ECSE 2500
Lec 21
April 17

Topic: Conditional PDF for Continuous RVs.

☐ Definition of conditional PDF

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} , \text{ when } f_X(x) > 0$$

Big case   Small case

This is the generalization of the conditional PMF.

As before, 2 importants things to note are

1) $f_{XY}(x,y) = f_{X|Y}(x,y) \cdot f_X(x)$ ← Useful

Conditional   Marginal

2) If X and Y are independent,

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

$$= f_Y(y)$$ ← Knowing X doesn't change the density of Y.

An implication of (1) is that

$$P(Y \in A) = \int_{-\infty}^{+\infty} f_{Y|X}(y \in A|x) f_X(x)\, dx$$

$$= \int_{y \in A} f_Y(y)\, dy$$

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

Example Select X uniformly at random from an interval [0, 1]. Then select Y uniformly at random from [0, X].

Q: What is the PDF of Y ?

Information we get from problem statement:

$$f_X(x) = 1, \quad x \in [0, 1]$$

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

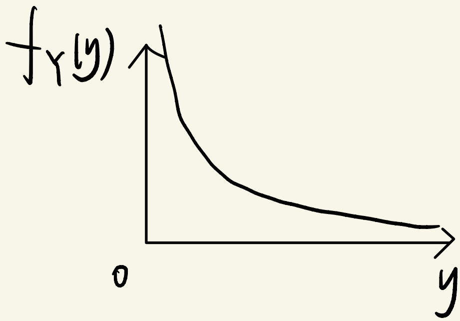
So $f_{XY}(x,y) = f_{Y|X}(y|x) \cdot f_X(x) = \frac{1}{x}, \quad 0 \leq y \leq x \leq 1$

Then we marginalize $f_{XY}(x,y)$ over x,

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx$$

$$= \int_y^1 \frac{1}{x}\,dx = \ln x \Big|_{x=y}^{x=1}$$

$$= \ln 1 - \ln y = -\ln y, \quad 0 \le y \le 1$$



That is, small values of Y are more likely than large values.

$\boxed{\text{Example}}$ Say X and Y are the light intensity measures from a laser at time t_1 and t_2. When the time are close together ($t_1 \approx t_2$), light intensity are correlated with each other. Say that we mode X and Y as jointly Gaussian with $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, correlation coefficient ρ. What is $f_{Y|X}(y|x)$?

To do so, we will use

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \leftarrow \text{already known}$$

Joint PDF $f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)}$

We neend to calculate the marginal PDF

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy$$

$$= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)}\,dy$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}(\underbrace{y^2-2\rho xy+\rho^2x^2}_{(y-\rho x)^2}-(\rho^2-1)x^2)}\,dy$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}[(y-\rho x)^2-(\rho^2-1)x^2]}\,dy$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \int_{-\infty}^{+\infty} \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{(y-\rho x)^2}{(\sqrt{1-\rho^2})^2}}}_{\text{View it as Gaussian with mean } \rho x \text{ and variance } (\sqrt{1-\rho^2})^2}\,dy$$

Integration over a PDF is 1 !

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$ $\leftarrow$ Marginal PDF of X is

with mean $\mu_X = 0$, and

variance $\sigma_X^2 = 1$

Now we can get the conditional PDF

$$f_{Y|X}(y|x) = \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}}$$

Gaussian
mean ρx
variance $1-\rho^2$ $\rightarrow$

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

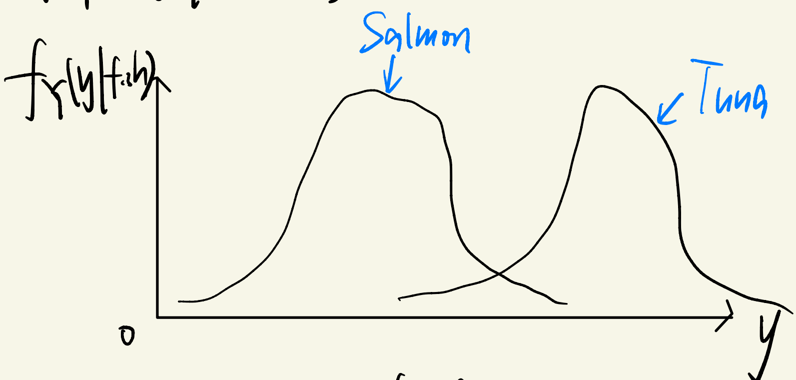
Remark: The conditional PDF of y is centered around ρx. If the observed RV X is $x > 0$, Y is biased to be positive. On the other hand, by the same reasoning, the marginal PDF $f_Y(y)$ is centered around mean 0. Thus, knowing X definitely changes the density of Y.

☐ Where the conditional PDF is useful?

In machine learning, we are often given the conditional PDFs (e.g., after observing realizations).

Classic example: We observe the length of a fish we caught in the Hudson river. We want to determine what kind of fish it is?

$f_Y(y \mid \text{salmon})$ $\qquad$ $f_Y(y \mid \text{Tuna})$



We want to estimate if fish is Tuna or Salmon based on its length y.

We apply Bayes Rule

$$\underbrace{P(\text{Salmon}\,|\,y)} = \frac{f_Y(y\,|\,\text{Salmon})\,P(\text{Salmon})}{f_Y(y)}$$

Prior

$$= \frac{f_Y(y\,|\,\text{salmon})\,P(\text{salmon})}{\sum\limits_{\text{all fish}} f_Y(y\,|\,\text{fish type})\cdot P(\text{fish})}$$

Total probability $\longrightarrow$

$$(*) \qquad = \frac{f_Y(y\,|\,\text{salmon})\,P(\text{salmon})}{f_Y(y\,|\,\text{salmon})\,P(\text{salmon}) + \underbrace{f_Y(y\,|\,\text{tuna})}_{\text{Likelihood}}\,P(\text{tuna})}$$

Likewise, we can also have

$$\underbrace{P(\text{tuna}\,|\,y)}_{\text{Posterior}} = \frac{f_Y(y\,|\,\text{tuna})\,P(\text{tuna})}{(*)}$$

This leads to the famous decision rule

Bayes Decision Rule

Decide Salmon if $P(\text{Salmon}\,|\,y) > P(\text{Tuna}\,|\,y)$

Tuna if $P(\text{Tuna}\,|\,y) > P(\text{Salmon}\,|\,y)$

From (*), we can choose Salmon, i.e.,

$$P(\text{Salmon} \mid y) > P(\text{Tuna} \mid y) \text{ iff}$$

$$f_Y(y \mid \text{salmon}) P(\text{Salmon}) > f_Y(y \mid \text{Tuna}) P(\text{Tuna})$$

$$\Leftrightarrow \frac{f_Y(y \mid \text{Salmon})}{f_Y(y \mid \text{Tuna})} > \frac{P(\text{Tuna})}{P(\text{Salmon})}$$

Two special cases:

1) If $P(\text{Salmon}) = P(\text{Tuna}) = \frac{1}{2}$, then we are just choosing PDF that has a large value at each point of y

$$f_Y(y \mid \text{Salmon}) > f_Y(y \mid \text{tuna})$$

2) If $f_Y(y \mid \text{Salmon})$ and $f_Y(y \mid \text{tuna})$ are same, we end up with choosing whatever class has higher prior.

$$P(\text{Salmon}) > P(\text{Tuna})$$

□ Conditional Expectation

(Discrete) $E[Y|X=x_k] = \sum_{y_j \in \text{All values } Y \text{ take } \{y_1, y_2, \cdots y_k\}} y_j \, P_{Y|X}(y_j|x_k)$

(Continuous) $E[Y|X=x] = \int_{-\infty}^{+\infty} y \, f_{Y|X}(y|x) \, dy$

Clearly, the conditional expectations are a function of x. We can actually think of this expectation as a new Random variable:

$$g(X) = E[Y|X]$$

Big case