Second half of the course:

- Instructor: Lele Wang (Please call my first name :))
- Office hour: Fridays 2:10-3:00pm online or by appointment.
- How to reach me: Email: lelewang@ece.ubc.ca
- Course material in the second half.
  - Conditional probability, Bayes rule (review)
  - Bayesian inference — MAP estimation
  - Conditional expectation, coverience
  - Bayesian inference — LMS & linear LMS estimation.
  - Joint Gaussian r.v.s.

The General Inference Problem

Ground truth (unknown to us)

data (known to us)

Estimator. (our guess about unknown)

dog

dog or cat ?

cat

$$\Theta \longrightarrow p(x|\theta) \longrightarrow X \longrightarrow f(x) \longrightarrow \hat{\Theta}$$

https://www.medicalnewstoday.com/articles/322868

https://www.theguardian.com/lifeandstyle/2020/sep/05/what-cats-mean-by-miaow-japans-pet-guru-knows-just-what-your-feline-friend-wants

Notation (used in 2nd half & different from 1st half)

- Probability mass function (pmf) of a discrete r.v. X :

lower case p   upper case P, "Probability of"

$$p_X(x) = \mathbb{P}(X=x).$$

$f(x)$ in Part 1

- Cumulative distribution function (cdf) of a discrete r.v. X

$$F_X(x) = \mathbb{P}(X \le x) = \sum_{t \le x} p_X(t)$$

$F(x)$ in Part 1.

- Probability density function of a continuous r.v. X

$$f_X(x)$$

$f(x)$ in Part 1.

- Cumulative distribution function (cdf) of a continuous r.v. X

$$F_X(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt.$$

$F(x)$ in Part 1.

- Variance of a r.v. X : $\mathrm{Var}(X)$ $V(X)$ in Part 1.

Notation

- When a random variable & its realization are the upper case & lower case of the same letter we may drop the subscript.

- eg: $p(x) \triangleq p_X(x) = \mathbb{P}(X = x)$.

$p(\theta|x) \triangleq p_{\Theta|X}(\theta|x) = \mathbb{P}(\Theta = \theta | X = x)$.

$f(x|\theta) \triangleq f_{X|\Theta}(x|\theta)$

but $p(1|x)$ can be confusing

$\rightarrow$ Write $p_{\Theta|X}(1|x)$

Conditional Probability

& Bayes Rule.

A summary of Bayes rules.

- Bayes rule for events. Let A, B be two events. Then,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(B)}$$

- Bayes rule for <u>discrete</u> random variables X, Y.
  - For each x, y, treat $\{X=x\}$ $\{Y=y\}$ as events.

short notation

  - $$\mathbb{P}(X=x|Y=y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)} = \frac{\mathbb{P}(X=x)\mathbb{P}(Y=y|X=x)}{\mathbb{P}(Y=y)} \Leftrightarrow \boxed{p(x|y) = \frac{p(x)p(y|x)}{p(y)}}$$

- Bayes rule in general
  - X continuous Y continuous: $\boxed{f(x|y) = \frac{f(x)f(y|x)}{f(y)}}$
  - X discrete Y continuous: $\boxed{p(x|y) = \frac{p(x)f(y|x)}{f(y)}}$
  - X continuous Y discrete: $\boxed{f(x|y) = \frac{f(x)p(y|x)}{p(y)}}$

Example: COVID-19 test. https://nyti.ms/31MTZgV

Suppose you think you may have contracted COVID-19.

You decide to take a diagnostic test to determine if you are infected.

The probability of contracting Covid-19 is 10%.

The false negative rate is 12.5%; The false positive rate is 2.5%.

What is the probability of being infected given a positive test result?

- $\Theta = 1$: infected, $\Theta = 0$: not infected. $P_\Theta(1) = 10\%$
- $X = 1$: test positive. $X = 0$: test negative. $P_{X|\Theta}(0|1) = 12.5\%$, $P_{X|\Theta}(1|0) = 2.5\%$

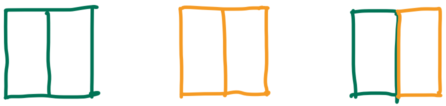
$$\mathbb{P}(\Theta = 1 \mid X = 1) \underset{\text{Bayes rule}}{\overset{\uparrow}{=}} \frac{P_\Theta(1) P_{X|\Theta}(1|1)}{P_X(1)} \underset{\text{law of total probability}}{\overset{\uparrow}{=}} \frac{P_\Theta(1) P_{X|\Theta}(1|1)}{P_\Theta(1) P_{X|\Theta}(1|1) + P_\Theta(0) P_{X|\Theta}(1|0)}$$

$$= \frac{0.1 \times 0.875}{0.1 \times 0.875 + 0.9 \times 0.025} = 0.795$$

Example: Three cards.

- There are 3 cards

① green on both sides
② yellow on both sides
③ green on one side and yellow on the other



1 2 3

- Pick a card and a side uniform at random. Let X be the color you get
- Let Y be the color on the back.
- Q: What is $\mathbb{P}(Y = \text{green} \mid X = \text{green})$? A. $> \frac{1}{2}$ ✓ B. $< \frac{1}{2}$ C. $= \frac{1}{2}$.

- card number $\Theta \sim \text{Unif}\{1, 2, 3\}$.
- $P_{X|\Theta}(\text{green} \mid 1) = 1$, $P_{X|\Theta}(\text{green} \mid 2) = 0$, $P_{X|\Theta}(\text{green} \mid 3) = \frac{1}{2}$
- $$P_{\Theta|X}(1 \mid \text{green}) = \frac{P_\Theta(1)\, P_{X|\Theta}(\text{green} \mid 1)}{\sum_\theta P_{X|\Theta}(\text{green} \mid \theta)\, P(\theta)} = \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3}} = \frac{2}{3}$$

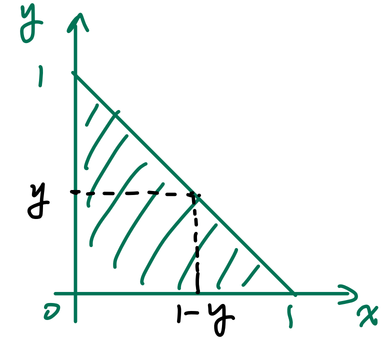
$$\Rightarrow \mathbb{P}(Y = \text{green} \mid X = \text{green}) = \mathbb{P}(\Theta = 1 \mid X = \text{green}) = \frac{2}{3}.$$

Example : Find conditional pdf for continuous r.v.s.

- Let $X, Y \sim f(x,y)$, where

$$f(x,y) = \begin{cases} 2, & x \geq 0,\ y \geq 0,\ x+y \leq 1 \\ 0, & \text{o.w.} \end{cases}$$

Find $f(x|y)$.

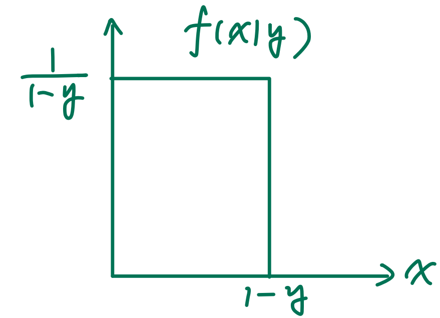


Solution : We first find the marginal pdf

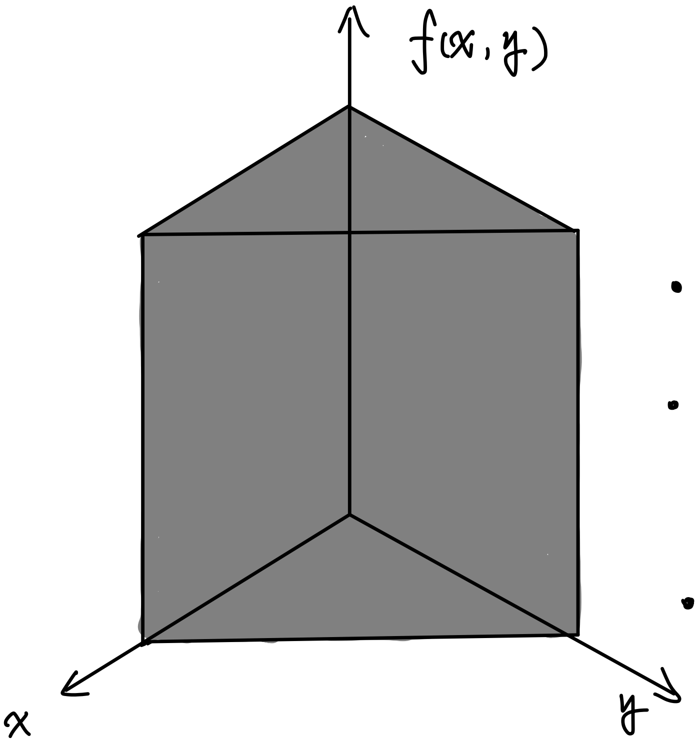
$$f(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^{1-y} 2\,dx = \begin{cases} 2(1-y), & 0 \leq y \leq 1 \\ 0, & \text{o.w.} \end{cases}$$

$$\Rightarrow \quad f(x|y) = \frac{f(x,y)}{f(y)} = \begin{cases} \frac{1}{1-y}, & 0 \leq y \leq 1,\ 0 \leq x \leq 1-y \\ 0, & \text{o.w.} \end{cases}$$

In other words, $X \mid \{Y=y\} \sim Unif[0, 1-y]$



Takeaway Message



continous r.v.s	disnete r.v.s.
• $f(x,y)$	• $p(x,y)$
• $f(y) = \int f(x,y)\,dx$	• $p(y) = \sum_x p(x,y)$
• $f(x\|y) = \dfrac{f(x,y)}{f(y)}$	• $p(x\|y) = \dfrac{p(x,y)}{p(y)}$
• For any fixed y $\int f(x\|y)\,dx = 1$	• For any fixed y $\sum_x p(x\|y) = 1$
• For any event A $\mathbb{P}(A) = \int \mathbb{P}(A\|y) f(y)\,dy$	• For any event A $\mathbb{P}(A) = \sum_x \mathbb{P}(A\|X=x)p(x)$

Example : Binary Symmetric Channel (BSC)

Consider the following binary communication channel

$$
\begin{tikzcd}
 & Z \in \{0,1\} \arrow[d] & \\
X \in \{0,1\} \arrow[r] & \oplus \arrow[r] & Y \in \{0,1\}
\end{tikzcd}
$$

The bit sent is $X \sim \text{Bern}(p)$, $0 \leq p \leq 1$.

The noise is $Z \sim \text{Bern}(\varepsilon)$, $0 \leq \varepsilon \leq 0.5$.

The bit received is $Y = (X + Z) \bmod 2 \triangleq X \oplus Z$

Assume X and Z are independent

Find

1. $p(x|y)$
2. $p(y)$
3. The probability of error $\mathbb{P}(X \neq Y)$

1. To find $p(x|y)$, we use Bayes rule

$$p(x|y) = \frac{p(y|x)\, p(x)}{\sum_{x' \in \mathcal{X}} p_{Y|X}(y|x')\, p_X(x')}$$

We know $p(x)$, but we need to find $p(y|x)$

$$\boxed{\begin{aligned} p(y|x) &= \mathbb{P}(Y=y \mid X=x) = \mathbb{P}(X \oplus Z = y \mid X=x) \\ &= \mathbb{P}(x \oplus Z = y \mid X=x) = \mathbb{P}(Z = (y-x) \bmod 2 \mid X=x) \\ &= \mathbb{P}(Z = y \oplus x \mid X=x) \underset{\substack{\uparrow \\ X \text{ and } Z \\ \text{independent}}}{=} \mathbb{P}(Z = y \oplus x) = p_Z(y \oplus x) \end{aligned}}$$

This is a derivation we will encounter many times through the course

$$\Rightarrow \quad \begin{aligned} p_{Y|X}(0|0) &= p_Z(0 \oplus 0) = p_Z(0) = 1-\varepsilon \\ p_{Y|X}(0|1) &= p_Z(0 \oplus 1) = p_Z(1) = \varepsilon \\ p_{Y|X}(1|0) &= p_Z(1 \oplus 0) = p_Z(1) = \varepsilon \\ p_{Y|X}(1|1) &= p_Z(1 \oplus 1) = p_Z(0) = 1-\varepsilon. \end{aligned}$$

Plugging into the Bayes rule equation, we obtain

$$P_{X|Y}(0|0) = \frac{P_{Y|X}(0|0)\,P_X(0)}{P_{Y|X}(0|0)\,P_X(0) + P_{Y|X}(0|1)\,P_X(1)} = \frac{(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)+\varepsilon p}$$

$$P_{X|Y}(1|0) = 1 - P_{X|Y}(0|0) = \frac{\varepsilon p}{(1-\varepsilon)(1-p)+\varepsilon p}.$$

$$P_{X|Y}(0|1) = \frac{P_{Y|X}(1|0)\,P_X(0)}{P_{Y|X}(1|0)\,P_X(0) + P_{Y|X}(1|1)\,P_X(1)} = \frac{\varepsilon(1-p)}{(1-\varepsilon)p+\varepsilon(1-p)}$$

$$P_{X|Y}(1|1) = 1 - P_{X|Y}(0|1) = \frac{(1-\varepsilon)p}{(1-\varepsilon)p+\varepsilon(1-p)}$$

2. We can find $P_Y(y)$ as

$$P_Y(y) = P_{Y|X}(y|0)\,P_X(0) + P_{Y|X}(y|1)\,P_X(1)$$

$$= \begin{cases} (1-\varepsilon)(1-p)+\varepsilon p & \text{for } y=0 \\ \varepsilon(1-p)+(1-\varepsilon)p & \text{for } y=1 \end{cases}$$

3. To find the probability of error, consider

$$\mathbb{P}(X \neq Y) = \mathbb{P}(\{X=0, Y=1\} \text{ or } \{X=1, Y=0\})$$
$$= p_{X,Y}(0,1) + p_{X,Y}(1,0)$$
$$= p_{Y|X}(1|0)\, p_X(0) + p_{Y|X}(0|1)\, p_X(1)$$
$$= \varepsilon(1-p) + \varepsilon p = \varepsilon.$$

An interesting special case is $\varepsilon = \frac{1}{2}$. Then $\mathbb{P}(X \neq Y) = \frac{1}{2}$ and

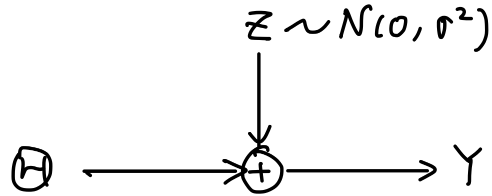
$$p_Y(0) = \tfrac{1}{2}p + \tfrac{1}{2}(1-p) = \tfrac{1}{2} = p_Y(1)$$

$\Rightarrow Y \sim \mathrm{Bern}(\frac{1}{2})$, independent of the value p!

In this case, X is independent of Y (check this)

Example: Additive Gaussian Noise Channel.

Consider the following communication channel:



The signal transmitted is a binary random variable Θ:

$$\Theta = \begin{cases} +1 & \text{, with probability } p \\ -1 & \text{, with probability } 1-p. \end{cases}$$

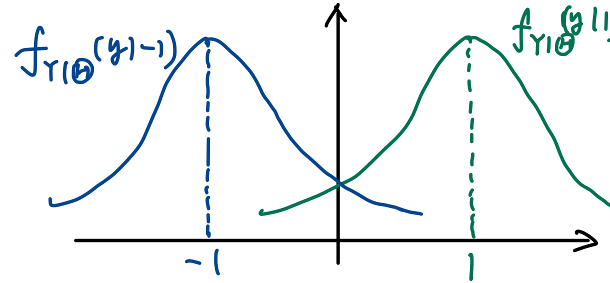
The received signal, also called the observation, is $Y = \Theta + Z$

We assume Θ and Z are independent.

Given $Y = y$ is received (observed), find $p_{\Theta|Y}(\theta|y)$.

Solution: We use the Bayes rule

$$p_{\Theta|Y}(\theta|y) = \frac{f_{Y|\Theta}(y|\theta)\, p_\Theta(\theta)}{\sum_{\theta'} f_{Y|\Theta}(y|\theta')\, p_\Theta(\theta')}$$



We know $p_\Theta(1) = p$ and $p_\Theta(0) = 1-p$.

$$\boxed{f_{Y|\Theta}(y|\theta) = f_{Z|\Theta}(y-\theta|\theta) = f_Z(y-\theta)}$$

recall the trick in the BSC example

Alternative explanation: Given $\Theta=1$, $Y = Z+1$. (Y is a shift of Z to the right by 1)
Given $\Theta=-1$, $Y = Z-1$ (Y is a shift of Z to the left by 1)

$$\Rightarrow Y|\{\Theta=+1\} \sim N(+1, \sigma^2) \quad \text{and} \quad Y|\{\Theta=-1\} \sim N(-1, \sigma^2)$$

$$\Rightarrow p_{\Theta|Y}(1|y) = \frac{\frac{p}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-1)^2}{2\sigma^2}}}{\frac{p}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-1)^2}{2\sigma^2}} + \frac{(1-p)}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+1)^2}{2\sigma^2}}} = \frac{p e^{\frac{y}{\sigma^2}}}{p e^{\frac{y}{\sigma^2}} + (1-p) e^{-\frac{y}{\sigma^2}}}$$

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