

Recall

Conditional stuff for continuous rvs

Let X be a continuous rv; A be an event with $P(\{A\}) > 0$.
Have $f_{X|A}(x)$.

For X, Y both continuous rvs have joint pdf $f_{X,Y}(x,y)$

Defined conditional pdf of X given $Y=y$ via

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \text{derived via limiting process}$$

(Generalizes to >2 rvs)

In both cases, conditional pdf is a "legit pdf" - i.e. it is ≥ 0 and integrates to 1

$$\int_{-\infty}^{+\infty} f_{X|A}(x) dx = 1 \quad \forall A \quad , \quad \int_{-\infty}^{+\infty} f_{X|Y}(x|y) dx = 1 \quad \forall y$$

Conditional Expected Values

Given X, A

$$E(X|A) = \int_{-\infty}^{+\infty} x f_{X|A}(x) dx$$

Given X, Y

$$E(X|Y=y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y) dx \quad \forall y$$

Expected Value Rule

$$\mathbb{E}(g(x)|A) = \int_{-\infty}^{+\infty} g(x) f_{x|A}(x) dx$$

$$\mathbb{E}(g(x)|Y=y) = \int_{-\infty}^{+\infty} g(x) f_{x|Y}(x|y) dx \quad \forall y$$

Recall the **Total Probability**-type results

- If events A_1, A_2, \dots, A_n have >0 probability and partition Ω , then

$$- f_x(x) = \sum_{k=1}^n f_{x|A_k}(x) P(A_k)$$

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

From these follows

Total Expectation Theorems

$$\mathbb{E}(X) = \sum_{k=1}^n \mathbb{E}(X|A_k) P(A_k)$$

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

What about **Independence**?

Super General Definition: X and Y are independent when every pair of events of form $\{X \in V\}$ and $\{Y \in W\}$ are independent events.

Earlier we saw that when X, Y discrete, they are independent iff

$$P_{X,Y}(x,y) = P_X(x)P_Y(y) \quad \forall x,y$$

For ANY pair X, Y (continuous, discrete, or "one and one")

← requires proof

$$X, Y \text{ independent} \iff F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad \forall x,y$$

↓ implies (take $\partial/\partial_x, \partial/\partial_y$)

For X, Y BOTH continuous w/ densities $f_X(x), f_Y(y); f_{X,Y}(x,y)$:

$$X, Y \text{ independent} \iff f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \forall x,y$$

Comment: When X, Y independent, we have

- $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
- $\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$
- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

Example - Gaussian Independence

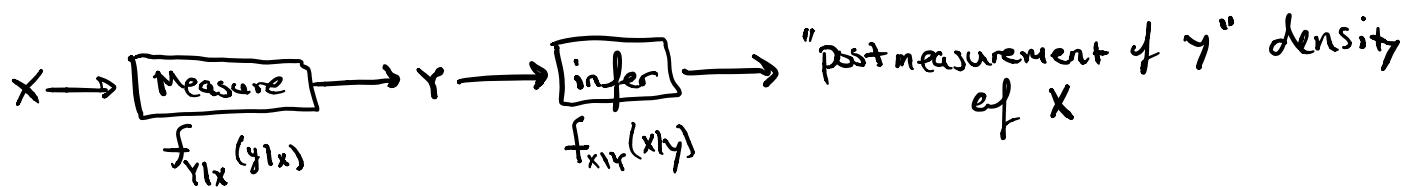
- $X \sim \text{Gaussian}(\mu_1, \sigma_1^2)$
- $Y \sim \text{Gaussian}(\mu_2, \sigma_2^2)$

Assume independent - find $f_{X,Y}(x,y)$

$$\begin{aligned}
 f_{X,Y}(x,y) &= f_X(x)f_Y(y) \\
 &= \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \right) \left(\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \right) \\
 &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x-\mu_1)^2 + (y-\mu_2)^2}{2\sigma_1^2 + \sigma_2^2}}
 \end{aligned}$$

Continuous Baye's Rule

Typical Setting:



Idea: - Have a grip on $f_X(x)$

- Have a good model for $f_{Y|x}(y|x) \neq x, y$
- Want $f_{x|y}(x|y)$

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

Continuous
Baye's
Rule

Example - λ on Some Interval

Say interval = $[1, \frac{3}{2}]$

and

$$f_\lambda(\lambda) = \begin{cases} \lambda, & \lambda \in [1, \frac{3}{2}] \\ 0, & \text{else} \end{cases}$$

X is exponential w/ rate parameter λ

λ = quality of a lightbulb (higher quality \Leftrightarrow lower λ)

Want to infer about λ given observations of X :

- Have $f_\lambda(\lambda)$

- Have $f_{X|\lambda}(x|\lambda)$

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

$$f_{\lambda|X}(\lambda|x) = \frac{f_{X|\lambda}(x|\lambda)f_\lambda(\lambda)}{\int_{-\infty}^{+\infty} f_{X|\lambda}(x|\lambda)f_\lambda(\lambda) d\lambda} = \begin{cases} \frac{2\lambda e^{-\lambda x}}{2 \int_1^{\frac{3}{2}} \lambda e^{-\lambda x} d\lambda}, & \lambda \in [1, \frac{3}{2}] \\ 0, & \text{else} \end{cases}$$

If x larger $\rightarrow f_{\lambda|X}(\lambda|x)$ more focused near $\lambda=1$

A different setting: the unobserved thing you want to infer about is discrete and the observed thing is continuous.

Example - ↓

Say you have an event A. Have $P(A)$, prior probability of A. Observe a continuous rv Y. Have a good model for $f_{Y|A}(y)$

Want: $P(A|Y=y)$. ← since $P(Y=y)=0$ we must define this using a limiting process (as in Lec 19)

$$P(A|Y=y) = \lim_{\delta \rightarrow 0} P(A|Y \in [y-\delta, y+\delta])$$

$$= \lim_{\delta \rightarrow 0} \left(\frac{P(y-\delta \leq Y \leq y+\delta) P(A)}{P(y-\delta \leq Y \leq y+\delta)} \right)$$

$$= \lim_{\delta \rightarrow 0} \frac{\left(\int_{y-\delta}^{y+\delta} f_{Y|A}(t) dt \right) P(A)}{\int_{y-\delta}^{y+\delta} f_Y(t) dt} \cdot \frac{\frac{1}{2\delta}}{\frac{1}{2\delta}}$$

$$= \frac{f_{Y|A}(y) P(A)}{f_Y(y)}$$

After using Total Probability Theorem

$$f_Y(y) = f_{Y|A}(y) P(A) + f_{Y|A^c}(y) P(A^c)$$

$$P(A|Y=y) = \frac{f_{Y|A}(y) P(A)}{f_{Y|A}(y) P(A) + f_{Y|A^c}(y) P(A^c)}$$

Example - Typical Context For This

- A is an event of the form $\{S=s\}$ - "Signa" - S a discrete rv
- Know $P(\{S=s\}) \neq s \in S$
- $Y =$ output of noisy channel you put S through - N
i.e $Y = S + N$
- $N \sim \text{Gaussian}(0, \sigma^2)$
- Observe Y . Want to infer about S

Baye's Rule says that for every s we have

$$P(\{S=s\} | \{Y=y\}) = \frac{f_{Y|S=s}(y) P(\{S=s\})}{\sum_s f_{Y|S=s}(y) P(\{S=s\})}$$

Specialize to $S = \pm 1$

$$Y = S + N; \quad N \sim \text{Gaussian}(0, \sigma^2)$$

Suppose

$$P(\{S=1\}) = p, \quad P(\{S=-1\}) = 1-p$$

Then

$$f_{Y|S=\pm 1}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\pm 1)^2}{2\sigma^2}} \sim \text{Gaussian}(1, \sigma^2)$$

finished in Lec21