, if } x \geq 0 \\ 0 & \text{, otherwise} \end{cases}$$

where $\Gamma(r) = \int_0^{\infty} t^{r-1} e^{-t} dt$

Erlang (n, λ)

$\downarrow$

$n = 1, 2, 3, \cdots$

$n = 1, 3, 5, 6$

- Gamma: an extension of Erlang
- Useful in Bayesian inference (will be discussed later)

6. Beta Random Variables

Beta Random Variables: A random variable X is Beta with parameters α, β ($\alpha > 0, \beta > 0$) if its PDF is

$$f_X(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{, if } 0 < x < 1 \\ 0 & \text{, otherwise} \end{cases}$$

where $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$

- Useful in Bayesian inference (will be discussed later)

Next Lecture

1. Bivariate random variables

2. Joint distributions (CDF / PMF / PDF)

1-Minute Summary

1. Normal and Mixture of Normal

- Visualization
- Linear transformation: standard normal ⇔ normal
- Uni-modal vs multi-modal

2. Exponential and Erlang Random Variables

- PDF and CDF / memoryless property
- Convolution theorem