5/7/19 Lecture 11

Technical Difficulties so used the blackboard to go over Quiz 5

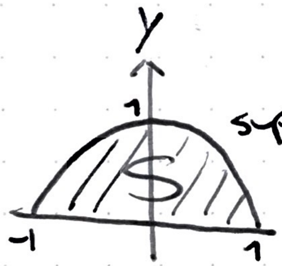
$$f_{XY}(x,y) = \begin{cases} cx^2 & \text{for } 0 \le y \le 1-x^2 \\ 0 & \text{else} \end{cases}$$

Bivariate distribution → 2 variables

Support is in two dimensions

All pdfs have 2 attributes

1) area = 1    2) never negative



support $S: \{(x,y): 0 \le y \le 1-x^2\}$

$$\iint_S f_{XY}(x,y)\,dy\,dx = 1$$

$$\int_0^1 \int_{-\sqrt{1-y}}^{\sqrt{1-y}} cx^2\,dx\,dy = \int_{-1}^{1} \int_0^{1-x^2} cx^2\,dy\,dx = \frac{4}{15}c = 1 \qquad c = \frac{15}{4}$$

Now back to normal lecture.

Consequences of the definition of bivariate CDFs:

1) If $(X, Y)$ has the joint CDF $F_{XY}(x,y)$, you can obtain the marginal CDF $F_X(x)$ from the joint CDF as $F_X(x) = \lim\limits_{y \to \infty} F_{XY}(x,y)$

2) The joint pdf and joint CDF are related in a manner similar to their relationship in the univariate (one rv at a time) case:

If $(X, Y)$ have a joint pdf $f_{XY}(x, y)$ then

$$F_{XY}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(r, s)\, dr\, ds$$

$$= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(r, s)\, ds\, dr$$

and $f_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y) = \frac{\partial^2}{\partial y \partial x} F_{XY}(x, y)$

at every $(x, y)$ where the partial derivatives exist

3) If $(X, Y)$ have a discrete joint distribution with joint pmf $f_{XY}(x, y)$, then the marginal pmf $f_X(x)$ of $X$ is $f_X(x) = \sum_y f_{XY}(x, y)$

(and similarly for $f_Y(y)$)

$(X, Y)$ : discrete  You have $f_{XY}(x, y)$

$$f_X(x) = \sum_{\text{any } y} f_{XY}(x, y)$$

The idea behind marginal distributions is that it's harder to visualize a joint (2-dimensional) distribution than it is to visualize each of its 1-dimensional marginal distributions

4) If $(X, Y)$ have a continuous joint distribution with joint pdf $f_{XY}(x,y)$ the marginal pdf $f_{XY}(x,y)$, the marginal pdf $f_X(x)$ of $X$ is

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy \quad (\text{for all } -\infty < x < \infty)$$

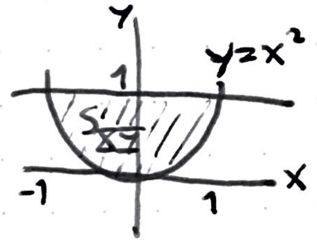
↑ marginalizing out $Y$

and the marginal pdf $f_Y(y)$ of $Y$ is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx \quad (\text{for all } -\infty < y < \infty)$$

Earlier example continued

$(X, Y)$ have joint pdf $f_{XY}(x,y) = \begin{cases} \frac{21}{4}x^2 y & \overbrace{0 \le x^2 \le y \le 1}^{S_{XY}} \\ 0 & \text{else} \end{cases}$

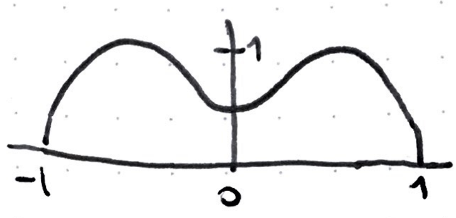



You can see from the sketch of the support $S_{XY}$ of $f_{XY}(x,y)$ that $-1 \le X \le 1$, so the support $S_X$ of $X$ is $(-1,1)$

and its marginal pdf is $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$

$$= \int_{x^2}^{1} \frac{21}{4}x^2 y\,dy = \frac{21}{8}x^2(1-x^4)$$

so $f_X(x) = \begin{cases} \frac{21}{8}x^2(1-x^4) & \text{for } -1 < x < 1 \\ 0 & \text{else} \end{cases}$

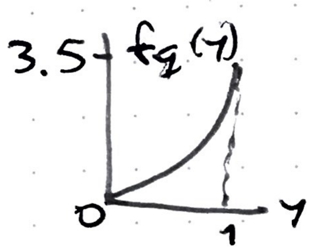



This is a weird pdf (supposed to be symmetric and bimodal)

Similarly, the support $S_{\mathbb{Y}}$ of $\mathbb{Y}$ is $(0,1)$ and its marginal pdf is

$$f_{\mathbb{Y}}(y) = \int_{-\infty}^{\infty} f_{\mathbb{X}\mathbb{Y}}(x,y)\,dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4} x^2 y \, dx$$

$$= \begin{cases} \frac{7}{2} y^{5/2} & \text{for } 0 < y < 1 \\ 0 & \text{else} \end{cases}$$




5) If you have the joint dist. $f_{\mathbb{X}\mathbb{Y}}(x,y)$ you can reconstruct the marginals $f_{\mathbb{X}}(x)$ and $f_{\mathbb{Y}}(y)$ but not the other way around:

If all you have is the marginals, in general they do not uniquely determine the joint.

<u>Ex</u>: Case 1: $\mathbb{X}$ = # heads in $n$ tosses of fair coin 1 and independently $\mathbb{Y}$ = # heads in $n$ tosses of fair coin 2

$\mathbb{X} \sim \text{Binomial}(n, \frac{1}{2})$ so

$$f_{\mathbb{X}}(x) = \begin{Bmatrix} \binom{n}{x}\left(\frac{1}{2}\right)^x \left(1-\frac{1}{2}\right)^{n-x} & x = 0,1,\ldots,n \\ 0 & \text{else} \end{Bmatrix}$$

and $\mathbb{Y} \sim \text{Binomial}(n, \frac{1}{2})$ also so

$$f_{\mathbb{Y}}(y) = \begin{Bmatrix} \binom{n}{y}\left(\frac{1}{2}\right)^n & y = 0,1,\ldots,n \\ 0 & \text{else} \end{Bmatrix}$$

Since $X$ and $Y$ are independent in case 1,

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

If $A, B$ independent $\rightarrow P(A \text{ and } B) = P(A) \cdot P(B)$

T/F statements

so $f_{XY}(x,y) = \begin{cases} \binom{n}{x}\binom{n}{y}\left(\frac{1}{2}\right)^{2n} & \text{for } x = 0,1,\dots,n \text{ and } y = 0,1,\dots,n \\ 0 & \text{else} \end{cases}$

Case 2: $X$ = # heads in $n$ tosses of fair coin 1

$Y = X$

$X \sim \text{Binomial}(n, \frac{1}{2})$ and so is $Y$ but their joint distribution (from $Y = X$) is

$f_{XY}(x,y) = \begin{cases} \binom{n}{x}\left(\frac{1}{2}\right)^n & \text{for } x = y = 0,\dots,n \\ 0 & \text{else} \end{cases}$

There is one situation in which the marginals do uniquely determine the joint: when $X$ and $Y$ are independent.

<u>Def</u>: rvs $X$ and $Y$ are independent if for every (non-weird) set $A$ and $B$ of real numbers

$P(X \in A \text{ and } Y \in B) = P(X \in A) \cdot P(Y \in B)$

<u>Consequences</u>

1) Immediately you get that if $X$ and $Y$ are independent

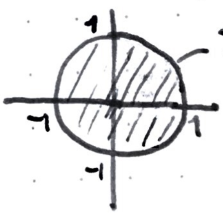
$$F_{XY}(x,y) = P(X \le x \text{ and } Y \le y)$$
$$= P(X \le x)\, P(Y \le y)$$
$$= F_X(x) \cdot F_Y(y)$$

This is iff the converse is also true

2) Differentiate this equation once with respect to y to get the result

$X, Y$ independent $\overset{\text{iff}}{\Longleftrightarrow}$ $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$

Ex: Suppose that continuous rvs $X$ and $Y$ have joint pdf




$$f_{XY}(x,y) = \begin{cases} kx^2y^2 & \text{for } 0 \le x^2+y^2 \le 1 \\ 0 & \text{else} \end{cases}$$

The support $S_{XY}$ of $f_{XY}$ is the region inside the unit circle.

You can evaluate the normalizing constant by computing $\iint_{S_{XY}} kx^2y^2\,dx\,dy$ and setting it equal to 1

$$1 = \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} kx^2y^2\,dy\,dx = \frac{\pi}{24} \quad \text{so} \quad k = \frac{24}{\pi}$$

Q: Are $X$ and $Y$ independent?

A: No, they can't be since the points $(x,y)$ with positive density satisfy $x^2+y^2 \le 1$, for any given value of $y$ of $Y$, the possible values of $X$ depend on $y$ and vice versa

Ex: Continuous rv $X$ and $Y$ have joint pdf

$$f_{XY}(x,y) = \begin{cases} ke^{-(x+2y)} & \text{for } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{else} \end{cases}$$

Q: Are $X$ and $Y$ independent?

A: Yes because $e^{-(x+2y)}$ factors into $(e^{-x})(e^{-2y})$ and the support $S_{XY}$ also factors: $(x \geq 0) \circ (y \geq 0)$

Just choose $(k, k_x, k_y)$ such that $\iint_{S_{XY}} ke^{-(x+2y)}\,dx\,dy = 1$

$\int_0^\infty k_x e^{-x}\,dx = 1$, $\int_0^\infty k_y e^{-2y}\,dy = 1$ and $k = k_x \cdot k_y$

You get $k_x = 1$, $k_y = 2$, $k = 2$

## Conditional probability distributions

Recalling that for two events $A$ and $B$,

$P(B|A) = \dfrac{P(A \text{ and } B)}{P(A)}$ (as long as $P(A) > 0$), we should be able to extend this idea to random variables

Start with $X$ and $Y$ both discrete, so that we can talk about $P(Y=y \mid X=x)$

Def: If $X$ and $Y$ have a discrete joint distribution w/ joint pmf $f_{XY}(x,y)$ and $X$ has marginal pmf $f_X(x)$ then for each $x$ such that $f_X(x) > 0$ define

$$f_{Y|X}(y|x) \triangleq \frac{f_{XY}(x,y)}{f_X(x)}$$ to be the conditional pmf of $Y$ given $X$

$= P(Y=y \mid X=x)$

Now let's do the analogous thing for continuous r.v.s.

Def: If $X$ and $Y$ have a continuous joint distribution with joint pdf $f_{XY}(x,y)$ and $X$ has continuous marginal pdf $f_X(x)$, then for each $x$ such that $f_X(x) > 0$, define $f_{Y|X}(y|x) = \left\{ \frac{f_{XY}(x,y)}{f_X(x)} \right.$ to be the conditional pdf of $Y$ given $X$.

Continuing an earlier ex.

$X, Y$ have joint pdf $f_{XY}(x,y) = \begin{cases} \frac{21}{4}x^2y & \text{for } 0 \le x^2 \le y \le 1 \\ 0 & \text{else} \end{cases}$

Let's work out $f_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$

Earlier we saw that $f_X(x) = \begin{cases} \frac{21}{8}x^2(1-x^4) & \text{for } -1 \le x \le 1 \\ 0 & \text{else} \end{cases}$

and $f_Y(y) = \begin{cases} \frac{7}{2}y^{5/2} & \text{for } 0 \le y \le 1 \\ 0 & \text{else} \end{cases}$

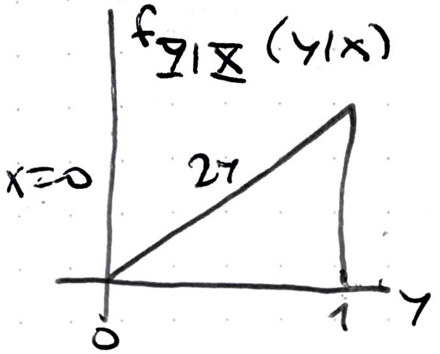
Immediately, then, for all $x$ for which $f_X(x) > 0$, namely $-1 < x < 1$

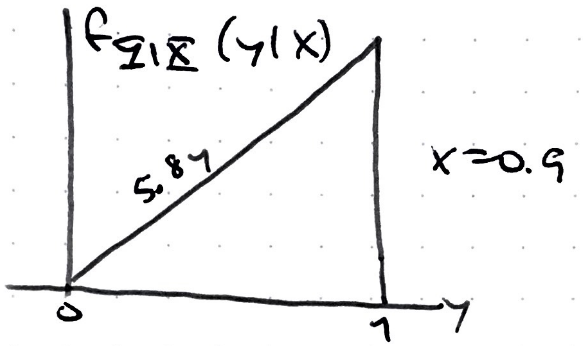
$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \begin{cases} \dfrac{\frac{21}{4}x^2y}{\frac{21}{8}x^2(1-x^4)} & \text{for } 0 \le x^2 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

This simplifies to

$$f_{Y|X}(y|x) = \begin{cases} \frac{2y}{1-x^4} & 0 \le x^2 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

A few "slices" of this:



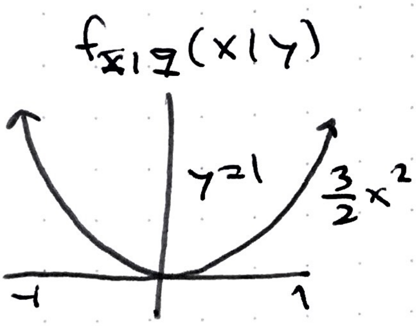

And in the other direction

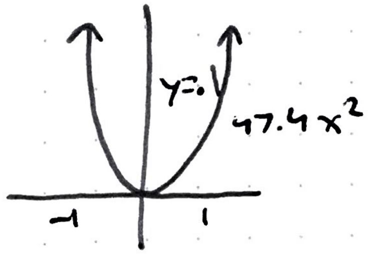
$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad \text{for } 0 \le y \le 1$$

$$= \frac{\frac{21}{4}x^2 y}{\frac{7}{2}y^{5/2}} = \frac{3x^2}{2y^{3/2}} = \frac{3}{2}x^2 y^{-3/2}$$

$$= \begin{cases} \frac{3}{2}x^2 y^{-3/2} & \text{for } 0 \le x^2 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

A few "slices" of this"



When $X$ and $Y$ are continuous, computing $f_{Y|X}(y|x)$ may seem to involve conditioning on the event $X=x$, which (as we saw earlier) has probability zero.

But that's not what's actually going on; strictly speaking $f_{Y|X}(y|x)$ is a limit:

$$f_{Y|X}(y|x^*) = \lim_{\epsilon \downarrow 0} \frac{d}{dy} P\left(Y \le y \mid x^* - \frac{\epsilon}{2} \le X \le x^* + \frac{\epsilon}{2}\right)$$

In other words, you take a little strip of $x$ values of width $\epsilon$ around $X = x^*$, compute $P(Y \le y | X$ is in the strip), differentiate the result with respect to $y$, and let $\epsilon$ go to 0.

Thus you can think of $f_{Y|X}(y|x)$ as the conditional pdf of $Y$ given that $X$ is close to $x$.

<u>Constructing a joint pdf from marginals & conditionals</u>

We know that as long as no divisions by 0 happen

$$f_{Y|X}(y|x) = f_{XY}(x,y) / f_X(x) \quad (1)$$

$$f_{X|Y}(x,y) = f_{XY}(x,y) / f_Y(y) \quad (2)$$

Multiply (1) by $f_X(x)$ and equation (2) by $f_Y(y)$ to get

$$f_{XY} = f_X(x) f_{Y|X}(y|x)$$

$$= f_Y(y) f_{X|Y}(x|y)$$

$$P(A \text{ and } B) = P(A)\,P(B|A)$$
$$= P(B)\,P(A|B)$$

So there are two ways to construct a joint pdf from a marginal pdf and a conditional pdf

<u>Case Study</u>: Bayesian Statistical Analysis

A machine produces nuts  and bolts , and the nut paired with a particular bolt in the manufacturing process is supposed to fit snuggly on the bolt.

Let's call a (nut, bolt) pair defective if the correct snug fit doesn't happen (e.g. bolt diameter either too big or too small, or nut diameter too small or too big)

Let $\theta$ = proportion of defective bolts if the machine were allowed to run for an indefinitely long period

Since we can only observe the machine for a finite (short) time interval, $\theta$ is unknown.

To learn about $\theta$, we could take a random sample of (nut, bolt) pairs of size $n$ ("with replacement") and count the # of defectives in the sample (call this $N$)

Let $D_i = \begin{cases} 1 & \text{if (nut, bolt) pair is defective} \\ 0 & \text{else} \end{cases}$

$(D_i \mid \theta) \overset{IID}{\sim} \text{Bernoulli}(\theta)$ for $(i = 1, \dots, n)$

$\underset{\uparrow}{I}\underset{\underset{\uparrow}{\sim}}{ID}$ $\quad N = \sum_{i=1}^{m} D_i$

v stationarity

So the conditional pmf of rv $N$ is $f_{N|\theta}(n \mid \underset{\text{fixed and known}}{m}, \theta)$

$= \begin{cases} \binom{m}{n} \theta^n (1-\theta)^{m-n} & \text{for } n = 0, 1, \dots, m \\ 0 & \text{else} \end{cases}$

Suppose that $m = 114$ $N = 3$

A reasonable estimate of $\theta$ would be $\hat{\theta} = \frac{N}{m} = \frac{3}{114} = 2.6\%$

but how much uncertainty do we have about $\theta$ on the basis of this dataset?

<u>Bayesian story</u> $\theta$ unknown continuous $\in (0,1)$

vector $\underset{\sim}{D} = (D_1, \dots, D_m)$ data set

Probability: $P(\text{data} \mid \text{unknown}) = \text{easy}$

$P(N \mid \theta) = (*)$

$P(\underset{\sim}{D} \mid \theta) \overset{?}{=} p(\theta \mid \underset{\sim}{D})$

Bayes Theorem: $P(\theta \mid \underset{\sim}{D}) = \dfrac{p(\theta)\, p(\underset{\sim}{D} \mid \theta)}{P(\underset{\sim}{D})}$

$P(\theta \mid N) = P(\theta)\, p(N \mid \theta) / P(N) \leftarrow$ normalizing constant

total info about $\theta$ — info about $\theta$ external to dataset — info about $\theta$ internal to dataset