If you model $(X \mid \theta)$ as Bernoulli(θ) 104

and $\theta \sim$ Uniform$(0,1)$

the joint pmf/pdf of (X, θ) would be

$$f_{X,\Theta}(x, \theta) = \begin{cases} \theta^x (1-\theta)^{1-x} & \text{for } \begin{pmatrix} x = 0, 1 \\ 0 < \theta < 1 \end{pmatrix} \\ 0 & \text{else} \end{cases}$$

pf/pdf

Then (e.g.) $P(X = 1) = P(X = 1$ and θ is anything between 0 and 1)

$$= \int_0^1 \theta^1 (1-\theta)^{1-1} \, d\theta = \int_0^1 \theta \, d\theta = \frac{1}{2}.$$

(2 May 19)

Bivariate CDFs

Def. The *joint* CDF of two rvs X and Y is the function $F_{X,Y}(x, y)$

satisfying $F_{X,Y}(x, y) = P(X \leq x$ *and* $Y \leq y)$

for all $-\infty < x < \infty$ and $-\infty < y < \infty$

Consequences of this definition

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 $F_X(x) = \lim_{y \to \infty} F_{XY}(x, y)$, and similarly the **marginal** CDF $F_Y(y)$ is just $F_Y(y) = \lim_{x \to \infty} F_{XY}(x, y)$

---

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).

~~Consequences~~ continued

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

(and similarly for $f_Y(y)$).

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 <u>marginal</u> pdf $f_X(x)$ of X is

(marginalizing out Y)

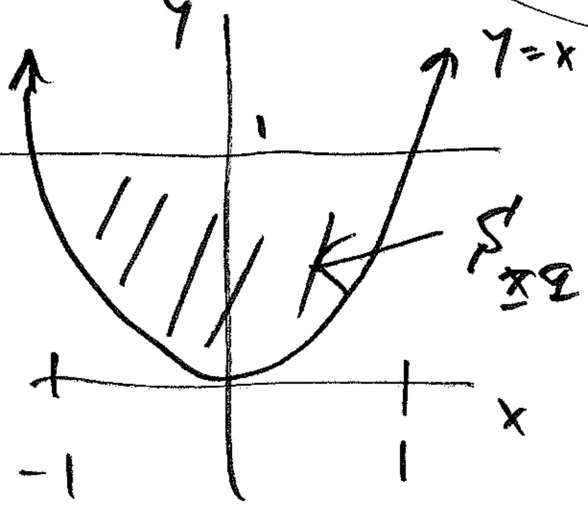
$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy \quad \left(\text{for all } -\infty < x < \infty\right)$$

and the marginal pdf $f_Y(y)$ of Y is $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx \quad \left(\text{for all } -\infty < y < \infty\right)$.

Earlier example, continued

(X, Y) have joint pdf

$$f_{XY}(x, y) = \begin{cases} \frac{21}{4} x^2 y, & \overset{S_{XY}}{0 \le x^2 \le y \le 1} \\ 0 & \text{else} \end{cases}$$



You can see from the sketch of the support S_{XY} of $f_{XY}(x, y)$ that

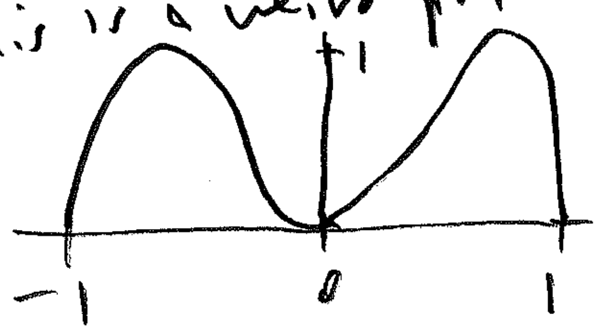
$-1 \le X \le 1$, so the support S_X of X is $(-1, 1)$ and its marginal pdf is

$\boxed{\text{wa}}$ integrate 21 x^2 y / 4 for y from x^2 to 1

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy = \int_{x^2}^{1} \frac{21}{4} x^2 y \, dy$$

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

this is a weird pdf



(supposed to be symmetric) & bimodal

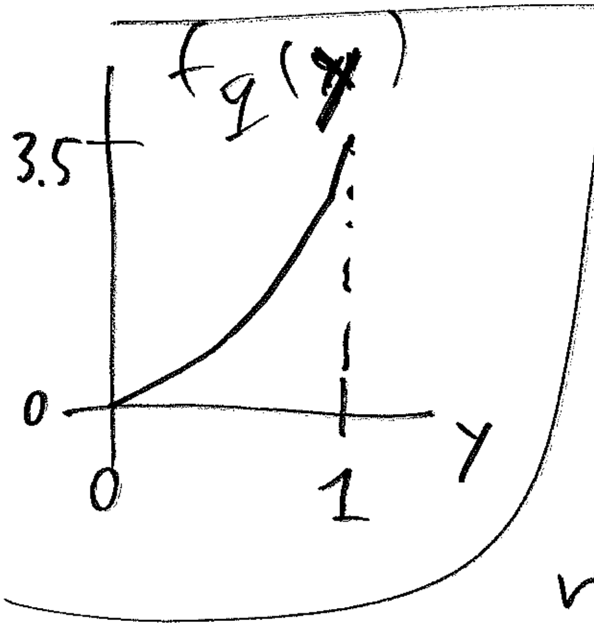
Similarly, the support S_Y of Y is $(0,1)$,

and its marginal pdf is

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

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

(4 Aug 17)



(5) Consequences, continued

If you have the joint dist. $f_{XY}(x,y)$, you can reconstruct the marginals $f_X(x)$ and $f_Y(y)$, but <u>not the other way around</u>: if all you have is the marginals, (in general) they do not uniquely determine the joint.

Example (DS p. 134)

Case 1: X = # heads in n tosses of fair coin 1 and independently Y = # heads in n tosses of fair coin 2

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

$Y = X$

Case 1:

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

so $f_X(x) = \begin{cases} \binom{n}{x}\left(\frac{1}{2}\right)^x\left(1-\frac{1}{2}\right)^{n-x} & x = 0, 1, \ldots, n \\ 0 & \text{else} \end{cases}$

$\binom{n}{x}\left(\frac{1}{2}\right)^n$

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

so $f_Y(y) = \begin{cases} \binom{n}{y}\left(\frac{1}{2}\right)^n & y = 0, 1, \ldots, n \\ 0 & \text{else} \end{cases}$

Since X and Y are independent in Case 1, $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$ (as we'll see in a minute), so in case 1 $f_{XY}(x,y) = \begin{cases} \binom{n}{x}\binom{n}{y}\left(\frac{1}{2}\right)^{2n} & \text{for } x=0,1,\dots,n \text{ and } y=0,1,\dots,n \\ 0 & \text{else} \end{cases}$

note: DS error

However: In case 2,

X is Binomial$(n, \frac{1}{2})$ and so is Y (same as in case 1), but their joint distribution (since $Y = X$) 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.

Def. rvs X and Y are <u>independent</u> if for every (non-weird) sets A and B of real numbers $P(X \in A \text{ and } Y \in B) = P(X \in A) \cdot P(Y \in B)$.

Consequences

① Immediately you get that if X and Y are indep.

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

$$= P(X \le x)\, P(Y \le y)$$

$$= F_X(x) \cdot F_Y(y).$$

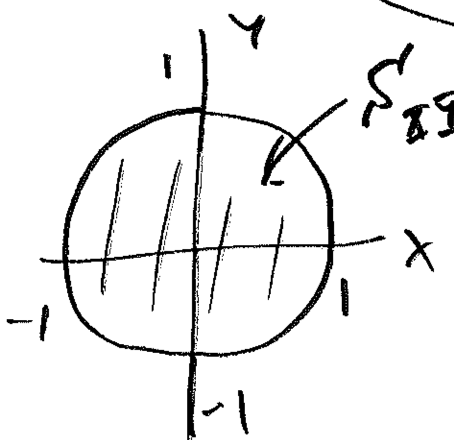
This is an iff: the converse is also true

② Differentiate this equation once with respect to x and once with respect to y

to get the result that

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

Example Suppose that continuous rvs X and Y have joint pdf



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

The support S_{XY} of f_{XY} is the region inside the unit circle. You can evaluate the normalizing constant by computing $\iint_{S_{XY}} k x^2 y^2 \, dx\, dy$ and setting it equal to 1:

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

so $k = \frac{24}{\pi}$

$\boxed{Q:}$ Are X and Y independent?

A: No, they can't be: since the only points (x,y) with positive density satisfy $x^2+y^2 \le 1$, for any given value y of Y, the possible values of X depend on y, & vice versa. (114)

Example:

Continuous rv X and Y have joint pdf

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

Q: Are X and Y independent?

A: Yes, because (a) $e^{-(x+2y)}$ factors into $(e^{-x})(e^{-2y})$ and (b) the support S_{XY} also "factors": $(x \ge 0)$ & $(y \ge 0)$

Just choose (k, k_x, k_y) such that

$$\iint_{S_{XY}} k e^{-(x+2y)} \, dx\,dy = 1, \quad \int_0^\infty k_x e^{-x} \, dx = 1,$$

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

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 with 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

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

$$\parallel$$

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

to be the conditional pmf of Y given X (12.56)

Example: gender & marijuana legalization preference at UCLA

(See doc. cam. notes 14 Aug 17)

& quiz 3

Now let's do the analogous thing for continuous rvs.

$\boxed{\text{Def.}}$ 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 example

X, Y have joint pdf

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

let's work out $f_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$. Earlier we saw that

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

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

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

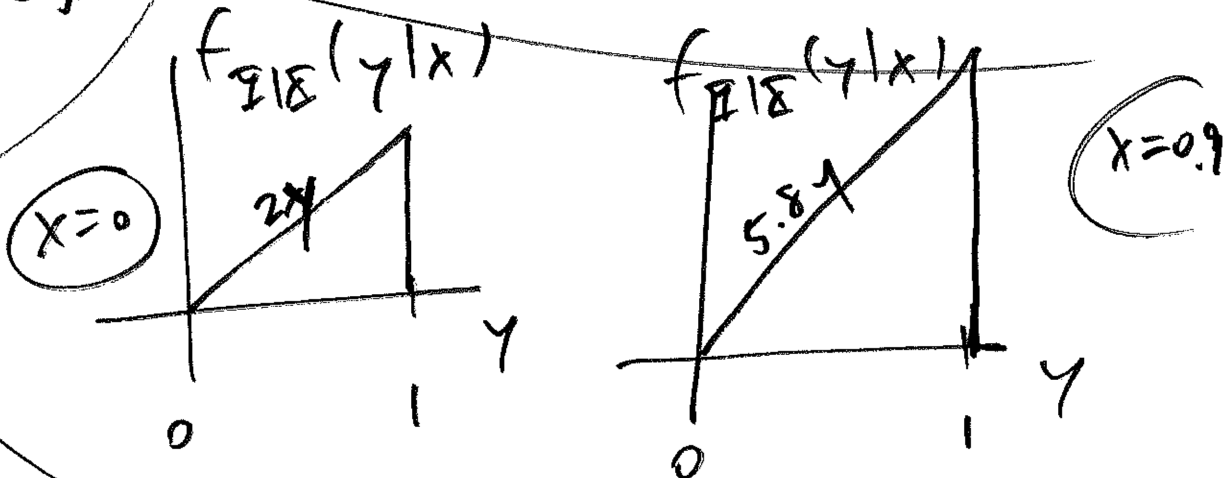
$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

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

and this simplifies to

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

A few "slices" of this:

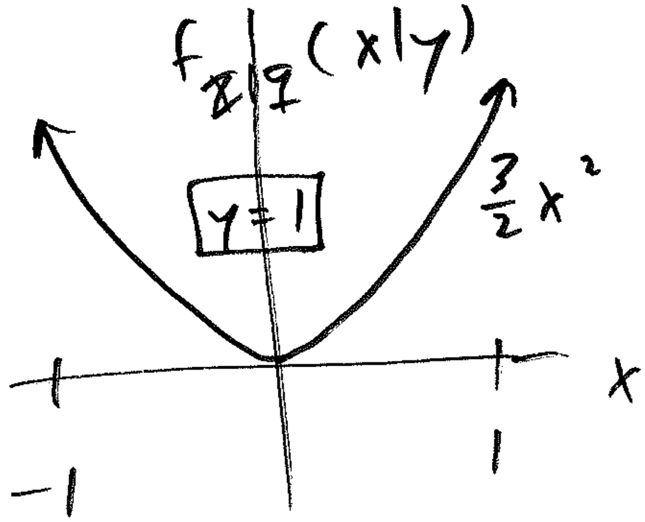


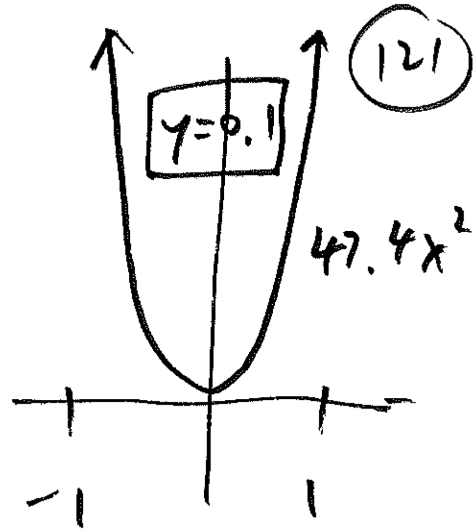
And in the other direction for $0 \le y \le 1$

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

$$= \begin{cases} \dfrac{\frac{21}{4}x^2 y}{\frac{7}{2}y^{5/2}} = \dfrac{3}{2}x^2 y^{-\frac{3}{2}} = \dfrac{3x^2}{2y^{3/2}} & \text{for } 0 \le x^2 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

A few "slices" of this





Note: 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 0. But that's not what's actually going on; strictly speaking $f_{Y|X}(y|x)$ is a limit:

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

In other words, you take a little strip

of x values of width ϵ around $X = x^*$, compute $P(Y \le y \mid X$ is in the strip$)$, differentiate the result with respect to y, and let ϵ 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.

constructing a joint pdf from marginals & conditionals

We know that (as long as no divisions by 0 happen)

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

and

$$f_{X|Y}(x|y) \overset{②}{=} \frac{f_{XY}(x,y)}{f_Y(y)}.$$

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

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

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

Case Study

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 snugly 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 θ = 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, θ is un*known*.

Implicit assumption (*stationarity*): θ is constant over the entire indefinite time period

To learn about θ, we could take a random sample of (nut, bolt) pairs of size m ("with replacement") (say) and count the # of defectives in the sample (call this N).

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

$(D_i \mid \theta) \overset{\text{IID}}{\sim} \text{Bernoulli}(\theta)$ stationarity

$(i = 1, \dots, n)$

$N = \sum_{i=1}^{m} D_i$

so the conditional pf of rv N is

fixed & known

$f_{N|\Theta}(n \mid m, \theta) = (*)$ (sampling dist.)

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

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

A reasonable estimate of θ would be

$\hat{\theta} = \frac{N}{m} = \frac{3}{114} \doteq \underline{2.6\%}$; but how much uncertainty do we have about θ on the basis of this dataset?

Bayesian story

θ unknown continuous $\in (0,1)$

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

probability AMS 131

$p(\text{data} \mid \text{unknown})$ easy

$p(N \mid \theta) = \circledast$

AMS 132/206 statistics

$p(\text{unknown} \mid \text{data})$

(stat. inference) harder

$$\boxed{p(\theta \mid \underset{\sim}{D})} =$$

$$p(\theta \mid \underline{\underline{N}})$$

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

because

Bernoulli: dataset

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

and the rv N

carry the same

info about θ

Bayes's Theorem

$$\underline{\underline{p(\theta \mid \underset{\sim}{D})}} = \frac{p(\theta) \overset{\circledast}{\boxed{p(\underset{\sim}{D} \mid \theta)}}}{p(\underset{\sim}{D})}$$

$\updownarrow$

$$p(\theta \mid N) = \frac{p(\theta) \overset{\circledast}{\boxed{p(N \mid \theta)}}}{p(N)}$$

total info about θ

info about θ external to dataset

(normalizing constant)

info about θ internal to dataset

(7 May 18)