

Lecture 4: Random Variable, Part II

Yi, Yung (이웅)

EE210: Probability and Introductory Random Processes
KAIST EE

August 31, 2021

August 31, 2021 1 / 45

Roadmap

- (1) Continuous Random Variable and PDF (Probability Density Function)
- (2) CDF (Cumulative Distribution Function)
- (3) Exponential RVs
- (4) Gaussian (Normal) RVs
- (5) Continuous RVs: Joint, Conditioning, and Independence
- (6) Bayes' rule for RVs

August 31, 2021 2 / 45

- (1) Continuous Random Variable and PDF (Probability Density Function)
- (2) CDF (Cumulative Distribution Function)
- (3) Exponential RVs
- (4) Gaussian (Normal) RVs
- (5) Continuous RVs: Joint, Conditioning, and Independence
- (6) Bayes' rule for RVs

L4(1)

August 31, 2021 3 / 45

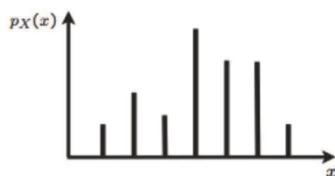
Continuous RV and Probability Density Function

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

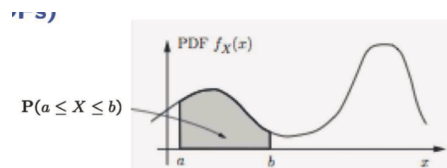
A rv X is **continuous** if \exists a function f_X , called **probability density function (PDF)**, s.t.

$$\mathbb{P}(X \in B) = \int_B f_X(x) dx, \quad \text{every subset } B \in \mathbb{R}$$

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



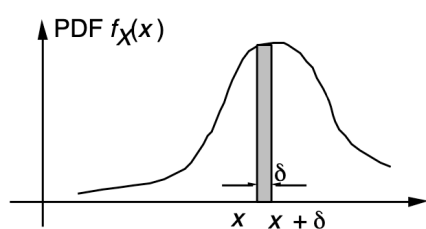
- $\mathbb{P}(a \leq X \leq b) = \sum_{x: a \leq x \leq b} p_X(x)$
- $p_X(x) \geq 0, \sum_x p_X(x) = 1$



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

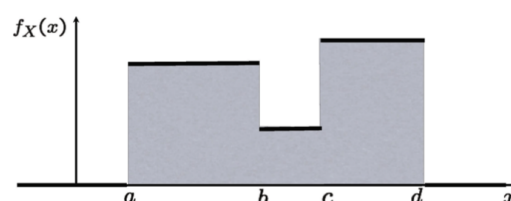
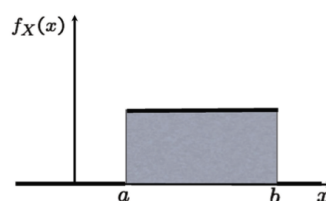
L4(1)

August 31, 2021 4 / 45



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

Examples

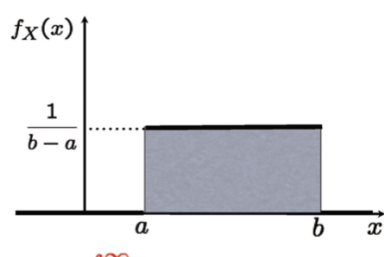


L4(1)

August 31, 2021

5 / 45

Expectation and Variance



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

L4(1)

August 31, 2021

6 / 45

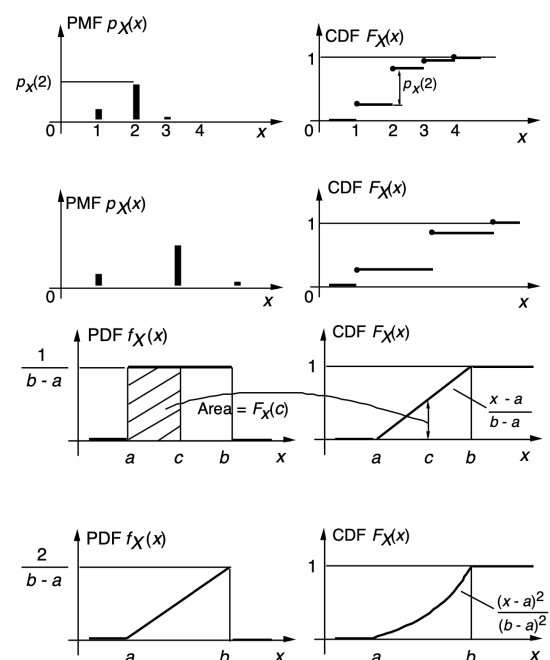
- (1) Continuous Random Variable and PDF (Probability Density Function)
- (2) CDF (Cumulative Distribution Function)
- (3) Exponential RVs
- (4) Gaussian (Normal) RVs
- (5) Continuous RVs: Joint, Conditioning, and Independence
- (6) Bayes' rule for RVs

L4(2)

August 31, 2021 7 / 45

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): $\mathbb{P}(X > x)$



L4(2)

August 31, 2021 8 / 45

- Non-decreasing
- $F_X(x)$ tends to 1, as $x \rightarrow \infty$ and $F_X(x)$ tends to 0, as $x \rightarrow -\infty$
- If X is discrete,
 - $F_X(x)$ is a piecewise constant function of x .
 - $p_X(k) = F_X(k) - F_X(k-1)$
- If X is continuous
 - $F_X(x)$ is a continuous function of x .
 - $F_X(x) = \int_{-\infty}^x f_X(t)dt$ and $f_X(x) = \frac{dF_X}{dx}(x)$

L4(2)

August 31, 2021 9 / 45

Example: Maximum of Random Variables

- Take a test three times, and your final score will be the maximum of test scores
- $X = \max\{X_1, X_2, X_3\}$, and $X_i \in \{1, 2, \dots, 10\}$ uniformly at random
- **Question.** $p_X(x)$?
- Approach 1: $\mathbb{P}(\max\{X_1, X_2, X_3\} = x)$?
- Approach 2

$$F_X(x) = \mathbb{P}(\max\{X_1, X_2, X_3\} \leq x) = \mathbb{P}(X_1 \leq x, X_2 \leq x, X_3 \leq x) \\ = \mathbb{P}(X_1 \leq x) \cdot \mathbb{P}(X_2 \leq x) \cdot \mathbb{P}(X_3 \leq x) = \left(\frac{x}{10}\right)^3$$

Thus,

$$p_X(x) = \left(\frac{x}{10}\right)^3 - \left(\frac{x-1}{10}\right)^3, \quad x = 1, 2, \dots, 10$$

L4(2)

August 31, 2021 10 / 45

- (1) Continuous Random Variable and PDF (Probability Density Function)
- (2) CDF (Cumulative Distribution Function)
- (3) Exponential RVs
- (4) Gaussian (Normal) RVs
- (5) Continuous RVs: Joint, Conditioning, and Independence
- (6) Bayes' rule for RVs

L4(3)

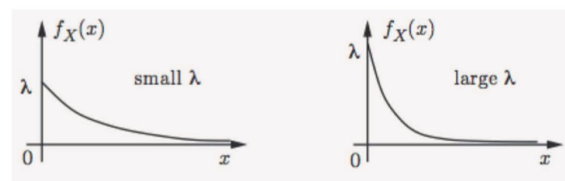
August 31, 2021

11 / 45

Exponential RV with parameter $\lambda > 0$

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

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



- CDF $F_X(x) = \int_0^x \lambda e^{-\lambda s} ds = 1 - e^{-\lambda x}$
- CCDF $\mathbb{P}(X > x) = e^{-\lambda x}$
- (Check) $\mathbb{E}[X] = 1/\lambda$, $\mathbb{E}[X^2] = 2/\lambda^2$, $\text{var}[X] = 1/\lambda^2$

L4(3)

August 31, 2021

12 / 45

- $\mathbb{E}(X) = 1/\lambda$. Use **integration by parts**: $\int u dv = uv - \int v du$

$$\int_0^{\infty} x \lambda e^{-\lambda x} dx = (-x e^{-\lambda x}) \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = 0 - \frac{e^{-\lambda x}}{\lambda} \Big|_0^{\infty} = \frac{1}{\lambda}$$

- $\mathbb{E}(X^2)$

$$\int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = (-x^2 e^{-\lambda x}) \Big|_0^{\infty} + \int_0^{\infty} 2x e^{-\lambda x} dx = 0 + \frac{2}{\lambda} \mathbb{E}(X) = \frac{2}{\lambda^2}$$

- $\text{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{1}{\lambda^2}$

L4(3)

August 31, 2021

13 / 45

- $\mathbb{P}(X > x) = e^{-\lambda x}$
- Appropriate for modeling a waiting time until an incident of interest takes place
 - $\mathbb{P}(X > x)$: exponentially decays
 - message arriving at a computer, some equipment breaking down, a light bulb burning out, etc
- **(Q)** What is the discrete rv which models a waiting time? **Geometric**
- What is the relationship between exponential rv and geometric rv? We will see this relationship soon, but let's look at an example first.

L4(3)

August 31, 2021

14 / 45



- A very small meteorite first lands anywhere in Korea
- Time of landing is modeled as an exponential rv with mean 10 days
- The current time is midnight. What is the probability that a meteorite first lands some time between 6 a.m. and 6 p.m. of the first day?

VIDEO PAUSE

- (Solution)

- $\mathbb{E}(X) = 1/\lambda = 10$. Thus, $\lambda = \frac{1}{10}$.
- 6 a.m. from midnight = 1/4 day, 6 p.m. from midnight = 3/4 day

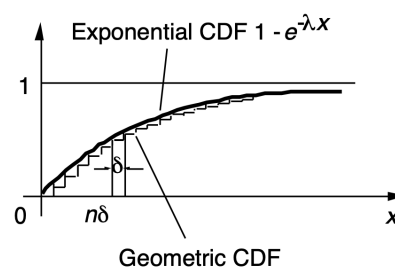
$$\mathbb{P}(1/4 \leq X \leq 3/4) = \mathbb{P}(X \geq 1/4) - \mathbb{P}(X \geq 3/4) = e^{-1/40} - e^{-3/40} = 0.0476$$

- Models a system evolution over time: Continuous time vs. Discrete time.
 - **Example.** Customer arrivals at my shop
 - **Modeling 1:** Every 30 minute I record the number of customers for each 30-min window
 - **Modeling 2:** I record the exact time of each customer's arrival
 - In modeling 1, every 10 minute? every 1 minute? every 1 sec? every 0.0000001 sec?
- In many cases, continuous case is some type of **limit** of its corresponding discrete case.
- Can we mathematically describe how geometric and exponential rvs meet each other in the limit?

- 'slot' is one unit time, e.g., 1 hour, 30 mins, 1 min, 10 sec, etc.
- Continuous system = Discrete system with
 - infinitely many slots whose duration is infinitely small.
 - success probability p over one slot decreases to 0 in the limit
- Given $X^{exp} \sim \exp(\lambda)$, let us construct a geometric RV X_δ^{geo}
 - Set the length of a slot to be δ , which is a parameter.
 - Set the success probability p_δ over a slot to be $p_\delta = 1 - e^{-\lambda\delta}$ (this looks magical, whose secret will be uncovered soon)
 - $\mathbb{P}(X_\delta^{geo} \leq n) = 1 - (1 - p_\delta)^n = 1 - e^{-\lambda\delta n}$

L4(3)

August 31, 2021 17 / 45



- Note that $\mathbb{P}(X^{exp} \leq x) = 1 - e^{-\lambda x}$. Then, when $x = n\delta$, $n = 1, 2, \dots$

$$\mathbb{P}(X^{exp} \leq x) = 1 - e^{-\lambda\delta n} = \mathbb{P}(X_\delta^{geo} \leq n)$$
- If we choose sufficiently small δ , the slot length \downarrow and $p_\delta \downarrow$

$$\mathbb{P}(X_\delta^{geo} \leq n) \xrightarrow{\delta \rightarrow 0} \mathbb{P}(X^{exp} \leq x), \quad x = n\delta$$

L4(3)

August 31, 2021 18 / 45

- (1) Continuous Random Variable and PDF (Probability Density Function)
- (2) CDF (Cumulative Distribution Function)
- (3) Exponential RVs
- (4) **Gaussian (Normal) RVs**
- (5) Continuous RVs: Joint, Conditioning, and Independence
- (6) Bayes' rule for RVs

L4(4)

August 31, 2021

19 / 45

Normal: PDF, Expectation, Variance

- **Standard** Normal $\mathcal{N}(0, 1)$

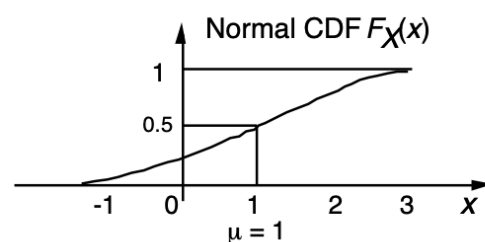
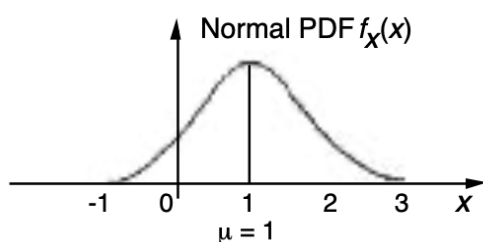
$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

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

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

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

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



L4(4)

August 31, 2021

20 / 45

- PDF's normalization property: $\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = 1$
 - A little bit boring :-). See Problem 14 at pp 189.
- Expectation
 - $f_X(x)$ is symmetric in terms of $x = \mu$. Thus, we should have $\mathbb{E}(X) = \mu$.
- Variance

$$\begin{aligned} \text{var}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx \stackrel{y=\frac{x-\mu}{\sigma}}{=} \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \left(-ye^{-y^2/2} \right) \Big|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = \sigma^2 \end{aligned}$$

$$\int u dv = uv - \int v du: u = y \text{ and } dv = ye^{-y^2/2} \rightarrow du = dy \text{ and } v = -e^{-y^2/2}$$

L4(4)

August 31, 2021 21 / 45

- Linear transformation preserves normality (we will verify this in Lecture 5)

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

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

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

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

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

L4(4)

August 31, 2021 22 / 45

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

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

	.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

L4(4)

August 31, 2021 23 / 45

Normal RVs: Why Important?

- Central limit theorem
 - One of the most remarkable findings in the probability theory
 - Sum of **any** random variables \approx Normal random variable
- Modeling aggregate noise with many small, independent noise terms
- Convenient analytical properties, allowing closed forms in many cases
- Highly popular in communication and machine learning areas

⁰Central limit theorem: 중심극한정리

L4(4)

August 31, 2021 24 / 45

- (1) Continuous Random Variable and PDF (Probability Density Function)
- (2) CDF (Cumulative Distribution Function)
- (3) Exponential RVs
- (4) Gaussian (Normal) RVs
- (5) Continuous RVs: Joint, Conditioning, and Independence
- (6) Bayes' rule for RVs

L4(5)

August 31, 2021

25 / 45

Continuous: Joint PDF and CDF (1)

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

L4(5)

August 31, 2021

26 / 45

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

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

L4(5)

August 31, 2021 27 / 45

* Conditional PDF, given an event A

- $f_X(x) \cdot \delta \approx \mathbb{P}(x \leq X \leq x + \delta)$
 $f_{X|A}(x) \cdot \delta \approx \mathbb{P}(x \leq X \leq x + \delta | A)$
- $\mathbb{P}(X \in B) = \int_B f_X(x) dx$
 $\mathbb{P}(X \in B | A) = \int_B f_{X|A}(x) dx$
- $\int f_{X|A}(x) dx = 1$

* Conditional PDF, given $\{X \in C\}$

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

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

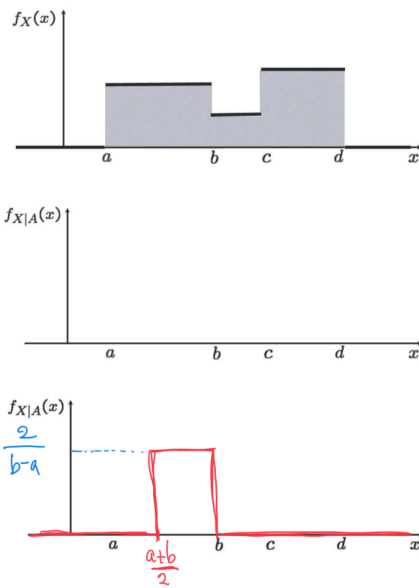
(Q) In the discrete, we consider the event $\{X = x\}$, not $\{X \in B\}$. Why?

Notation: A is an event, but B and C is a subset that includes the possible values which can be taken by the rv X . Sorry for the confusion, if any.

L4(5)

August 31, 2021 28 / 45

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



L4(5)

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

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

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

August 31, 2021 29 / 45

Exponential RV: Memoryless

- **Remember:** Exponential rv is a continuous counterpart of geometric rv.
- Thus, expected to be memoryless. Remember the definition?

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 the exponential rv's CCDF $\mathbb{P}(X > x) = e^{-\lambda x}$. Then,

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

L4(5)

August 31, 2021 30 / 45

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

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

L4(5)

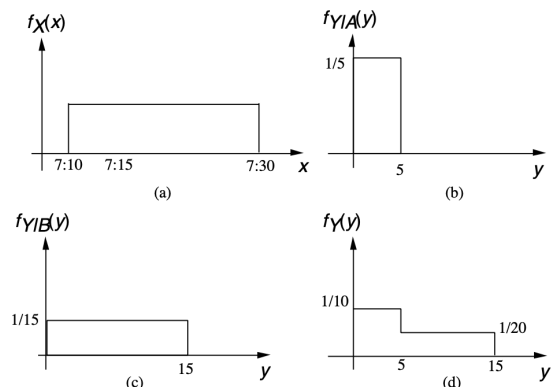
August 31, 2021 31 / 45

Example: Train Arrival

- The train's arrival every quarter hour (0, 15min, 30min, 45min).
- Your arrival $\sim \mathcal{U}(7:10, 7:30)$ 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:

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

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



VIDEO PAUSE

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

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

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

L4(5)

August 31, 2021 32 / 45

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

- 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.**

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

- **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**

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

L4(5)

August 31, 2021

33 / 45

Example: Stick-breaking (1)

(Prob 21 at pp. 191)

- Break a stick of length l twice
 - first break at $Y \sim \mathcal{U}[0, l]$
 - second break at $X \sim \mathcal{U}[0, Y]$

(a) joint PDF $f_{X,Y}(x,y)$?

$$f_Y(y) = \frac{1}{l}, \quad 0 \leq y \leq l$$

$$f_{X|Y}(x|y) = \frac{1}{y}, \quad 0 \leq x \leq y$$

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

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{l} \cdot \frac{1}{y}, & 0 \leq x \leq y \leq l, \\ 0, & \text{otherwise} \end{cases}$$

(b) marginal PDF $f_X(x)$?

$$\begin{aligned} f_X(x) &= \int f_{X,Y}(x,y) dy = \int_x^l \frac{1}{ly} dy \\ &= \frac{1}{l} \ln(l/x), \quad 0 \leq x \leq l \end{aligned}$$

⁰ $\mathcal{U}[a, b]$: continuous uniform random variable over the interval $[a, b]$

L4(5)

August 31, 2021

34 / 45

(c) Evaluate $\mathbb{E}(X)$, using $f_X(x)$

$$\begin{aligned}\mathbb{E}(X) &= \int_0^l x f_X(x) dx = \int_0^l \frac{x}{l} \ln(l/x) dx \\ &= \frac{l}{4}\end{aligned}$$

(d) Evaluate $\mathbb{E}(X)$, using $X = Y \cdot (X/Y)$

If $Y \perp\!\!\!\perp X/Y$, it becomes easy, but true?
Yes, because whatever Y is, the fraction X/Y does not depend on it.

$$\mathbb{E}(X) = \mathbb{E}(Y)\mathbb{E}(X/Y) = \frac{l}{2} \cdot \frac{1}{2} = \frac{l}{4}$$

(e) Evaluate $\mathbb{E}(X)$, using TET

$$\begin{aligned}0\mathbb{E}[X] &= \int_{-\infty}^{\infty} f_Y(y)\mathbb{E}[X|Y=y]dy \\ &= \int_0^l \frac{1}{l}\mathbb{E}[X|Y=y]dy = \int_0^l \frac{1}{l} \frac{y}{2} dy = \frac{l}{4}\end{aligned}$$

- **Message.** There are many ways to reach our goal. Of crucial importance is how to find the best way!

- (1) Continuous Random Variable and PDF (Probability Density Function)
- (2) CDF (Cumulative Distribution Function)
- (3) Exponential RVs
- (4) Gaussian (Normal) RVs
- (5) Continuous RVs: Joint, Conditioning, and Independence
- (6) **Bayes' rule for RVs**

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

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

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

L4(6)

August 31, 2021 37 / 45

Example

- A light bulb $Y \sim \exp(\lambda)$. However, there are some quality control problems. So, the parameter λ of Y is actually a random variable, denoted by Λ , which is $\Lambda \sim \mathcal{U}[1, 3/2]$. We test a light bulb and record its lifetime.
- **Question.** What can we say about the underlying parameter λ ? In other words, what is $f_{\Lambda|Y}(\lambda|y)$?
- $f_{\Lambda}(\lambda) = 2$ for $1 \leq \lambda \leq 3/2$ and $f_{Y|\Lambda}(y|\lambda) = \text{pdf of } \exp(\lambda)$. Then, the inference about the parameter given the lifetime of a light bulb is:

$$f_{\Lambda|Y}(\lambda|y) = \frac{f_{\Lambda}(\lambda)f_{Y|\Lambda}(y|\lambda)}{\int_{-\infty}^{\infty} f_{\Lambda}(t)f_{Y|\Lambda}(y|t)dt}$$

L4(6)

August 31, 2021 38 / 45

- X : **parameter** $\rightarrow Y$: result of **my model**
- Given: $\mathbb{P}(X)$ and $\mathbb{P}(Y|X)$ (parameter \rightarrow model)
- Inference: $\mathbb{P}(X|Y)$? Probabilistic feature of the parameter given the result of the model?

Example.

1. Light bulb's lifetime $Y \sim \exp(\lambda)$. Given the lifetime y , the modified belief about λ ?
2. Romeo and Juliet start dating, but Romeo will be late by a random variable $Y \sim \mathcal{U}[0, \theta]$. Given the time of being late y , the modified belief about θ ?

L4(6)

August 31, 2021

39 / 45

Bayes Rule for Mixed Case

 K : discrete, Y : continuous

- Inference of K given Y

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

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

- $f_{Y|K}(y|k) = f_{Y|A}(y)$, where $A = \{K = k\}$

- Inference of Y given K

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

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

- Wait! $p_{K|Y}(k|y)$? Well-defined?

$$p_{K|Y}(k|y) = \frac{\mathbb{P}(K = k, Y = y)}{\mathbb{P}(Y = y)} = \frac{0}{0}$$

L4(6)

August 31, 2021

40 / 45

- For small δ (in other words, taking the limit as $\delta \rightarrow 0$).

Let $A = \{K = k\}$.

$$\begin{aligned}
 p_{K|Y}(k|y) &\approx \mathbb{P}(A|y \leq Y \leq y + \delta) \\
 &= \frac{\mathbb{P}(A)\mathbb{P}(y \leq Y \leq y + \delta|A)}{\mathbb{P}(y \leq Y \leq y + \delta)} \\
 &\approx \frac{\mathbb{P}(A)f_{Y|A}(y)\delta}{f_Y(y)\delta} \\
 &= \frac{\mathbb{P}(A)f_{Y|A}(y)}{f_Y(y)}
 \end{aligned}$$

L4(6)

August 31, 2021

41 / 45

Example: Signal Detection (1)

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, $Y = K + W$, $W \sim \mathcal{N}(0, 1)$
- 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**
- Your intuition: If positive received signal, +1. If negative received signal, -1. How can we mathematically verify this?

L4(6)

August 31, 2021

42 / 45

- $Y|\{K = 1\} \sim \mathcal{N}(1, 1)$ and $Y|\{K = -1\} \sim \mathcal{N}(-1, 1)$.
(Remind: linear transformation preserves normality.)

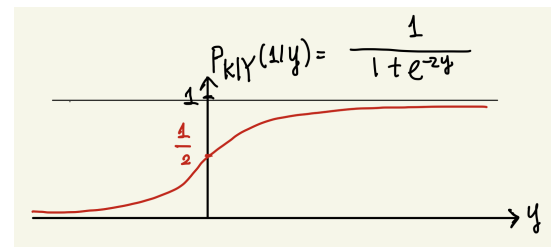
$$f_{Y|K}(y|k) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-k)^2}, \quad k = 1, -1$$

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

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

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

- If $y > 0$, the inference probability for $K = 1$ exceeds $\frac{1}{2}$. So, original signal = 1.
- Similarly, compute $p_{K|Y}(-1|y)$ and then do the inference



Questions?

- 1) What do we mean by “continuous” in continuous random variables?
- 2) Explain PDF and CDF. Why do we need CDF?
- 3) What are joint/marginal/conditional PDFs?
- 4) Explain the relationship between Geometric rvs and Exponential rvs.
- 5) Explain how normality is preserved under linear transformation for Normal (Gaussian) rvs.
- 6) Explain how we can use Bayes’ rule for parameter learning.
- 7) Explain the version of Bayes’ rule for continuous and mixed random variables.