Conditional pmfs (section 3.4)
Conditional cdfs (section 4.2.2)
Conditional pdfs (section 4.2.2)

$\Bigg\}$ Conditioned on an event

These combine the concept of RVs with the concept of conditional probability.
New notation and new insights – same fundamentals.

The conditioning event, C describes the partial information we may have about the RV X or its underlying experiment.

Let X be a RV with pmf $p_X(x)$
cdf $F_X(x)$
and/or pdf $f_X(x)$

C is an event, $P(C)>0$.

Definitions

conditional pmf $\quad p_X(x|C) = P(\{X=x\}|C)$

conditional cdf $\quad F_X(x|C) = \dfrac{P(\{X\le x\}\cap C)}{P(C)}$

conditional pdf $\quad f_X(x|C) = \dfrac{d}{dx} F_X(x|C)$

Note: $p_X(x|C)$ is a pmf and has all the properties of a pmf.
$F_X(x|C)$ is a cdf. $f_X(x|C)$ is a pdf.

Recall: for conditional probability we had

1. definition
2. theorem of total probability (building more complicated models)
3. Bayes Rule / inference (improving our knowledge)

we'll consider these again for this new scenario w/ RVs.

we'll also consider conditional expectation and conditional variance

Theorem of total probability for RVs.

Suppose $B_1, B_2, \ldots, B_n$ partition the sample space.

$$\boxed{\begin{aligned} p_X(x) &= \sum_{i=1}^{n} p_X(x \mid B_i)\, P(B_i) \\ F_X(x) &= \sum_{i=1}^{n} F_X(x \mid B_i)\, P(B_i) \\ f_X(x) &= \sum_{i=1}^{n} f_X(x \mid B_i)\, P(B_i) \end{aligned}}$$

(3.24) on page 113

(4.25) on page 153

(4.26) on page 153

Use this to build more complicated models from several simpler ones.

Example You want to model the height of the trees in a forest, and you know there are 3 types of trees, each with their own height distribution. $F_H(h \mid T_i)$ where T_i is the tree type

and you know $P(T_i)$.

Then overall model of tree height in the forest is

$$F_H(h) = F_H(h \mid T_1)\, P(T_1) + F_H(h \mid T_2)\, P(T_2) + F_H(h \mid T_3)\, P(T_3)$$

Example A production line creates two kinds of devices. Type 1 devices occur with probability α and have a life time governed by

$$p_X(x \mid B_1) = (1-r)^{x-1} r \qquad x = 1, 2, \dots$$

Type 2 devices occur w/ prob. $1-\alpha$ and have life time

$$p_X(x \mid B_2) = (1-s)^{x-1} s \qquad x = 1, 2 \dots$$

Select a device and observe its lifetime

$$p_X(x) = p_X(x \mid B_1) P(B_1) + p_X(x \mid B_2) P(B_2)$$

Gaussian mixtures are a popular model.

- communication signals
- speech signals
- pixel values in image and video

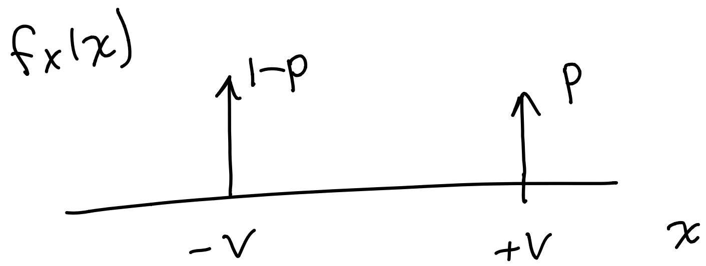
Example 4.11: Communication system with Gaussian noise

- send a 0 by sending a signal w/ $-v$ volts
- send a 1 by sending a signal w/ $+v$ volts.

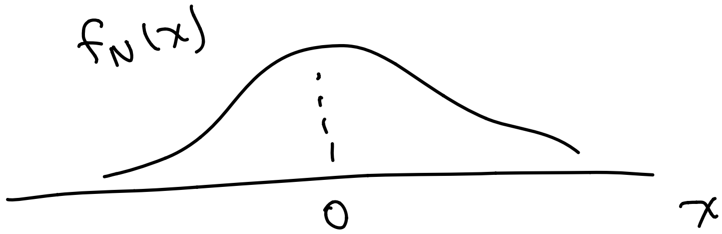
$P("1") = p = 1 - P("0")$. $B_1 = \{\text{send } 1\}$ $B_0 = \{\text{send } 0\}$

Send signal X, receive signal $X + N$

N is a random noise voltage, Gaussian $N(0, \sigma^2)$



pdf of signal X

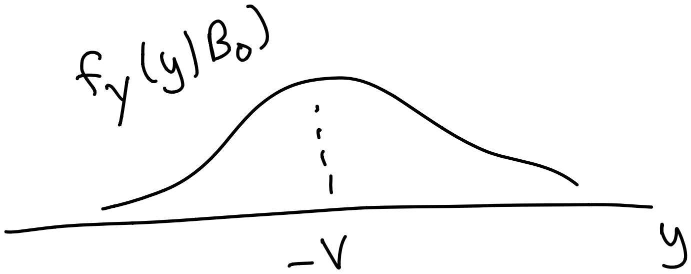


pdf of noise N

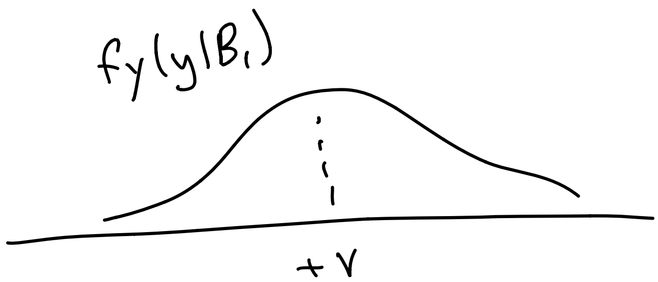
If $B_1 = \{\text{send } 1\}$, then $X = V$ and $Y = +V + N$

so $Y|B_1$ is gaussian $N(v, \sigma^2)$

and similarly $Y|B_0 \sim N(-v, \sigma^2)$



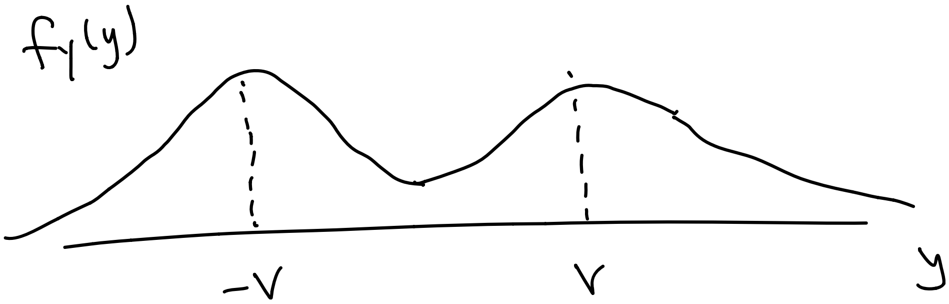
$Y|B_0 \sim N(-v, \sigma^2)$



$Y|B_1 \sim N(+v, \sigma^2)$

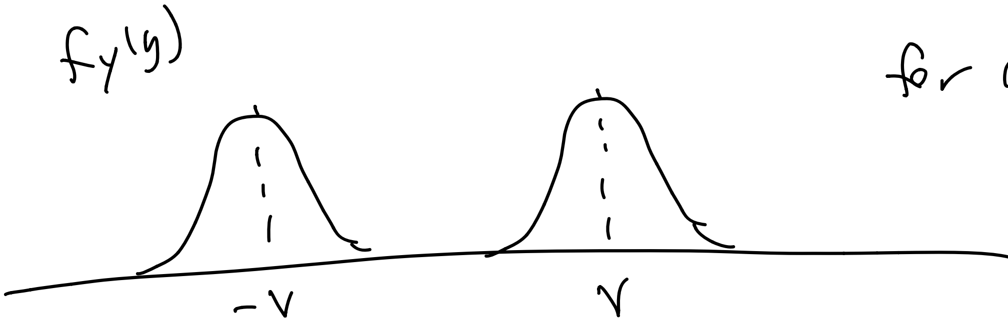
And the received signal Y is in general a mixture of Gaussians

$$f_Y(y) = f_Y(y|B_1)\,P(B_1) + f_Y(y|B_0)\,P(B_0)$$



pdf of signal Y

Note: the relative values of v and σ^2 control the shape of the pdf and the separability of the 2 cases



for a smaller σ^2.

Inference : what happens if our conditioning event depends on the values of the RV X?

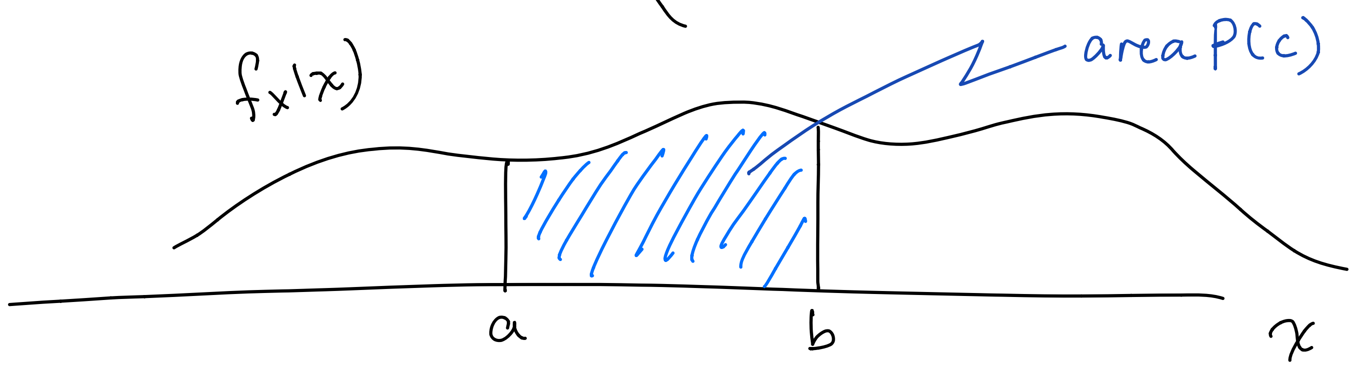
This is an important special case
- allows us to update our understanding of X, and incorporate knowledge we gain by observing the out come of our experiment.

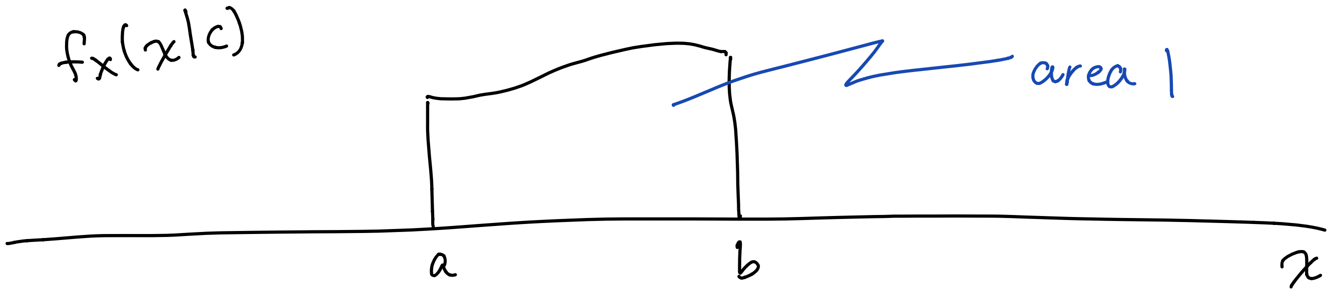
Given RV X, event C, where C depends on X

given $f_X(x)$ and $C = \{a < X \le b\}$

we can easily compute $P(C) = \int_a^b f_X(x)\,dx$

Then
$$f_X(x|C) = \begin{cases} \dfrac{f_X(x)}{P(C)} & \text{when } x \in C \\ 0 & \text{otherwise} \end{cases}$$





Intuition: if C happened, then X could not have taken values outside of $C = \{a < X \le b\}$,

so $f_X(x|C) = 0$ for x not in C.

But $f_X(x|C)$ must be a pdf,

so $$\int_{-\infty}^{\infty} f_X(x|C)\,dx = 1$$

And we know $$\int_a^b f_X(x)\,dx = P(C)$$

so $$\int_a^b \frac{f_X(x)}{P(C)}\,dx = 1$$ and $$f_X(x|C) = \frac{f_X(x)}{P(C)}$$ when $x \in C$.

You need to remember both pieces!

Chop (narrow the sample space)

Scale (renormalize)

These are the same 2 steps we used when we first considered conditional probability

Why chop and scale? The mathematical derivation

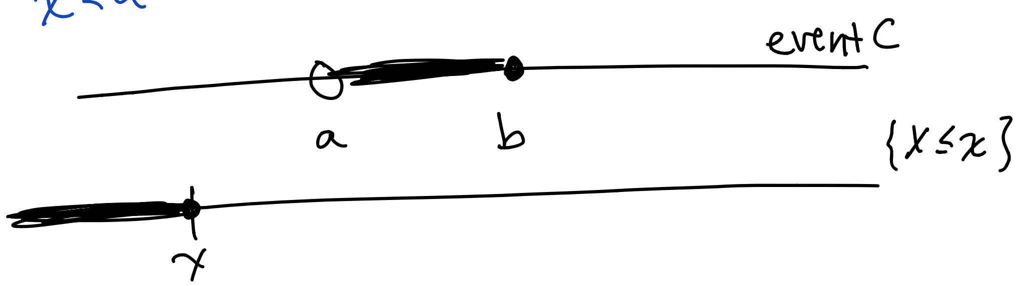
recall conditional CDF

$$F_X(x \mid C) = P(X \le x \mid C)$$

$$= \frac{P(\{X \le x\} \cap \{a < X \le b\})}{P(\{a < X \le b\})}$$

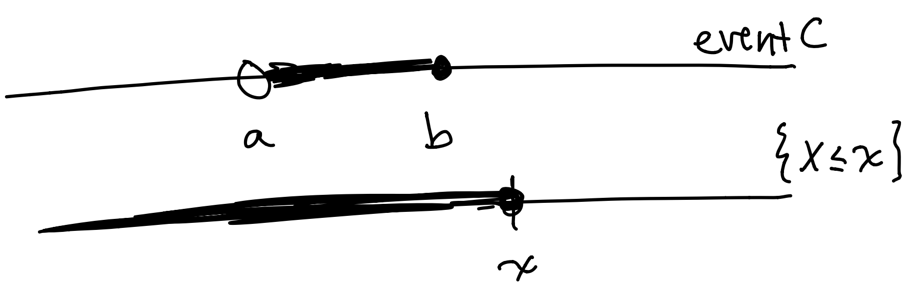
To compute this for all values of x, we will need to look at the different ways that $\{X \le x\}$ and $\{a < X \le b\}$ overlap

① $x < a$



$F_X(x \mid C) = 0$ if $x < a$

② $x > b$

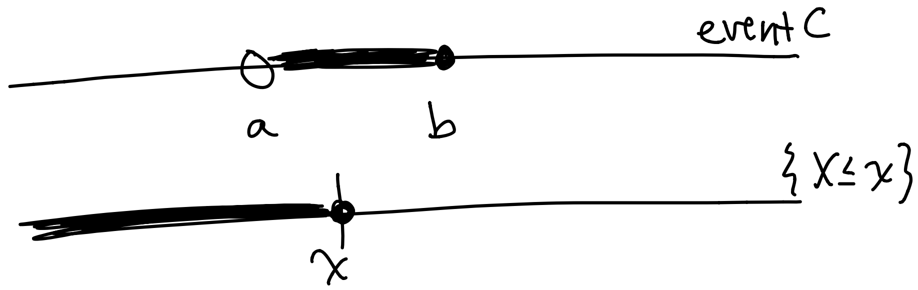


$F_X(x \mid C) = 1$ if $x > b$

Because $\{X \le x\} \cap C = C$ in this region

$$F_X(x \mid C) = \frac{P(X \le x \cap C)}{P(C)} = \frac{P(C)}{P(C)} = 1 \quad \text{if } x > b$$

③ $a < x \le b$



$$F_X(x|C) = \frac{P(a < X \le x)}{P(C)} = \frac{F_X(x) - F_X(a)}{P(C)}$$

for $a \le x \le b$

Combining

$$F_X(x|C) = \begin{cases} 0 & x < a \\ \dfrac{F_X(x) - F_X(a)}{P(C)} & a \le x \le b \\ 1 & x > b \end{cases}$$

if $C = \{a < X \le b\}$

Differentiate to get conditional pdf:

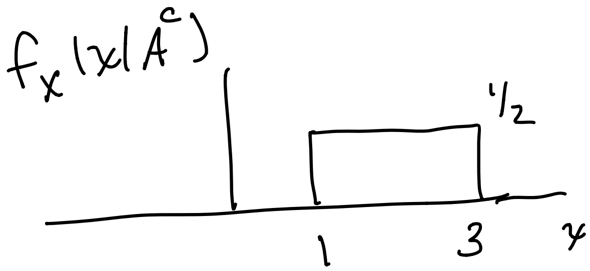
$$f_X(x|C) = \begin{cases} 0 & x < a \\ \dfrac{f_X(x)}{P(C)} & a < x \le b \\ 0 & x > b \end{cases}$$

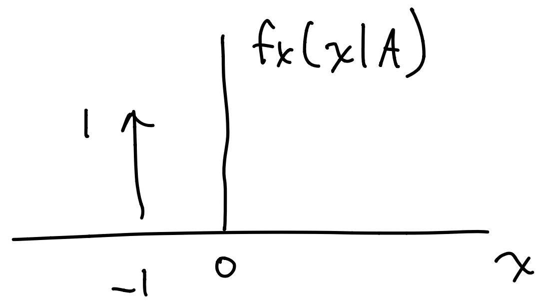
chop

scale

These are "2 sides of the same coin".

Example. Suppose when event A^c happens, a RV X is uniformly distributed on $(1,3)$, but when A happens, X is a constant -1.



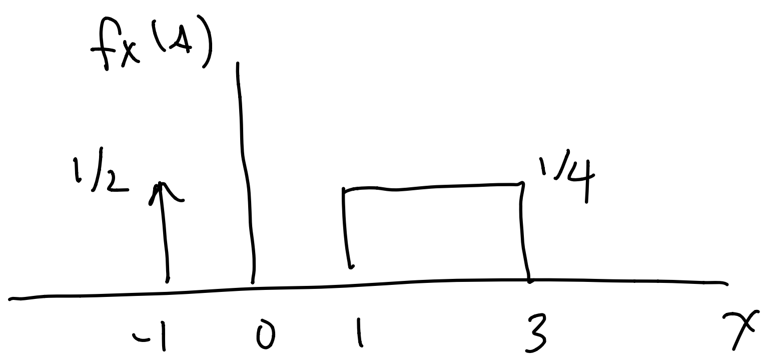


$f_X(x|A^c) = [u(x-1) - u(x-3)]\frac{1}{2}$ $\qquad$ $f_X(x|A) = \delta(x+1)$

If $P(A) = P(A^c) = 1/2$, then

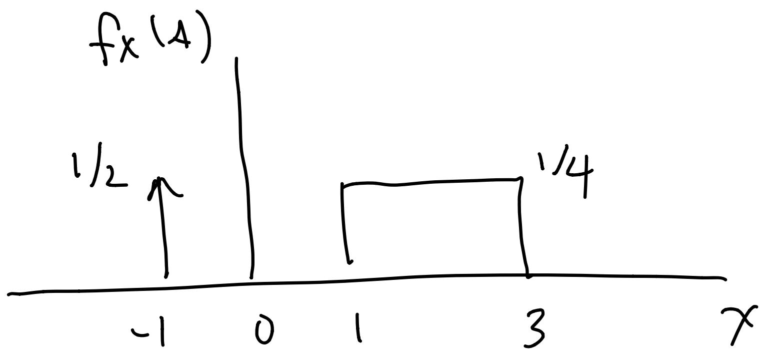
$$f_X(x) = P(A)\, f_X(x|A) + P(A^c)\, f_X(x|A^c)$$

$$= \frac{1}{2}\left[\frac{1}{2}u(x-1) - \frac{1}{2}u(x-3) + \delta(x+1)\right]$$



Note: $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

The flip side: Take $f_X(x)$ from the previous example and let $A = \{X < 0\}$.



what's the conditional pdf of X given A?

Answer The "long way": when $x \in A^c$, $f_X(x|A) = 0$.

$$P(A) = P(X<0) = 1/2$$

when $x \in A$, the only nonzero component is at $x=-1$, which is $\frac{1}{2}\delta(x+1)$.

so
$$f_X(x|A) = \begin{cases} \dfrac{\frac{1}{2}\delta(x+1)}{P(A)} & x \in A \\ 0 & x \in A^c \end{cases}$$

$$= \delta(x+1)$$

Note: $\int f_X(x|A)\,dx = 1$

similarly, $P(A^c) = \int_0^\infty f_X(x)\,dx = 1/4(3-1) = 1/2$

so
$$f_X(x|A^c) = \begin{cases} \dfrac{\frac{1}{2}[u(x-1)-u(x-3)]}{P(A^c)} & x \in A^c \\ 0 & x \in A \end{cases}$$

$$f_X(x|A^c) = \frac{1}{2}[u(x-1)-u(x-3)]$$

Note: $\int f_X(x|A^c)\,dx = 1$

The intuitive way.

When $A = \{X < 0\}$, we have just the $\delta(x+1)$ shape and it needs to be a pdf, so

$$f_X(x|A) = \delta(x+1).$$

When A^c happens, the remaining uncertainty has the uniform shape and it needs to be a pdf, so

$$f_X(x|A^c) = \frac{1}{2}\left[u(x-1) - u(x-3)\right].$$

The scaling happens automatically, and "just so happens" equals $1/P(A)$ (if A is the conditioning event.)

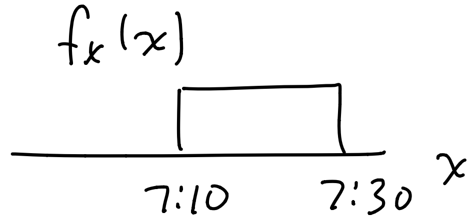
Another example of a mixed RV

The City Bus arrives at the bus stop by your home every 15 minutes starting at 6am. You arrive at the bus stop between 7:10 and 7:30, with the time being a uniform random variable in this interval. What is the pdf of the time you have to wait?

Answer:

X = your arrival time

Y = your wait time



Let $A = \{$ you get on the 7:15 bus $\}$

$B = \{$ you get on the 7:30 bus $\}$

$A \cup B = S$.

$A = \{ 7{:}10 \le x \le 7{:}15 \}$

$B = \{ 7{:}15 < x \le 7{:}30 \}$

Conditioned on A, your arrival is uniform $[7{:}10, 7{:}15]$.

so your wait conditioned on A is also uniform $[0,5]$

$$f_Y(y|A) = \begin{cases} 1/5 & 0 \le y \le 5 \\ 0 & \text{else} \end{cases}$$

Conditioned on B, your wait is uniform $[0,15]$.

$$f_Y(y) = f_Y(y|A)P(A) + f_Y(y|B)P(B)$$

$$= \begin{cases} 1/5 \cdot 1/4 + 1/15 \cdot 3/4 & 0 \le y \le 5 \\ 1/15 \cdot 3/4 & 5 < y \le 15 \end{cases} = \begin{cases} 1/10 \\ 1/20 \end{cases}$$