Two very useful uses for conditional probability

① Build complex models from simpler ones
use the theorem of total probability

② make inferences (based on partial observation of the experiment)
Bayes Rule

Theorem of total probability, for 2 RVs.

start with $f_{XY}(x,y) = f_X(x|y)\, f_Y(y)$

and integrate over y to get $f_X(x)$

$$\boxed{f_X(x) = \int_{-\infty}^{\infty} f_X(x|y)\, f_Y(y)\, dy}$$

or

$$\boxed{f_Y(y) = \int_{-\infty}^{\infty} f_Y(y|x)\, f_X(x)\, dx}$$

Recall we also had

$$f_X(x) = \sum_{i=1}^{n} f_X(x|B_i)\, P(B_i)$$ if B_i's form a partition

These are connected by considering the partitions B_i to be narrower and narrower slices of Y

Building probability models using conditional probability.

$$\boxed{\begin{aligned} f_{XY}(x,y) &= f_X(x|y)\, f_Y(y) \\ &= f_Y(y|x)\, f_X(x) \end{aligned}}$$

← this is actually simpler than the thm total prob, if you want to find the <u>joint</u> prob

<u>Example</u>: X is exponentially distributed with mean 1

Y is normally distributed with mean 0 and variance $x+1$

↑ value

what is the joint pdf of X and Y?

$$f_X(x) = e^{-x} \quad \text{for} \quad x > 0$$

$$f_Y(y|x) = \frac{1}{\sqrt{2\pi(x+1)}} \exp\left[-\frac{y^2}{2(x+1)}\right] \quad \text{for all } y$$

so $$f_{XY}(x,y) = \frac{e^{-x}}{\sqrt{2\pi(x+1)}} \exp\left[-\frac{y^2}{2(x+1)}\right]$$

for $x > 0$

and $-\infty < y < \infty$

Now if you want $f_Y(y)$, using the theorem of total probability, you have to integrate over x. Details omitted

More examples of building models

Sometimes, we know the RV would be well-modeled by a particular family of RVs (ie, binomial, exponential, Gaussian), but we don't know what the parameters should be (ie, (n,p), λ, (μ, σ^2)).

We can assign a probability to the parameter.

Example: Y is Gaussian mean 0, variance $x+1$
X is exponential mean 1

Example: Y is Bernoulli parameter $p = x$
X is uniform $[0,1]$
Z is Geometric with same p as Y

Example: Y = # likes on social media
(Poisson parameter $\alpha = x$)
X uniform $[3, 5]$
(from an exam question a few years ago)

It is straightforward to form a joint model

$$f_{XY}(x,y) = f_Y(y|x) f_X(x)$$

More examples of defining conditional models

X is the observed value at time t_0

Y is the observed value at time t_1

Example: fraction of pizza remaining

$0 \le x \le 1$, $0 \le y \le 1$

and $y \le x$

Example: A first measurement of anything
A second measurement confirming the same thing

X = underlying actual length (a RV)

Y_1 = 1st measurement
= X + error

Y_2 = 2nd measurement
= X + different error

Once you've constructed the joint model, it is straightforward to apply everything we've learned to compute quantities of interest,

ex $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$, $f_X(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$
(when $f_Y(y) > 0$)

Example: building a more complex model from simpler parts.

Video delivery. Suppose the quality of a video is rated on a Mean Opinion Score (MOS) on a scale from 1 to 5. (MOS is also used to rate the quality of voice calls.) (1 bad, 5 excellent)

Video that is stored on a server already has an upper bound on quality. Denote this X.

When video is delivered over the Internet, quality cannot improve, and often gets worse (due to packet loss, inadequate bandwidth, or recompression at a lower bandwidth).

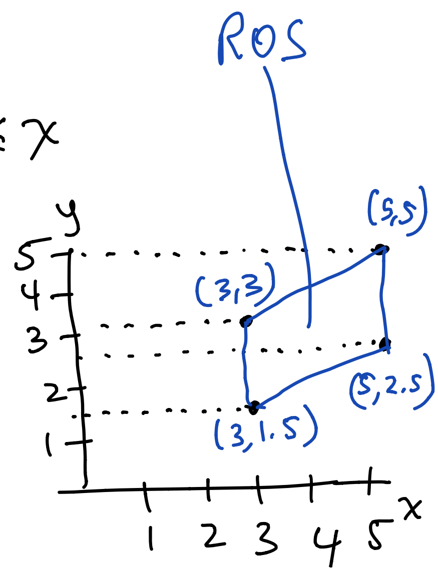
Denote the received quality Y, with $Y \le X$.

Suppose X is uniformly distributed between 3 and 5 and Y is uniformly distributed between $X/2$ and X (a simple but plausible model)

Then $f_X(x) = \begin{cases} 1/2 & \text{when } 3 \le x \le 5 \\ 0 & \text{else} \end{cases}$

and $f_{Y|X}(y) = \begin{cases} \dfrac{1}{x/2} & \frac{x}{2} \le y \le x \\ 0 & \text{else} \end{cases}$

$f_{XY}(x,y) = \begin{cases} 1/x & 3 \le x \le 5 \text{ and } \frac{x}{2} \le y \le x \\ 0 & \text{else} \end{cases}$



Inference – infer the probability of some (possibly unobservable) RV from another related RV that can be observed or measured.

Use Bayes Rule.

Ex: the voltage across some inaccessible piece of a circuit based on a measurement of voltage across a measurable / accessible part of the same circuit.

Recall $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$ if $P(B)>0$

$$\boxed{\text{now: } f_X(x|y) = \frac{f_Y(y|x)f_X(x)}{f_Y(y)} \quad \text{if } f_Y(y)>0}$$

also $P(C|x) = \dfrac{P_X(x|C)P(C)}{P_X(x)}$ for discrete RV

$P(C|x) = \dfrac{f_X(x|C)P(C)}{f_X(x)}$ for continuous RV

This is very useful in cases where we've constructed the model using conditional probability.

Example Binary communications (examples 5.31 and 5.35)

X is input to channel $\qquad p_X(x) = \begin{cases} 1/3 & x=+1 \\ 2/3 & x=-1 \\ 0 & \text{else} \end{cases}$

Y is output, $Y = X + N$

N is noise, Gaussian $N(0,1)$

X and N are independent

At the receiver, we receive Y and want to decide which X was sent.

ie, determine $f_X(x/y)$ and use it...

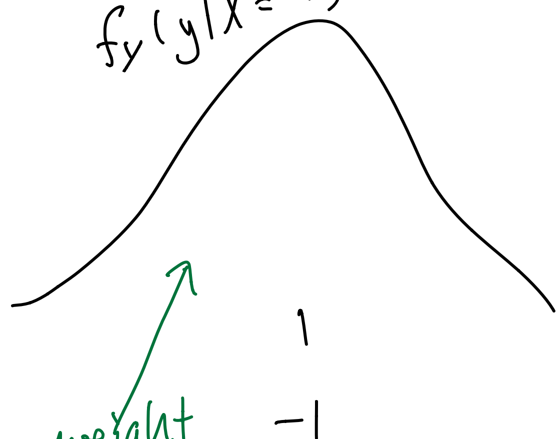
$$f_Y(y) = f_Y(y \mid X=+1)\, P(X=+1) + f_Y(y \mid X=-1)\, P(X=-1)$$

Given $X=+1$, $Y = 1+N \Rightarrow Y \sim N(1,1)$

Given $X=-1$, $Y = -1+N \Rightarrow Y \sim N(-1,1)$

$f_Y(y \mid x=+1)$

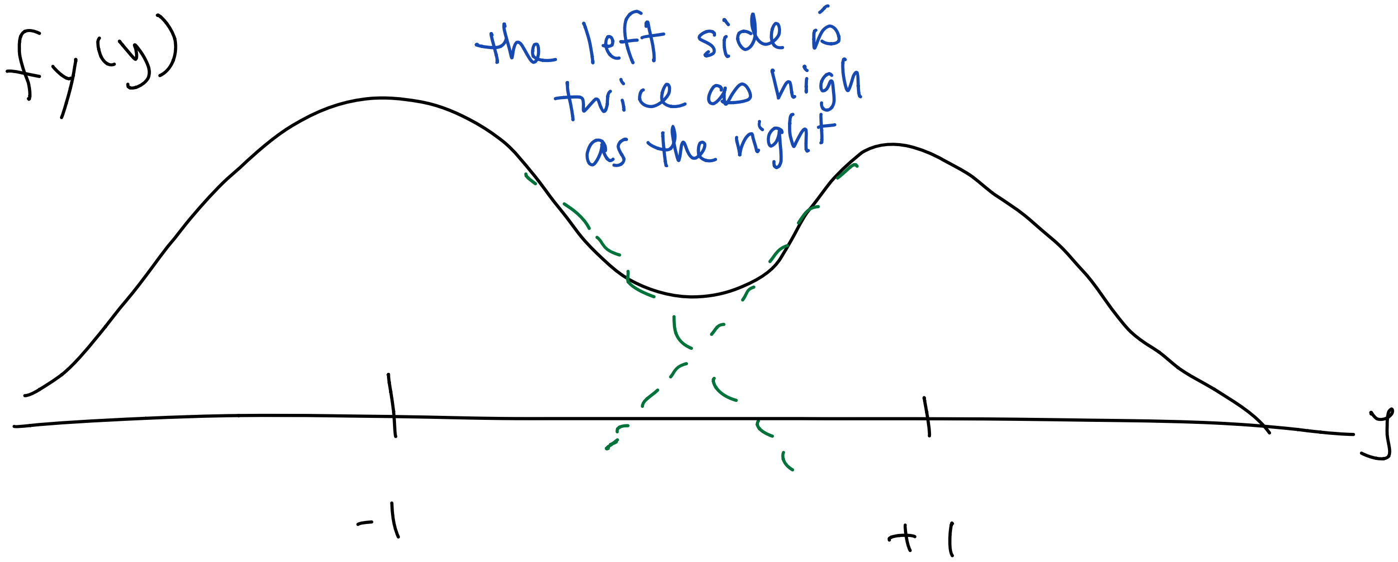
$f_Y(y \mid X=-1)$



weight by $p_X(+1)$

weight by $p_X(-1)$

sum to get $f_Y(y)$



Recall $\Phi(-1)$ is 0.1587, not very close to zero, so there is significant overlap.

What's the probability the signal was $X=+1$ if we receive the value y?

Use Bayes Rule

$$P_X(X=1 \mid y) = \frac{f_Y(y \mid X=+1)\, P(X=+1)}{f_Y(y)}$$

use theorem of total probability to get $f_Y(y)$

$$= \frac{\frac{1}{3} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-1)^2}{2}\right)}{\frac{1}{3} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-1)^2}{2}\right) + \frac{2}{3} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y+1)^2}{2}\right)}$$

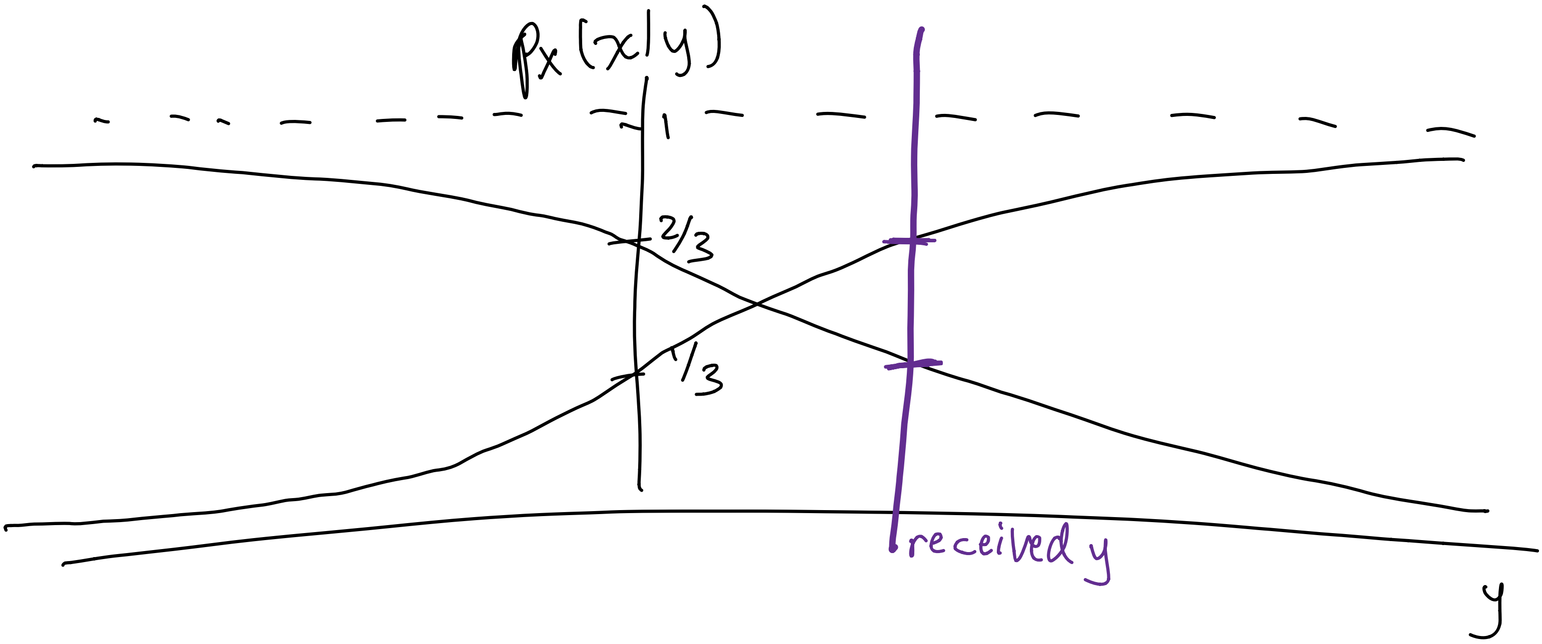
simplify: cancel $\exp\left(-\frac{(y-1)^2}{2}\right)$ from numerator + denominator

$$= \frac{1}{1 + 2\exp(-2y)} = P_X(X=1 \mid y)$$

Similarly, we can show

$$p_X(X=-1|y) = \frac{2\exp(-2y)}{1+2\exp(-2y)}$$

$$= \frac{2}{2+\exp(2y)}$$



Choose $\hat{X} = +1$ when $P(X=1|y) > P(X=-1|y)$

choose $\hat{X} = -1$ when $P(X=1|y) < P(X=-1|y)$

Maximum Aposteriori Probability (MAP) detector