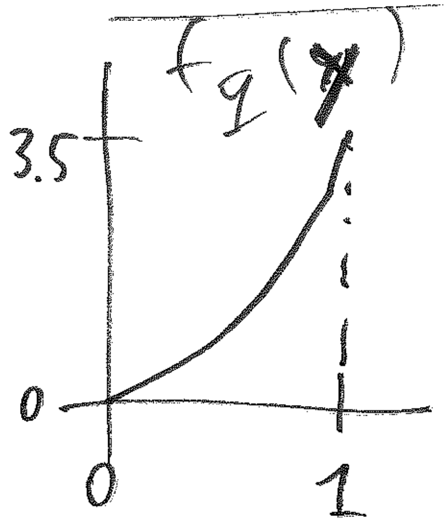
Similarly, the support S_Y of Y is $(0, 1)$,

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)



Consequences, continued

(5) If you have the joint dist. $f_{XY}(x, y)$, you can reconstruct the marginals $f_X(x)$ and $f_Y(y)$, <u>but not the other way around</u>: if all you have is the marginals, 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 $f_X(x) = \begin{cases} \binom{n}{x}\left(\frac{1}{2}\right)^x\left(1-\frac{1}{2}\right)^{n-x} & x=0,1,\dots,n \\ 0 & \text{else} \end{cases}$

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

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

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

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: $\exists$ 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 <u>independent</u>.

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.

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

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

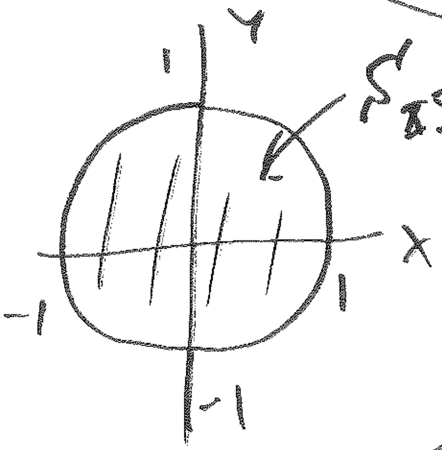
$$= F_X(x) \cdot F_Y(y).$$

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

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

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 $k = \dfrac{24}{\pi}$

Q: Are X and Y independent?

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

Example:

Continuous rv X and Y have joint pdf

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

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

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

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 $k_x = 1$, $k_y = 2$, $k = 2$. ✓

$\boxed{\text{Conditional probability distributions}}$ Recalling that for two events A and B, $P(B \mid A) = \dfrac{P(A \text{ and } B)}{P(A)}$ (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

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

to be

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

the conditional pmf of Y given X (10.56)

Example: gender & marijuana legalization preference at UCLA

(See d.c. com. notes 14 Ans 17) & quiz 3

117

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

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

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

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

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

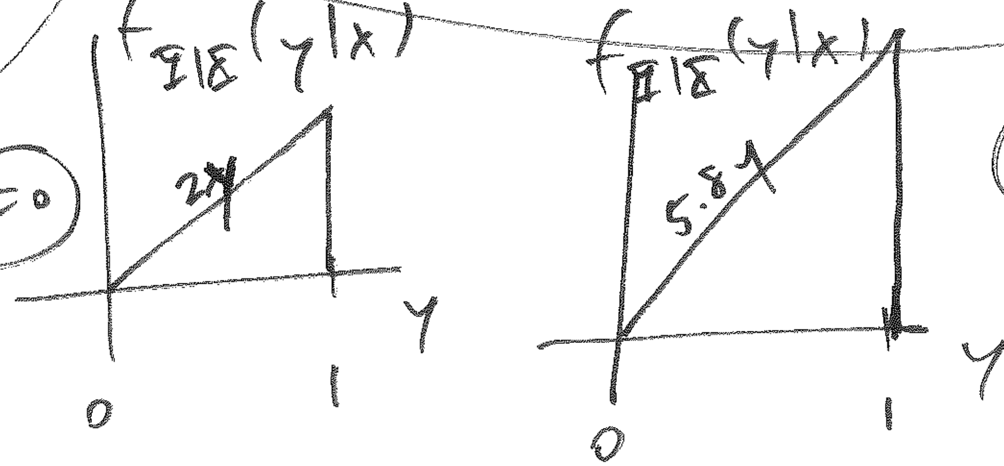
$$= \begin{cases} \dfrac{\frac{21}{4}x^2 y}{\frac{21}{8}x^2(1-x^4)} & \text{for } 0 \le x^2 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

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:

$x=0$ $\qquad$ $x=0.9$

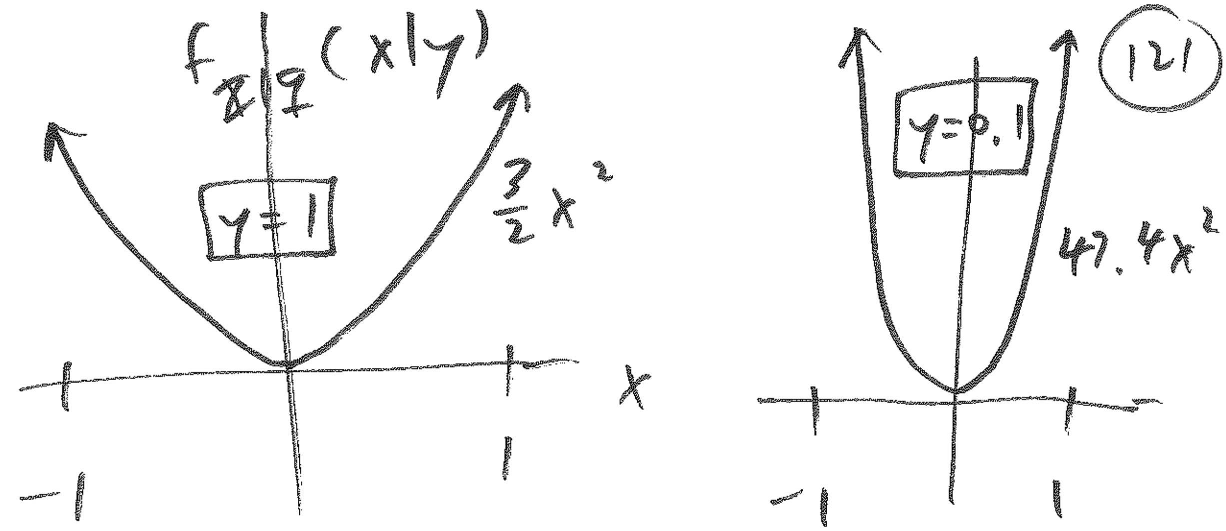


And in the other direction for $0 \le y \le 1$

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

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

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:

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

In other words, you take a little strip

of x values of width ϵ around $X = x^*$, compute $P(Y \le y \mid X \text{ is in the strip})$, 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)

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

and

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

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; (124)

let's call a (nut, bolt) pair <u>defective</u> 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 <u>unknown</u>.

Implicit assumption (<u>stationarity</u>): 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 (say) and count the # of defectives in the sample (call this N). Let $D_i = \begin{cases} 1 & \text{if (nut, bolt) pair } i \text{ is defective} \\ 0 & \text{else} \end{cases}$

$(D_i \mid \theta) \overset{\text{IID}}{\sim} \text{Bernoulli}(\theta)$ $(i = 1, \dots, m)$ — stationarity

$N = \sum_{i=1}^{m} D_i$ so the pf of rv N is

fixed & known

$f_N(n \mid m, \theta) = (*)$ (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

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

probability

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

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

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

Bayes's Theorem

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

$\updownarrow$

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

total info about θ

info about θ external to dataset

(normalizing constant)

info about θ internal to dataset

statistics

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

(stat. inference)

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

because

Bernoulli: dataset

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

and the rv N

carry the same

info about θ

Multivariate distributions

So far we've looked at one and then two rvs at a time; easy to generalize to a finite number of rv $Y_1, \ldots, Y_n$, n a positive finite integer.

Def. The *joint CDF* of n rvs $Y_1, \ldots, Y_n$ is the function $F_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)$ specified by
$$F_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = P(Y_1 \le y_1 \text{ and } \ldots, Y_n \le y_n)$$

More compact to use vector notation: $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$, $\underset{\sim}{y} = (y_1, \ldots, y_n)$

$$F_{\underset{\sim}{Y}}(\underset{\sim}{y}) = P(Y_1 \le y_1, \ldots, Y_n \le y_n)$$

$\underset{\sim}{Y}$ is said to be a *random vector* taking values in $\mathbb{R}^n$.

Def. n rv $(Y_1, \ldots, Y_n) = \underset{\sim}{Y}$ have a discrete joint distribution if the random vector $\underset{\sim}{Y}$ can only take on a finite or countably infinite # of possible values $(y_1, \ldots, y_n) \in \mathbb{R}^n$.

The joint PMF (probability mass function) of $\underset{\sim}{Y}$

is $f_{Y_1 \cdots Y_n}(y_1, \ldots, y_n) = P(Y_1 = y_1, \ldots, Y_n = y_n)$

or equivalently $f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = P(\underset{\sim}{Y} = \underset{\sim}{y})$.

Example n patients in treatment group of a randomized clinical trial; $B_i = \begin{cases} 1 & \text{if patient } i \text{ has a good outcome} \\ 0 & \text{else} \end{cases}$.

If nothing else is known about the patients (e.g., age, disease burden at start of trial, ...) it would be reasonable to model the B_i as IID Bernoulli (θ) same success probability.

$\underset{\sim}{B} = (B_1, \ldots, B_n)$; $\underset{\sim}{b} = (b_1, \ldots, b_n)$; $\underset{\sim}{B}$ has a discrete joint distribution $f_{\underset{\sim}{B}}(\underset{\sim}{b}) = P(B_1 = b_1, \ldots, B_n = b_n)$.

PF

1 or 0

If θ were known you could use $f_{\underset{\sim}{B}}(\underset{\sim}{b})$ to predict the dataset before it arrives: by the IID assumption $P(B_1 = b_1, \ldots, B_n = b_n) =$

$$P(B_1 = b_1) \cdots P(B_n = b_n)$$

Recall that

$$P(B_i = b_i) = \theta^{b_i}(1-\theta)^{1-b_i} \text{ for } b_i = 0, 1$$

so

$$f_{\underset{\sim}{B}}(\underset{\sim}{b}) = \prod_{i=1}^{n} \theta^{b_i}(1-\theta)^{1-b_i} = \theta^{\sum_{i=1}^{n} b_i}(1-\theta)^{n-\sum_{i=1}^{n} b_i}$$

$$= \theta^{s}(1-\theta)^{n-s}$$

with $s = \sum_{i=1}^{n} b_i$.

Def. n rv $\Sigma_1, \ldots, \Sigma_n$ have a continuous joint distribution if you can find a function $f_{\underset{\sim}{\Sigma}}$ on $\mathbb{R}^n$ such that for every (non-weird) subset $G \subset \mathbb{R}^n$

$$P[(\mathcal{I}_1, \ldots, \mathcal{I}_n) \in G] = \int_G \cdots \int f_{\mathcal{I}_1 \cdots \mathcal{I}_n}(y_1, \ldots, y_n)\, dy_1 \cdots dy_n$$

$f_{\underset{\sim}{\mathcal{I}}}(\underset{\sim}{y})$ is the <u>joint PDF</u> (probability density function) of $\underset{\sim}{\mathcal{I}}$.

more compactly

$$P(\underset{\sim}{\mathcal{I}} \in G) = \int_G \cdots \int f_{\underset{\sim}{\mathcal{I}}}(\underset{\sim}{y})\, d\underset{\sim}{y}.$$

Consequence of this def.

① If the joint dist. of $\underset{\sim}{\mathcal{I}}$ is continuous, then

$$f_{\underset{\sim}{\mathcal{I}}}(\underset{\sim}{y}) = \frac{\partial^n}{\partial y_1 \cdots \partial y_n} F_{\underset{\sim}{\mathcal{I}}}(\underset{\sim}{y}).$$

mixed discrete/continuous random vectors behave just as they do with n rv with 2 rv.

more realistically, θ would be unknown, and you can think about the

Example clinical (trial continued)

joint dist. of $(\underset{\sim}{B}, \theta) = (B_1, \dots, B_n, \theta)$, in which the B_i are discrete and $0 < \theta < 1$ is continuous.

Marginal distributions

If you know the joint PDF $f_{\underset{\sim}{Y}}$ of $\underset{\sim}{Y}$, you can work out the marginal distribution of any subset of $(Y_1, \dots, Y_n)$ by integrating $f_{\underset{\sim}{Y}}(\underset{\sim}{y})$ over the elements of $(Y_1, \dots, Y_n)$ that are not in the subset.

Example

$\underset{\sim}{Y} = (Y_1, Y_2, Y_3, Y_4)$

$$f_{Y_1}(y_1) = \iiint f_{\underset{\sim}{Y}}(\underset{\sim}{y}) \, dy_2 \, dy_3 \, dy_4$$

$$f_{Y_2, Y_3}(y_2, y_3) = \iint f_{\underset{\sim}{Y}}(\underset{\sim}{y}) \, dy_1 \, dy_4$$ and so on.

Similarly, you can work out a marginal CDF by sending the other components to ∞: for example

$$F_{Y_1}(y_1) = P(Y_1 \le y_1) = P(Y_1 \le y_1, Y_2 < \infty, \ldots, Y_n < \infty)$$

$$= \lim_{y_2 \to \infty, \ldots, y_n \to \infty} F_{\underset{\sim}{Y}}(\underset{\sim}{y}).$$

Definition

n rvs $Y_1, \ldots, Y_n$ are <u>independent</u> if for any non-weird sets $A_1, \ldots, A_n$ of real numbers

$$P(Y_1 \in A_1, \ldots, Y_n \in A_n) = \prod_{i=1}^{n} P(Y_i \in A_i).$$

Immediate consequences

① $Y_1, \ldots, Y_n$ independent iff

$$F_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \prod_{i=1}^{n} F_{Y_i}(y_i)$$

② $Y_1, \ldots, Y_n$ independent iff $f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \prod_{i=1}^{n} f_{Y_i}(y_i)$.

[illegible]

Def. Starting with a univariate PMF or PDF $f_{Y_i}(y_i)$, n rvs $(Y_1, \dots, Y_n)$ form a random sample of size n from f_{Y_i} if the Y_i are independent and all of them have marginal PMF or PDF f_{Y_i} $\longleftrightarrow$ i.e., if the Y_i are an independent identically distributed (IID) sample from f_{Y_i}.

Example deer at UCSC: some have a disease (chronic wasting disease)

population

all deer living within UCSC boundary on 8 Aug 2017

Sample

the observed deer

1 = Y

0 = N

$$N = ? \ (\approx 800) \quad \begin{array}{c} \text{disease?} \\ \begin{pmatrix} 1\text{s} \\ \& \\ 0\text{s} \end{pmatrix} \end{array} \quad \xrightarrow{\text{IID}} \quad \begin{array}{c} \text{disease?} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ n \end{array} \begin{pmatrix} 1\text{s} \\ \& \\ 0\text{s} \end{pmatrix} \end{array} \ n = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

mean $\theta = ?$ (unknown)

mean $\bar{y} = \hat{\theta}$ ← estimate of θ

"y-bar" "theta-hat"

Shorthand for the diagram:

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

Definition

Start with random vector $\underset{\sim}{X} = (X_1, \dots, X_n)$; partition it into 2 subvectors $\underset{\sim}{X} = (\underset{\sim}{Y}, \underset{\sim}{Z})$, $\underset{\sim}{Y} = (Y_1, \dots, Y_k)$, $1 \le k \le n-1$,

$\underset{\sim}{Z} = (Z_1, \dots, Z_{n-k})$

Then for every point $\underset{\sim}{z}$ for which $f_{\underset{\sim}{Z}}(\underset{\sim}{z}) > 0$, the conditional distribution of $\underset{\sim}{Y}$ given $\underset{\sim}{Z}$ is

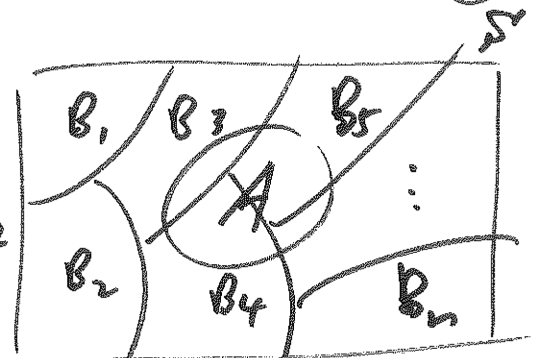
$$f_{\underset{\sim}{Y} \mid \underset{\sim}{Z}}(\underset{\sim}{y} \mid \underset{\sim}{z}) = \frac{f_{\underset{\sim}{Y}\underset{\sim}{Z}}(\underset{\sim}{y}, \underset{\sim}{z})}{f_{\underset{\sim}{Z}}(\underset{\sim}{z})}, \quad \underset{\sim}{y} \in \mathbb{R}^k$$

from which

$$f_{\underset{\sim}{Y}\underset{\sim}{Z}}(\underset{\sim}{y}, \underset{\sim}{z}) = f_{\underset{\sim}{Z}}(\underset{\sim}{z}) \, f_{\underset{\sim}{Y} \mid \underset{\sim}{Z}}(\underset{\sim}{y} \mid \underset{\sim}{z}).$$

Multivariate law of total probability

You'll recall that if A is an event & you're



trying to compute $P(A)$ & it's hard, one idea is to find another aspect of the world B upon which A depends, such that the events $B_1, \dots, B_n$ form a <u>partition</u>; then

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(B_i)\, P(A \mid B_i).$$

This has an analogue with continuous r.v.s:

using the notation in the definition of conditional distributions

$$\underbrace{f_{\underset{\sim}{Y}}(\underset{\sim}{y})}_{\text{like } A} = \int \cdots \int_{\mathbb{R}^{n-k}} \underbrace{f_{\underset{\sim}{Z}}(\underset{\sim}{z})}_{\text{like } B_i} \overbrace{f_{\underset{\sim}{Y} \mid \underset{\sim}{Z}}(\underset{\sim}{y} \mid \underset{\sim}{z})}^{\text{like } P(A \mid B_i)} \, d\underset{\sim}{z}.$$

Multivariate Bayes's theorem

Using the same notation,

$$\underbrace{f_{\underset{\sim}{Z}|\underset{\sim}{Y}}(\underset{\sim}{z}|\underset{\sim}{y})}_{\text{(posterior info)}} = \underbrace{f_{\underset{\sim}{Z}}(\underset{\sim}{z})}_{\text{(prior info)}} \, \underbrace{f_{\underset{\sim}{Y}|\underset{\sim}{Z}}(\underset{\sim}{y}|\underset{\sim}{z})}_{\text{(likelihood info)}} \Big/ f_{\underset{\sim}{Y}}(\underset{\sim}{y}) \quad \text{(normalizing constant)}.$$

The usual application of this in statistics is as follows.

Def. $\underset{\sim}{Z}$ a random vector with multivariate distribution $f_{\underset{\sim}{Z}}(\underset{\sim}{z})$; then random variables $X_1, \dots, X_n$ are <u>conditionally independent given $\underset{\sim}{Z}$</u> if for all $\underset{\sim}{z}$ with $f_{\underset{\sim}{Z}}(\underset{\sim}{z}) > 0$,

$$f_{\underset{\sim}{X}|\underset{\sim}{Z}}(\underset{\sim}{x} \mid \underset{\sim}{z}) = \prod_{i=1}^{n} f_{X_i|\underset{\sim}{Z}}(x_i \mid \underset{\sim}{z}).$$

Earlier example, revisited

Remember the machine with a θ dial that can make IID coin tossing with $P(\text{heads}) = \theta$? 137

We agreed earlier that, if θ is unknown to you, (1) the results of the coin tossing $Y_1, Y_2, \ldots$ are <u>dependent</u>, because there is useful information in any subset of them for predicting any other subset, but (2) the Y_i become <u>conditionally independent given θ</u>, because once you know θ there's no longer any useful information in the Y_i to predict other Y_i.

This is why — in both the clinical trial example & the (nuts & bolts) example — we model the data values Y_i as

$$(Y_i \mid \theta) \overset{\text{conditionally IID}}{\sim} \text{Bernoulli}(\theta).$$

Functions of a rv

Case 1: discrete

(univariate) X discrete rv with PMF $f_X(x)$; $Y = h(X)$ for some function h defined on {possible values of X}. Then

$$f_Y(y) = P(Y = y) = P[h(X) = y]$$

$$= \sum_{\{x : h(x) = y\}} f_X(x)$$

Example

$X \sim \text{Uniform}\{1, 2, \dots, 9\}$

The median of this distribution is 5;

$Y = |X - 5| = h(X)$ keeps track of how far X is from the median.

(139)

y	X such that $Y = y$	$P(Y = y)$
0	5	1/9
1	4 or 6	2/9
2	3 or 7	2/9
3	2 or 8	2/9
4	1 or 9	2/9
		1

Case 2: continuous

X continuous rv with PDF $f_X(x)$; $Y = h(X)$ as before.

The CDF $F_Y(y)$ can be worked out as follows:

$$F_Y(y) = P(Y \le y) = P[h(X) \le y]$$

$$= \int_{\{x : h(x) \le y\}} f_X(x)\,dx$$

and if Y is also continuous

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

(at every point y where F_Y is differentiable).

Example) X = rate at which customers served in a queue at the bank (140)

Natural to model X as continuous, with CDF F_X.
(also, $X > 0$)

Turns out that the average waiting time is $Y = \frac{1}{X} = h(X)$. You can get the PDF of Y

in 2 steps: ① work out CDF of Y
② differentiate with respect to y

① (for $y > 0$)

$$F_Y(y) = P(Y \le y) = P[h(X) \le y]$$

$$= P\left(\frac{1}{X} \le y\right) = P\left(X \ge \frac{1}{y}\right)$$

since X is continuous

$$= 1 - P\left(X < \frac{1}{y}\right) = 1 - P\left(X \le \frac{1}{y}\right)$$

$$= 1 - F_X\left(\frac{1}{y}\right) \quad \text{and now}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}\left(1 - F_X\left(\frac{1}{y}\right)\right)$$

chain rule

$$= -f_X\left(\frac{1}{y}\right)\left(-y^{-2}\right) = \frac{f_X\left(\frac{1}{y}\right)}{y^2}$$

Example $X \sim \text{Uniform}[-1, +1]$ (14 Aug 17) (continuous)

$Y = X^2$ find PDF of Y

First note that Y's possible values are $[0, 1]$.

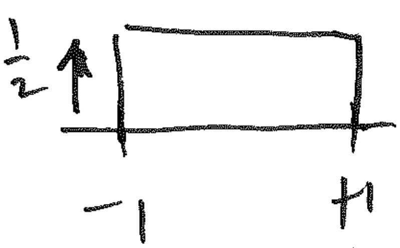
~~for $0 < y < 1$~~

① $F_Y(y) = P(Y \le y) = P(X^2 \le y)$

$$= P(-\sqrt{y} \le X \le \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)\, dx$$

$$= \frac{1}{2} x \Big|_{-\sqrt{y}}^{\sqrt{y}}$$

$$= \sqrt{y}$$



$$f_X(x) = \begin{cases} \frac{1}{2} & -1 \le x \le 1 \\ 0 & \text{else} \end{cases}$$

② Thus

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$