

Lecture 4: Random Variable, Part II

Yi, Yung (이육)

EE210: Probability and Introductory Random Processes
KAIST EE

MONTH DAY, 2021

- Continuous Random Variable
- PDF (Probability Density Function)
- CDF (Cumulative Distribution Function)
- Exponential and Normal Distribution
- Joint PDF, Conditional PDF
- Bayes' rule for continuous and even mixed cases

- Famous discrete random variables used in the community
 - Bernoulli, Uniform, Binomial, Geometric, Poisson, etc.
- Summarizing a random variable: Expectation and Variance
- Functions of a single random variable, Functions of multiple random variables
- Conditioning for random variables, Independence for random variables
- Continuous random variables
 - Normal, Uniform, Exponential, etc.
- Bayes' rule for random variables

Continuous RV and Probability Density Function

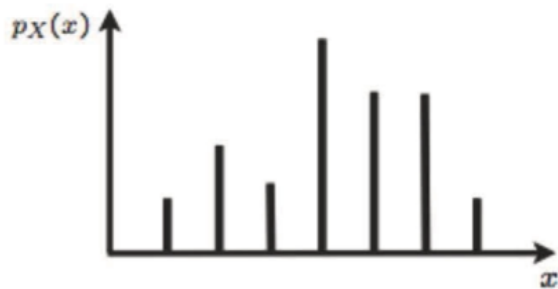
- Many cases when random variable have “continuous values”, e.g., velocity of a car

Continuous Random Variable

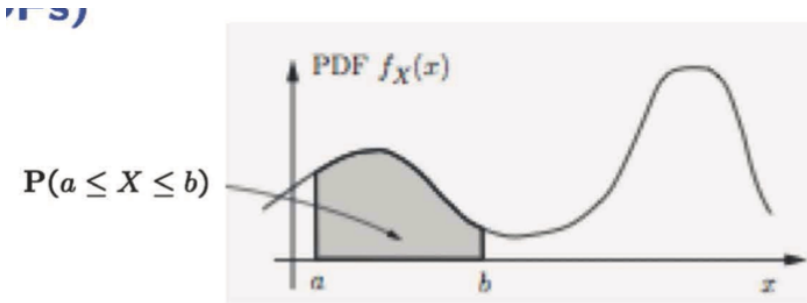
A rv X is **continuous** if \exists a function f_X , called **probability density function (PDF)**, s.t.

$$\mathbb{P}(X \in B) = \int_B f_X(x) dx$$

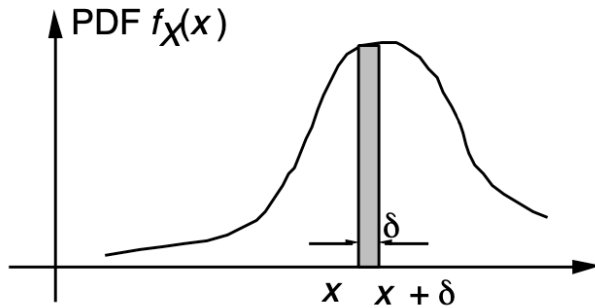
- All of the concepts and methods (expectation, PMFs, and conditioning) for discrete rvs have continuous counterparts



- $\mathbb{P}(a \leq X \leq b) = \sum_{x: a \leq x \leq b} p_X(x)$
- $p_X(x) \geq 0, \sum_x p_X(x) = 1$

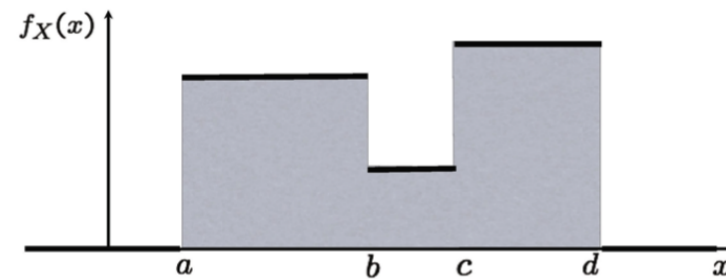
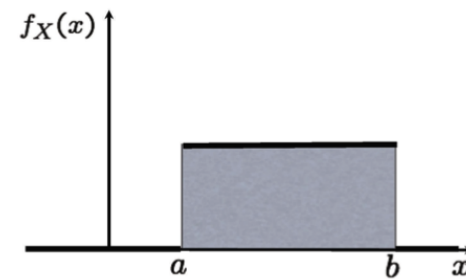


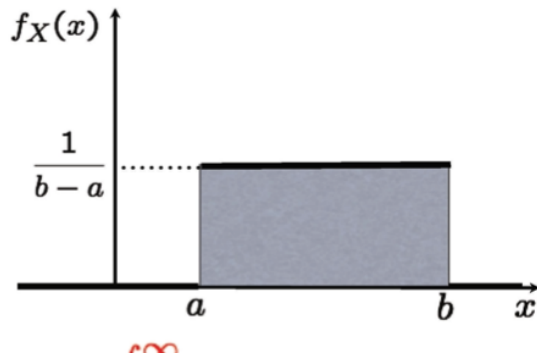
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx$
- $f_X(x) \geq 0, \int_{-\infty}^{\infty} f_X(x) dx = 1$



- $\mathbb{P}(a \leq X \leq a + \delta) \approx f_X(a) \cdot \delta$
- $\mathbb{P}(X = a) = 0$

Examples





- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{b+a}{2}$
- $\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}$
- $\text{var}[X] = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4}$

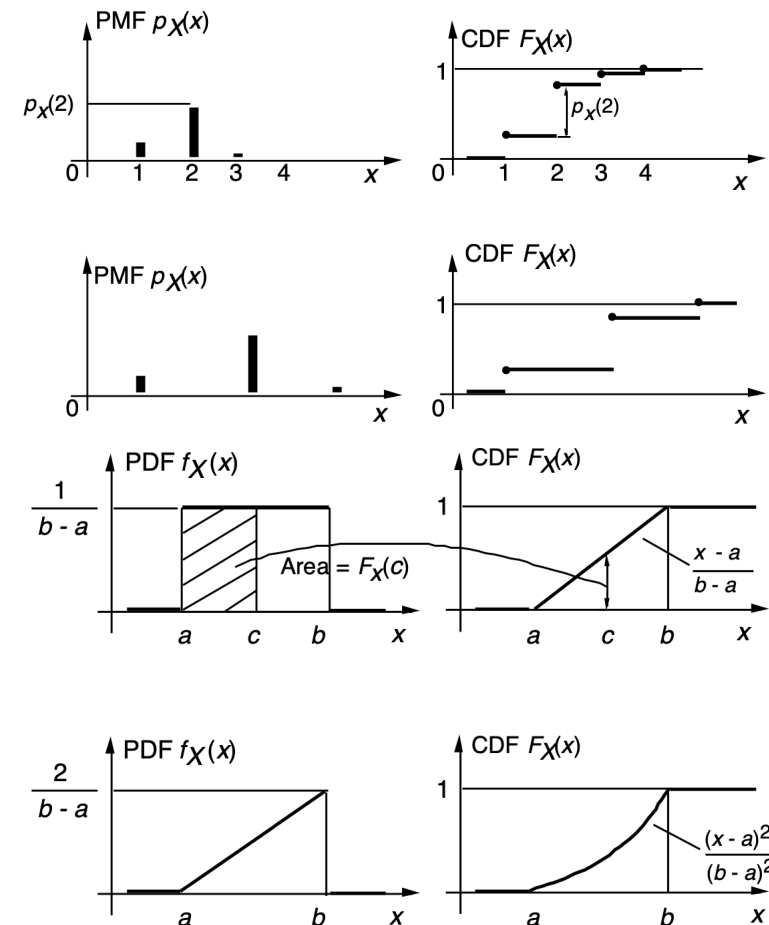
Cumulative Distribution Function (CDF)

- Discrete: PMF, Continuous: PDF
- Can we describe all rvs with a single mathematical concept?

$$F_X(x) = \mathbb{P}(X \leq x) =$$

$$\begin{cases} \sum_{k \leq x} p_X(k), & \text{discrete} \\ \int_{-\infty}^x f_X(t) dt, & \text{continuous} \end{cases}$$

- always well defined, because we can always compute the probability for the event $\{X \leq x\}$
- CCDF (Complementary CDF): $\mathbb{P}(X > x)$



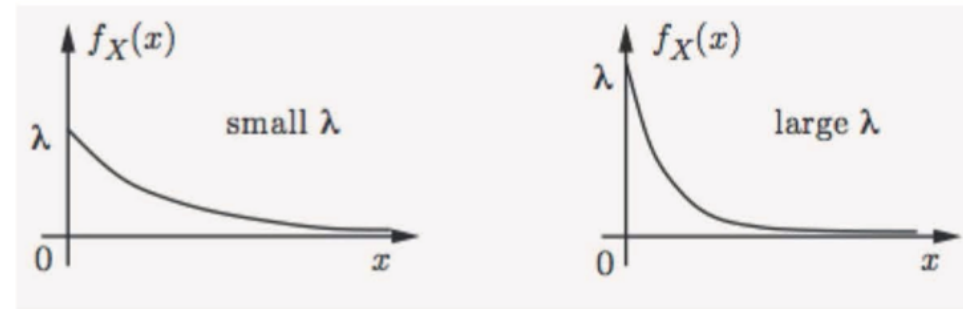
- Non-decreasing
- $F_X(x)$ tends to 1, as $x \rightarrow \infty$
- $F_X(x)$ tends to 0, as $x \rightarrow -\infty$

Now, let's look at famous continuous random variables popularly used in our life.

- A rv X is called **exponential with λ** , if

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad \text{or} \quad F_X(x) = 1 - e^{-\lambda x}$$

- Models a waiting time
- CCDF $\mathbb{P}(X \geq x) = e^{-\lambda x}$ (waiting time decays exponentially)
- $\mathbb{E}[X] = 1/\lambda$, $\mathbb{E}[X^2] = 2/\lambda^2$, $\text{var}[X] = 1/\lambda^2$
- (Q) What is the discrete rv which models a waiting time?



- A discrete twin for modeling waiting times is **geometric** rvs.
- Models a system evolution over time: Continuous time vs. Discrete time. In many cases, continuous case is the some type of **limit** of its corresponding discrete case.
- Can you make mathematical description, where geometric and exponential rvs meet each other in the limit?
- **Key idea.**
 - Continuous system: Discrete system with **infinitely many slots whose duration is infinitely small.**
- limiting system: $X_{exp}(\lambda)$ with CDF $F_{exp}(\cdot)$
- n -th system: $X_{geo}^n(p_n)$ with CDF $F_{geo}^n(\cdot)$

Modeling Waiting Time? A Discrete Twin (2)

For a given $x > 0$,

- Define $\delta = \frac{x}{n}$ (a slot length in the n -th system)

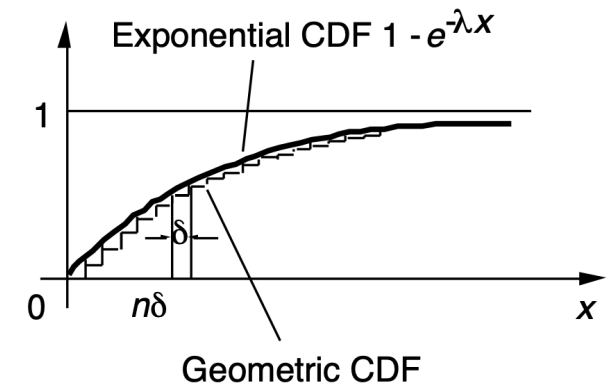
- Remember

$$F_{\text{exp}}(x) = 1 - e^{-\lambda x}$$

$$F_{\text{geo}}^n(n) = 1 - (1 - p_n)^n$$

- Choose $p_n = 1 - e^{-\lambda\delta} = 1 - e^{-\lambda\frac{x}{n}}$.
- As $n \rightarrow \infty$, the slot length $\delta \rightarrow 0$ thus $p_n \rightarrow 0$
- The CDF values of exponential and n -th geometric rvs become equal whenever $x = \delta, 2\delta, 3\delta, \dots$, i.e.,

$$F_{\text{exp}}(n\delta) = F_{\text{geo}}^n(n), \quad n = 1, 2, \dots$$



- As n grows, the number of slots grows, but the success probability over one slot decreases, so that everything is balanced up.
- As n grows, $F_{\text{geo}}^n(n)$ approaches $F_{\text{exp}}(n\delta)$.

Why important?

- Central limit theorem (중심극한정리)
 - One of the most remarkable findings in the probability theory
- Convenient analytical properties
- Modeling aggregate noise with many small, independent noise terms

- Standard Normal $N(0, 1)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

- $\mathbb{E}[X] = 0$
- $\text{var}[X] = 1$

- General Normal $N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

- $\mathbb{E}[X] = \mu$
- $\text{var}[X] = \sigma^2$

Need to check:

- a legitimate PDF or not
- expectation/variance

- Linear transformation preserves normality

Linear transformation of Normal

If $X \sim \text{Norm}(\mu, \sigma^2)$, then for $a \neq 0$ and b $Y = aX + b \sim \text{Norm}(a\mu + b, a^2\sigma^2)$.

- Thus, every normal rv can be **standardized** :

If $X \sim \text{Norm}(\mu, \sigma^2)$, then $Y = \frac{X - \mu}{\sigma} \sim \text{Norm}(0, 1)$

- Thus, we can make the **table** which records the following CDF values:

$$\Phi(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(Y < y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-t^2/2} dt$$

Example

- Annual snowfall X is modeled as $Norm(60, 20^2)$. What is the probability that this year's snowfall is at least 80 inches?
- $Y = \frac{X-60}{20}$.

$$\begin{aligned}\mathbb{P}(X \geq 80) &= \mathbb{P}\left(Y \geq \frac{80 - 60}{20}\right) \\ &= \mathbb{P}(Y \geq 1) = 1 - \Phi(1) \\ &= 1 - 0.8413 = 0.1587\end{aligned}$$

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986

- Famous discrete random variables used in the community
 - Bernoulli, Uniform, Binomial, Geometric, Poisson, etc.
- Summarizing a random variable: Expectation and Variance
- Functions of a single random variable, Functions of multiple random variables
- Conditioning for random variables, Independence for random variables
- Continuous random variables
 - Normal, Uniform, Exponential, etc.
- Bayes' rule for random variables

** Continuous counterparts are intuitively understandable. So, we will be quick at reviewing them.

Jointly Continuous

Two continuous rvs are **jointly continuous** if a non-negative function $f_{X,Y}(x, y)$ (called joint PDF) satisfies: for **every** subset B of the two dimensional plane,

$$\mathbb{P}((X, Y) \in B) = \iint_{(x,y) \in B} f_{X,Y}(x, y) dx dy$$

1. The joint PDF is used to calculate probabilities

$$\mathbb{P}((X, Y) \in B) = \iint_{(x,y) \in B} f_{X,Y}(x, y) dx dy$$

Our particular interest: $B = \{(x, y) \mid a \leq x \leq b, c \leq y \leq d\}$

2. The marginal PDFs of X and Y are from the joint PDF as:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

3. The joint CDF is defined by $F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y)$, and determines the joint PDF as:

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{x,y}}{\partial x \partial y}(x,y)$$

4. A function $g(X, Y)$ of X and Y defines a new random variable, and

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$$

* Conditional PDF, given an event

- $f_X(x) \cdot \delta \approx \mathbb{P}(x \leq X \leq x + \delta)$
 $f_{X|A}(x) \cdot \delta \approx \mathbb{P}(x \leq X \leq x + \delta | A)$

- $\mathbb{P}(X \in B) = \int_B f_X(x) dx$
 $\mathbb{P}(X \in B | A) = \int_B f_{X|A}(x) dx$

Note: A is an event, but B is a subset that includes the possible values which can be taken by the rv X .

- $\int f_{X|A}(x) = 1$

* Conditional PDF, given $X \in B$

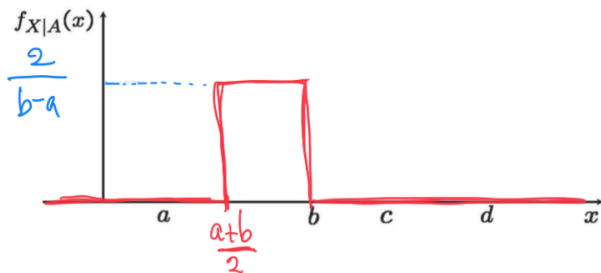
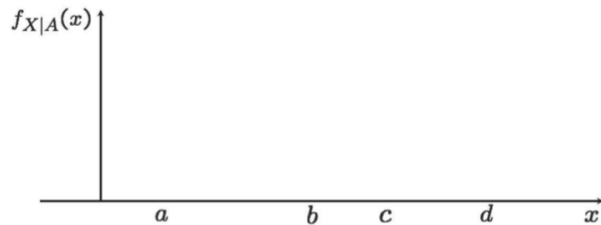
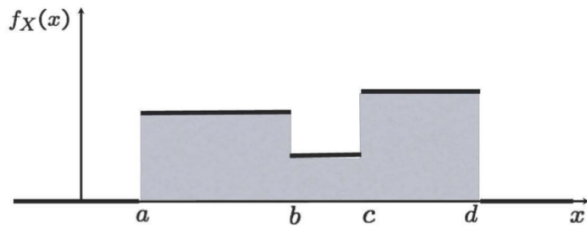
$$\mathbb{P}(x \leq X \leq x + \delta | X \in B) \approx f_{X|X \in B}(x) \cdot \delta$$

$$f_{X|X \in B}(x) = \begin{cases} 0, & \text{if } x \notin B \\ \frac{f_X(x)}{\mathbb{P}(B)}, & \text{if } x \in B \end{cases}$$

(Q) In the discrete, we consider the event $\{X = x\}$, not $\{X \in B\}$. Why?

Continuous: Conditional Expectation

$$A = \left\{ \frac{a+b}{2} \leq X \leq b \right\}$$



- $\mathbb{E}[X] = \int x f_X(x) dx$
 $\mathbb{E}[X|A] = \int x f_{X|A}(x) dx$
- $\mathbb{E}[g(X)] = \int g(x) f_X(x) dx$
 $\mathbb{E}[g(X)|A] = \int g(x) f_{X|A}(x) dx$

$$\mathbb{E}[X|A] = \int_{(a+b)/2}^b x \frac{2}{b-a} dx = \frac{a}{4} + \frac{3b}{4}$$

$$\mathbb{E}[X^2|A] = \int_{(a+b)/2}^b x^2 \frac{2}{b-a} dx =$$

- Exponential rv is a continuous counterpart of geometric rv.
- Thus, expected to be memoryless.

Definition. A random variable X is called **memoryless** if, for any $n, m \geq 0$,

$$\mathbb{P}(X > n + m | X > m) = \mathbb{P}(X > n)$$

- **Proof.** Note that $\mathbb{P}(X > x) = e^{-\lambda x}$. Then,

$$\mathbb{P}(X > n + m | X > m) = \frac{\mathbb{P}(X > n + m)}{\mathbb{P}(X > m)} = \frac{e^{-\lambda(n+m)}}{e^{-\lambda m}} = e^{-\lambda n} = \mathbb{P}(X > n)$$

Total Probability/Expectation Theorem

Partition of Ω into A_1, A_2, A_3, \dots

* Discrete case

Total Probability Theorem

$$\begin{aligned} p_X(x) &= \sum_i \mathbb{P}(A_i) \mathbb{P}(X = x | A_i) \\ &= \sum_i \mathbb{P}(A_i) p_{X|A_i}(x) \end{aligned}$$

Total Expectation Theorem

$$\mathbb{E}[X] = \sum_i \mathbb{P}(A_i) \mathbb{E}[X | A_i]$$

* Continuous case

Total Probability Theorem

$$f_X(x) = \sum_i \mathbb{P}(A_i) f_{X|A_i}(x)$$

Total Expectation Theorem

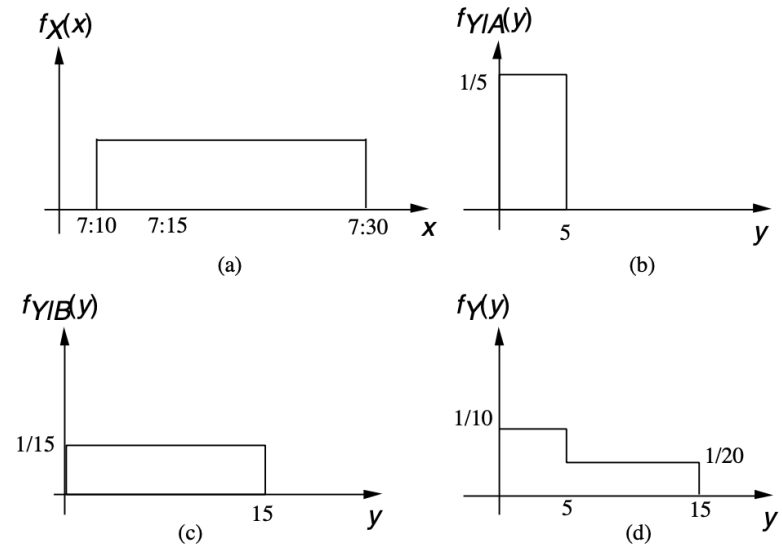
$$\mathbb{E}[X] = \sum_i \mathbb{P}(A_i) \mathbb{E}[X | A_i]$$

Example

- Your train's arrival every quarter hour (0, 15min, 30min, 45min).
- Your arrival \sim uniform(7:10, 7:30) am.
- What is the PDF of waiting time for the first train?
- X : your arrival time, Y : waiting time.
- The value of X makes a different waiting time. So, consider two events:

$$A = \{7:10 \leq X \leq 7:15\}$$

$$B = \{7:15 \leq X \leq 7:30\}$$



$$f_Y(y) = \mathbb{P}(A)f_{Y|A}(y) + \mathbb{P}(B)f_{Y|B}(y)$$

$$f_Y(y) = \frac{1}{4} \frac{1}{5} + \frac{3}{4} \frac{1}{15} = \frac{1}{10}, \quad \text{for } 0 \leq y \leq 5$$

$$f_Y(y) = \frac{1}{4} 0 + \frac{3}{4} \frac{1}{15} = \frac{1}{20}, \quad \text{for } 5 < y \leq 15$$

Continuous: Conditional PDF given a RV

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

- Similarly, for $f_Y(y) > 0$,

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

- Remember: For a fixed event A , $\mathbb{P}(\cdot|A)$ is a legitimate probability law.
- Similarly, For a fixed y , $f_{X|Y}(x|y)$ is a legitimate PDF, since

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y) dx = \frac{\int_{-\infty}^{\infty} f_{X,Y}(x,y) dx}{f_Y(y)} = 1$$

- Multiplication rule.

$$\begin{aligned} f_{X,Y}(x,y) &= f_Y(y) \cdot f_{X|Y}(x|y) \\ &= f_X(x) f_{Y|X}(y|x) \end{aligned}$$

- Total prob./exp. theorem.

$$f_X(x) = \int_{-\infty}^{\infty} f_Y(y) f_{X|Y}(x|y) dy$$

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

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} f_Y(y) \mathbb{E}[X|Y = y] dy$$

- Independence.

$$f_{X,Y}(x,y) = f_X(x) f_Y(y), \quad \text{for all } x \text{ and } y$$

Example: Stick-breaking (Ch 3. Prob 21)

- Break a stick of length l twice
 - first break at $X \sim \text{uniform}[0, l]$
 - second break at $Y \sim \text{uniform}[0, X]$
- (Q) What is $\mathbb{E}[Y]$?
- Since Y depends on X , the total expectation theorem seems useful.

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} f_X(x) \mathbb{E}[Y|X = x] dx$$

- Using the TET,

$$\begin{aligned} \mathbb{E}[Y] &= \int_0^l \frac{1}{l} \mathbb{E}[Y|X = x] dx \\ &= \int_0^l \frac{1}{l} \frac{x}{2} dx = \frac{l}{4} \end{aligned}$$

- $f_X(x)$ and $f_{Y|X}(y|x)$ seems easy to compute. Thus,

$$f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y|x) = \frac{1}{l} \cdot \frac{1}{x}$$

You can do many other things with the joint PDF.

- Famous discrete random variables used in the community
 - Bernoulli, Uniform, Binomial, Geometric, Poisson, etc.
- Summarizing a random variable: Expectation and Variance
- Functions of a single random variable, Functions of multiple random variables
- Conditioning for random variables, Independence for random variables
- Continuous random variables
 - Normal, Uniform, Exponential, etc.
- Bayes' rule for random variables

- X : state/cause/original value \rightarrow Y : result/resulting action/noisy measurement
- Model: $\mathbb{P}(X)$ (prior) and $\mathbb{P}(Y|X)$ (cause \rightarrow result)
- Inference: $\mathbb{P}(X|Y)$?

$$\begin{aligned}p_{X,Y}(x,y) &= p_X(x)p_{Y|X}(y|x) \\&= p_Y(y)p_{X|Y}(x|y) \\ \textcolor{red}{p_{X|Y}(x|y)} &= \frac{p_X(x)p_{Y|X}(y|x)}{p_Y(y)} \\ p_Y(y) &= \sum_{x'} p_X(x')p_{Y|X}(y|x')\end{aligned}$$

$$\begin{aligned}f_{X,Y}(x,y) &= f_X(x)f_{Y|X}(y|x) \\&= f_Y(y)f_{X|Y}(x|y) \\ \textcolor{red}{f_{X|Y}(x|y)} &= \frac{f_X(x)f_{Y|X}(y|x)}{f_Y(y)} \\ f_Y(y) &= \int f_X(x')f_{Y|X}(y|x')dx'\end{aligned}$$

K : discrete, Y : continuous

- Inference of K given Y

$$p_{K|Y}(k|y) = \frac{p_K(k)f_{Y|K}(y|k)}{f_Y(y)}$$

$$f_Y(y) = \sum_{k'} p_K(k')f_{Y|K}(y|k')$$

- Inference of Y given K

$$f_{Y|K}(y|k) = \frac{f_Y(y)p_{K|Y}(k|y)}{p_K(k)}$$

$$p_K(k) = \int f_Y(y')p_{K|Y}(k|y')dy'$$

Inference of discrete K given continuous Y :

$$p_{K|Y}(k|y) = \frac{p_K(k)f_{Y|K}(y|k)}{f_Y(y)}, \quad f_Y(y) = \sum_{k'} p_K(k')f_{Y|K}(y|k')$$

- K : -1, +1, original signal, equally likely. $p_K(1) = 1/2, p_K(-1) = 1/2$.
- Y : measured signal with Gaussian noise, $Y = K + W, W \sim N(0, 1)$
- Your received signal = 0.7. What's your guess about the original signal? **+1**
- Your received signal = -0.2. What's your guess about the original signal? **-1**

Example: Signal Detection (2)

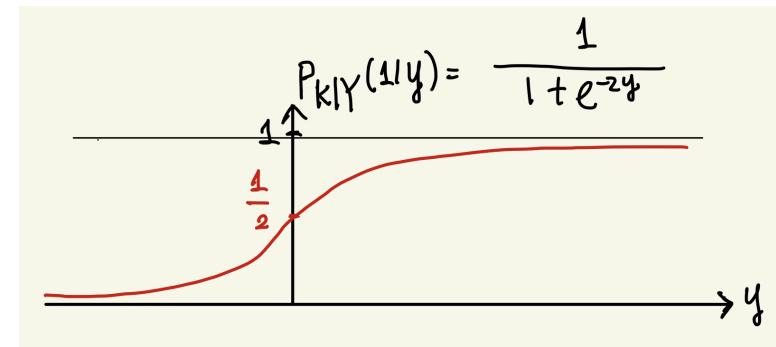
- $Y|K = 1 \sim N(1, 1)$ and $Y|K = -1 \sim N(-1, 1)$.

$$f_{Y|K}(y|k) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-k)^2}, \quad k = 1, -1$$

$$f_Y(y) = \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y+1)^2} + \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-1)^2}$$

- Probability that $K = 1$, given $Y = y$? After some algebra,

$$p_{K|Y}(1|y) = \frac{1}{1 + e^{-2y}}$$



Questions?

- 1) What is PDF and CDF?
- 2) Why do we need CDF?
- 3) What are joint/marginal/conditional PDFs?
- 4) Explain memorylessness of exponential random variables.
- 5) Explain the version of Bayes' rule for continuous and mixed random variables.