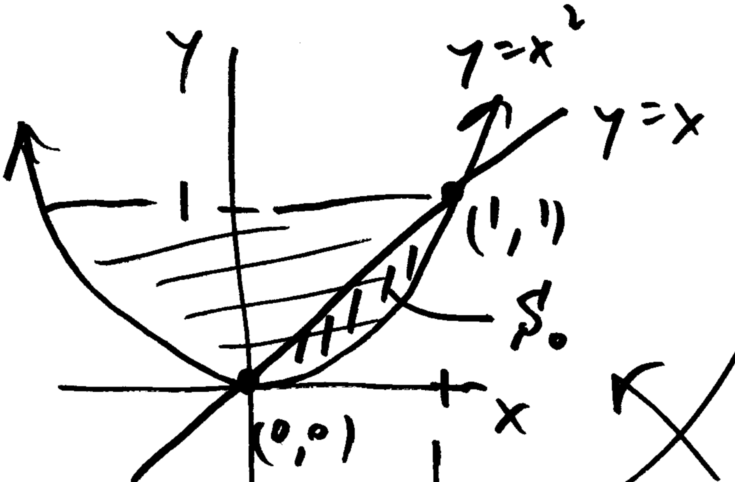
$$= \frac{c}{3}\int_0^1 2y^{5/2}\,dy = \frac{2c}{3}\left(\frac{y^{7/2}}{7/2}\Big|_0^1\right)$$ 101

$= \frac{4}{21}c$ as before ($\iint dx\,dy$ and $\iint dy\,dx$ always have to agree, of course).

Example, continued | Let's compute $P(X \geq Y)$



The relevant part S_0 of S where $x \geq y$ is sketched here, so

Wed

integrate $21x^2y/4\,dy$, $y = x^2 .. x$

then integrate result $x =$ from $0..1$

$$P(X \geq Y) = \iint_{S_0} f_{XY}(x,y)\,dy\,dx$$

$$= \int_0^1\int_{x^2}^{x} \frac{21}{4}x^2 y\,dy\,dx \underset{(\ldots)}{=} \frac{3}{20}.$$

You can have bivariate distributions in which one of (X, Y) is discrete and the other is continuous.

Definition: $\boxed{\text{Case 3}}$ mixed bivariate distribution

(X, Y) rv such that X is discrete and Y is continuous → suppose you can find a function $f_{XY}(x, y)$ defined on $\mathbb{R}^2$ such that for every pair of (non-weird) subsets A and B of $\mathbb{R}$ (assume integral exists)

$$P(X \in A \text{ and } Y \in B) = \int_B \sum_{x \in A} f_{XY}(x, y)\, dy.$$

Then f_{XY} is the joint pmf/pdf of (X, Y)

Immediate consequence: If X takes on values $x_1, x_2, \dots$, then

$$\int_{-\infty}^{\infty} \sum_{i=1}^{\infty} f_{XY}(x_i, y)\, dy = 1.$$

Example | Randomized controlled (clinical) trial; patients in $\textcircled{T}$ get a treatment, patients in $\textcircled{C}$ get a placebo. Outcome is success (e.g., cancer goes into remission) or failure; let $X_i = \begin{cases} 1 & \text{if patient } i \text{ in } \textcircled{T} \text{ is a success} \\ 0 & \text{else} \end{cases}$

and let θ ← (unknown) be the proportion of patients in the population of all patients who might get the treatment who would have no relapse if they had been in the study. Then our uncertainty about θ is continuous on $(0, 1)$ and (X_i, θ) has a mixed bivariate distribution.

If you model $(X \mid \theta)$ as Bernoulli(θ) 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 $F_{X,Y}(x,y) = P(X \le x$ and $Y \le y)$

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 <u>marginal</u> CDF $F_X(x)$ from the joint CDF as $F_X(x) = \lim_{y \to \infty} F_{XY}(x, y)$, and similarly the <u>marginal</u> 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~~ [illegible] 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

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

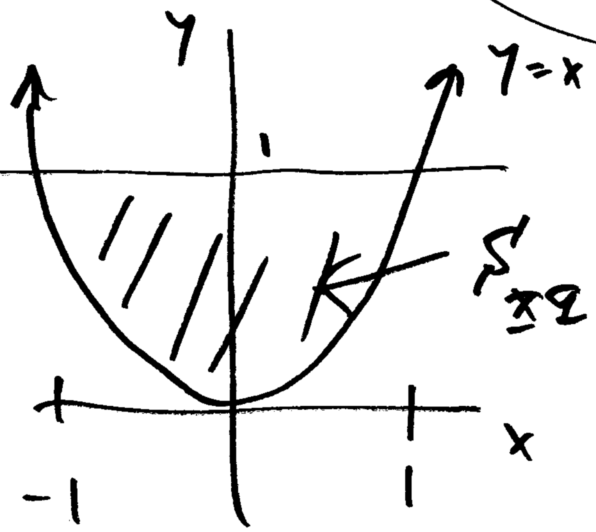
and the marginal pdf $f_Y(y)$ of Y is

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

Earlier example, continued

(X, Y) have joint pdf

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

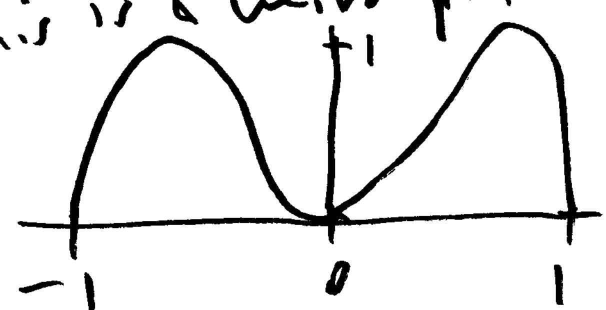


You can see from the sketch of the support S_{XY} of $f_{XY}(x,y)$ that $-1 \le X \le 1$, so the support S_X of X is $(-1, 1)$ and its marginal pdf is

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

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

This is a weird pdf



(supposed to be symmetric) & bimodal

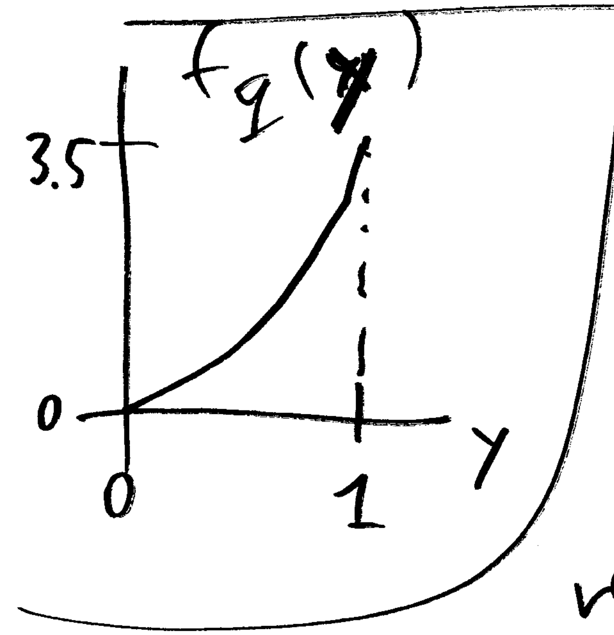
$$= \begin{cases} \frac{21}{8} x^2 (1 - x^4), & -1 < x < 1 \\ 0 & \text{else} \end{cases}$$

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 $f_Y(y)$, but not the other way around: 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:

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

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

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

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

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

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 $\quad f_{XY}(x,y) = \begin{cases} \binom{n}{x}\binom{n}{y}\left(\frac{1}{2}\right)^{2n} & \text{for } x = 0,1,\dots,n \text{ and } y = 0,1,\dots,n \\ 0 & \text{else} \end{cases}$ (11)

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 <u>independent</u>.

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

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

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

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

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

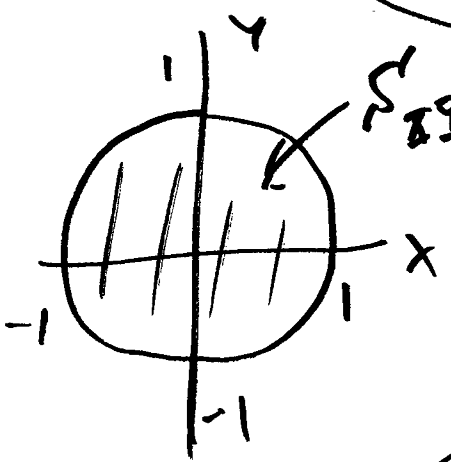
This is an iff: the converse is also true

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 $x^2+y^2 \leq 1$, 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

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

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": $(x \geq 0)$ & $(y \geq 0)$

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 $k_X = 1$, $k_Y = 2$, $k = 2$ ✓.

Conditional probability distributions

Recalling that for two events A and B, $P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$ (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

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

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

the conditional pmf of Y given X (12.56)

Example: gender & marijuana legalization preference at UCLA

(See doc. cam. notes [redacted] 14 Aug [redacted] 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

$$f_{Y|X}(y|x) = \left\{ \frac{f_{XY}(x,y)}{f_X(x)} \right\} \text{ to be}$$

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 $-1 < x < 1$)

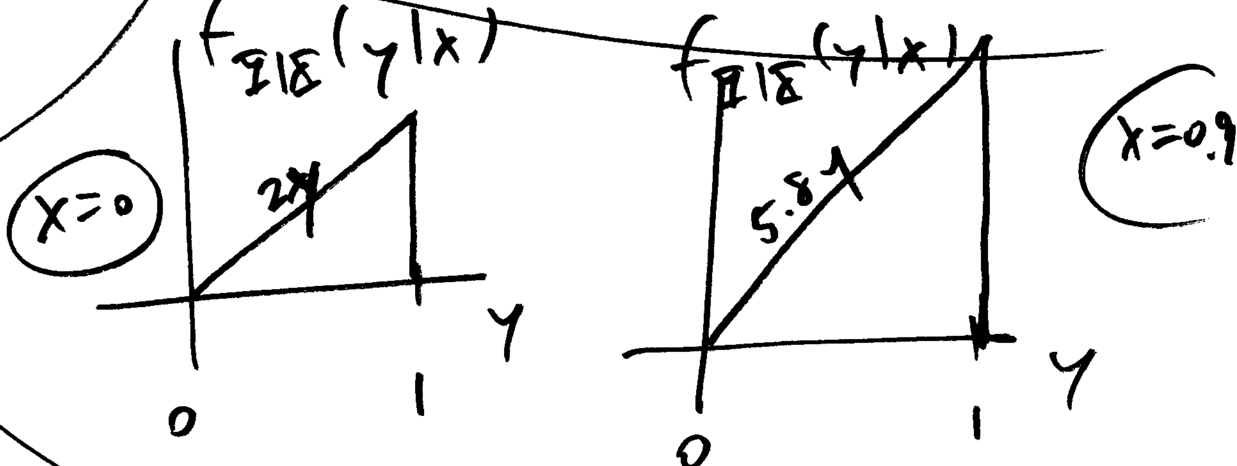
$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

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

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:



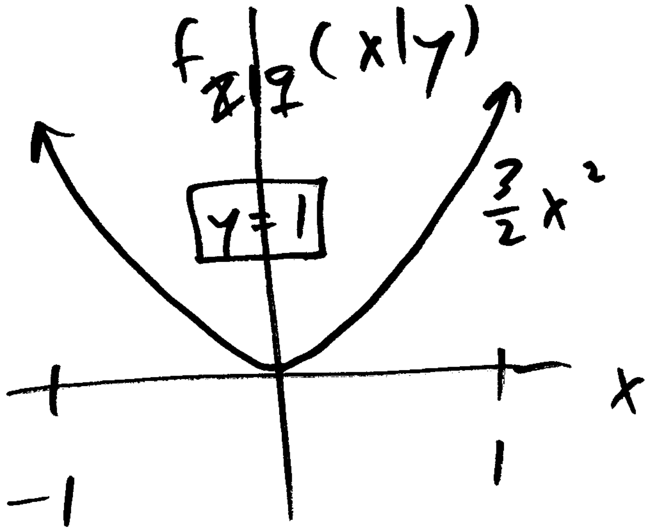
And in the other direction

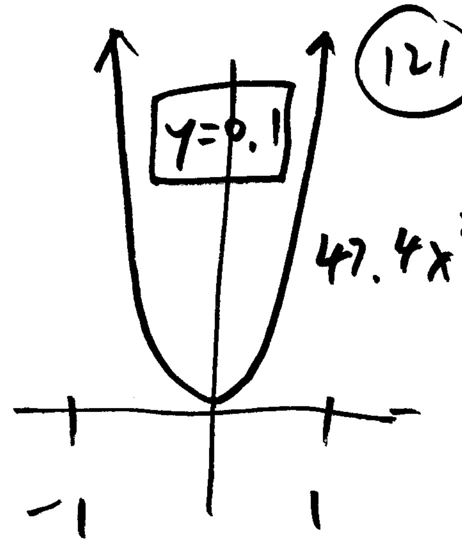
for $0 \le y \le 1$

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

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

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 $X = x^*$, (122)

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)

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

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; (24)
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) ("with replacement") pairs of size m (say) and count the # of defectives in the sample (call this N).

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

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

stationarity

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

so the conditional pf of rv N is

m fixed & known

$f_{N|\Theta}(n \mid m, \theta) = \circledast$ (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

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

Bayesian story

θ unknown $\in (0,1)$ continuous

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

probability AMS 131

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

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

AMS 132/206 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

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

and the rv N

carry the same

info about θ

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

Bayes's Theorem

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

$\updownarrow$

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

total info about θ

info about θ external to dataset

(normalizing constant)

info about θ internal to dataset

(7 May 18)

Multivariate distributions

So far we've looked at 127
one and then two rvs at
a time; easy to generalize to a finite
number of rv $X_1, \dots, X_n$, n a positive finite integer.

Def. The <u>joint CDF</u> of n rvs
$X_1, \dots, X_n$ is the function $F_{X_1, \dots, X_n}(y_1, \dots, y_n)$
specified by $F_{X_1, \dots, X_n}(y_1, \dots, y_n) = P(X_1 \le y_1$, and $\dots, X_n \le y_n)$

More compact to use vector
notation: $\underset{\sim}{X} = (X_1, \dots, X_n)$, $\underset{\sim}{y} = (y_1, \dots, y_n)$

$F_{\underset{\sim}{X}}(\underset{\sim}{y}) = P(X_1 \le y_1, \dots, X_n \le y_n)$ $\underset{\sim}{X}$ is
said to be a <u>random vector</u> 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 $f_{\underset{\sim}{B}}(\underset{\sim}{b}) = P(B_1 = b_1, \ldots, B_n = b_n)$.

1 or 0

PF

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

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

Recall that

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

so

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

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

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

Def. n rv $Z_1, \ldots, Z_n$ have a continuous joint distribution if you can find a function $f_{\underset{\sim}{Z}}$ on $\mathbb{R}^n$ such that for every (non-weird) subset $G \subset \mathbb{R}^n$

130

$$P[(X_1, \dots, X_n) \in G] = \int \cdots \int_G f_{X_1 \cdots X_n}(y_1, \dots, y_n)\, dy_1 \cdots dy_n$$

$f_{\underset{\sim}{X}}(\underset{\sim}{y})$ is the joint PDF (probability density function) of $\underset{\sim}{X}$.

more compactly

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

Consequences of this def.

① If the joint dist. of $\underset{\sim}{X}$ is continuous, then $f_{\underset{\sim}{X}}(\underset{\sim}{y}) = \frac{\partial^n}{\partial y_1 \cdots \partial y_n} F_{\underset{\sim}{X}}(\underset{\sim}{y})$.

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

Example clinical trial (continued)

more realistically, θ would be unknown, and you can think about the joint dist. of $(\underset{\sim}{B}, \theta) = (B_1, \dots, B_n, \theta)$, (131)

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

$$F_{X_1}(y_1) = P(X_1 \le y_1) = P(X_1 \le y_1, X_2 < \infty, \dots, X_n < \infty)$$

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

Definition

n rvs $X_1, \dots, X_n$ are <u>independent</u> if for any non-weird sets $A_1, \dots, A_n$ of real numbers

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

Immediate consequences

① $X_1, \dots, X_n$ independent iff

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

② $X_1, \dots, X_n$ independent iff $f_{\underset{\sim}{X}}(\underset{\sim}{y}) = \prod_{i=1}^{n} f_{X_i}(y_i)$.

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 ***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 9 May 2019

Sample: the observed deer

1 = Y
0 = N

disease?

$N = ?$ (≈ 800)

$$\begin{bmatrix} 1s \\ \& \\ 0s \end{bmatrix}$$

mean $\theta = ?$ (unknown)

IID

disease?

$$\begin{array}{c|c} 1 & \\ 2 & 1s \\ \vdots & \& \\ & 0s \\ n & \end{array} \qquad n = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

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

"y-bar"   "theta-hat"

Shorthand for the diagram:

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

(end example for now)

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 <u>conditional distribution</u> of $\underset{\sim}{Y}$ given $\underset{\sim}{Z}$ is

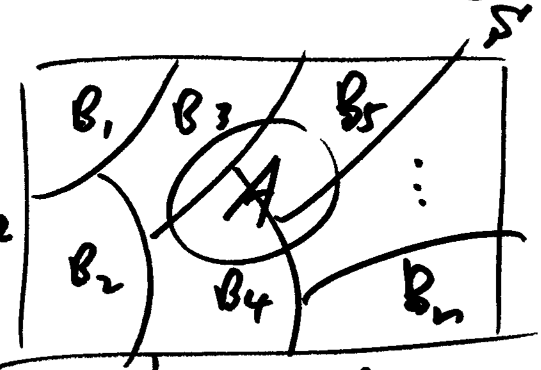
$$f_{\underset{\sim}{Y} \mid \underset{\sim}{Z}}(\underset{\sim}{y} \mid \underset{\sim}{z}) = \frac{f_{\underset{\sim}{Y}\underset{\sim}{Z}}(\underset{\sim}{y}, \underset{\sim}{z})}{f_{\underset{\sim}{Z}}(\underset{\sim}{z})}, \quad \underset{\sim}{y} \in \mathbb{R}^k$$

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 partition;

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

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

Multivariate Bayes's theorem

Using the same notation,

$$\underset{\text{(posterior info)}}{f_{\underset{\sim}{Z}|\underset{\sim}{Y}}(\underset{\sim}{z}|\underset{\sim}{y})} = \frac{\underset{\text{(prior info)}}{f_{\underset{\sim}{Z}}(\underset{\sim}{z})}\;\underset{\text{(likelihood info)}}{f_{\underset{\sim}{Y}|\underset{\sim}{Z}}(\underset{\sim}{y}|\underset{\sim}{z})}}{\underset{\text{(normalizing constant)}}{f_{\underset{\sim}{Y}}(\underset{\sim}{y})}}.$$

unknown data

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 $f_{\underset{\sim}{Z}}(\underset{\sim}{z}) > 0$,

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

Earlier example, revisited

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

We agreed earlier that, if θ is unknown to you, (1) the results of the coin tosses $Y_1, Y_2, \ldots$ 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

(univariate)

Case 1: discrete

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

Discrete

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

The median of this distribution is 5;

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

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

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

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 (14 Aug 17)

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

$Y = X^2$ find PDF of Y

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

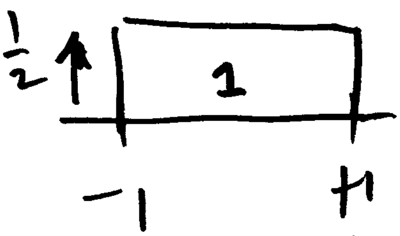
~~for $0 < y < 1$~~

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

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

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

$$= \sqrt{y}$$

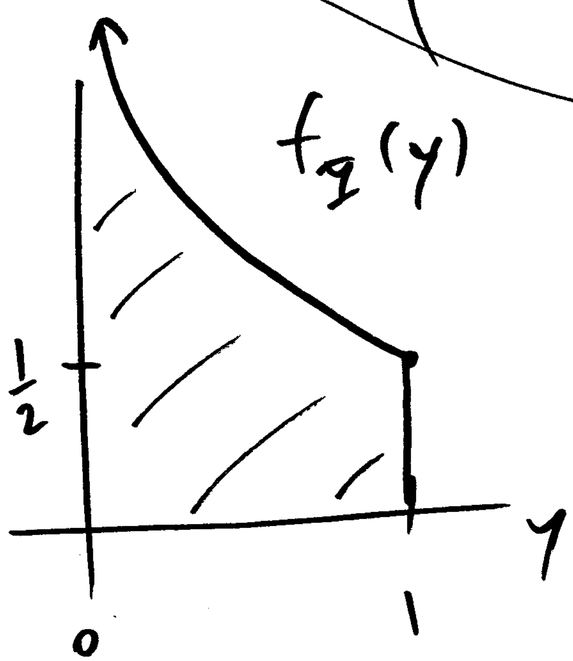


$$f_X(x) = \begin{cases} \frac{1}{2} & -1 \le x \le 1 \\ 0 & \text{else} \end{cases}$$

② Thus

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

so $f_Y(y) = \begin{cases} \frac{d}{dy} y^{\frac{1}{2}} = \frac{1}{2\sqrt{y}} & \text{for } 0 < y < 1 \\ 0 & \text{else} \end{cases}$



This density is unbounded at 0 (!).

Easy theorem

X continuous rv with pdf $f_X(x)$,

$Y = aX + b \quad (a \neq 0)$ $\boxed{\text{linear transformation}}$

$$\to f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Interesting and useful fact

X continuous with CDF $F_X(x)$; what's the distribution of $Y = F_X(X)$?

$F_Y(y) = P(Y \le y) = P[F_X(X) \le y)$ (143)

for $0<y<1$ $= P[X \le F_X^{-1}(y)] = F_X[F_X^{-1}(y)] = y$

But the dist. with $F_Y(y) = y$ for $0<y<1$

is the Uniform$(0,1)$ distribution (!)

probability Integral Transform

X continuous with CDF F_X, $Y = F_X(X)$

$\to$ $Y \sim$ Uniform$(0,1)$ or $[\ \ ]$

why is this useful?

Converse is also true:

$Y \sim$ Uniform$[0,1]$, F_X a

continuous CDF with quantile function

F_X^{-1} $\to$ $X = F_X^{-1}(Y) \sim F_X$

This is the practical basis for the generation of many forms of pseudo-random numbers:

It turns out to be easy to generate pseudo-Uniform$(0,1)$ values; therefore if you want to generate pseudo-random #s from a distribution with CDF F_X and F_X^{-1} is easy & fast to compute,

$\boxed{\text{Algorithm}}$ (Quiz 6)

$$u_1, \dots, u_n \stackrel{IID}{\sim} \text{Uniform}(0,1)$$

$$\rightarrow F_X^{-1}(u_1), \dots, F_X^{-1}(u_n) \stackrel{IID}{\sim} F_X$$

~~Earlier~~ Example ~~revisited~~

If $X \sim$ Exponential(λ), $(\lambda > 0)$, its PDF is $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{else} \end{cases}$

(p. 91)

Earlier we saw that $F_X(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$

and $F_X^{-1}(p) = -\dfrac{\log(1-p)}{\lambda}$

$(0 < p < 1)$

(R demo)

Now you can see immediately

that if $U \sim$ Uniform$(0,1)$ so is $(1-U)$,

so to generate IID Exponential (λ) rv you

just compute $-\dfrac{1}{\lambda} \log U_i$, $U_i \overset{IID}{\sim}$ Uniform$(0,1)$

(rexp) R

why do people want/need pseudo-random numbers?

Some stochastic (probabilistic) models of real-world phenomena are too complicated to fully characterize mathematically in closed form; one highly useful method in such situations is to conduct a (computer-based) simulation study driven by pseudo-random numbers.

Bedrock method in data science today.

The method used above for working out the distribution of $Y = \frac{1}{X}$ can be generalized, as follows. Some functions $h(X)$ are <u>nice</u>, in that they are both <u>differentiable</u> and <u>one-to-one (invertible)</u>

Calculus reminder

If $h(x)$ real-valued is differentiable and one-to-one (1-1) for x in the open interval (a, b), then h is either <u>monotonically increasing</u> or <u>decreasing</u>, and h is also continuous, so it transforms the interval (a, b) to another open interval $h[(a, b)] = (\alpha, \beta)$ called the <u>image</u> of (a, b) under h. Since h is invertible, it makes sense to talk about $y = h(x) \leftrightarrow x = h^{-1}(y)$.

Theorem: $\underline{X}$ continuous rv with PDF $f_{\underline{X}}(x)$

and for which $P(\overset{\text{could}}{a} < \underline{X} < \overset{\text{be infinite}}{b}) = 1$; $\underline{Y} = h(\underline{X})$,

with h differentiable and 1-1 for $a < x < b$;

(α, β) image of (a, b) under h; $h^{-1}(y)$ inverse

function of $h(x)$ for $\alpha < y < \beta$ $\longrightarrow$ PDF

of $\underline{Y}$ is

$$f_{\underline{Y}}(y) = \begin{cases} f_{\underline{X}}[h^{-1}(y)] \underbrace{\left| \frac{d h^{-1}(y)}{dy} \right|}_{\text{(chain rule)}} & \text{for } \alpha < y < \beta \\ 0 & \text{else} \end{cases}$$

Easy short-hand way to render this: "Multiply" both sides

by $|dy|$ to get $f_{\underline{Y}}(y)\,|dy| = f_{\underline{X}}(x)\,|dx|$

$y = h(x)$
$x = h^{-1}(y)$

Earlier example, revisited

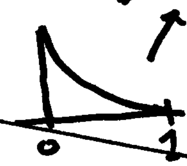
$\underline{Y} = h(\underline{X}) = \frac{1}{\underline{X}}$: average waiting time in the bank queue

Here $y = h(x) = \frac{1}{x}$ so $x = h^{-1}(y) = \frac{1}{y}$

and $\frac{d}{dy} \frac{1}{y} = -\frac{1}{y^2}$; thus $f_Y(y) = \dfrac{f_X\left(\frac{1}{y}\right)}{y^2}$ as before

Example: At time 0, population of V organisms introduced into large tank of water with nutrients; X = rate of growth. Under one model that's realistic in some circumstances, at time t the predicted population size would be $Y = V e^{Xt}$ (exponential growth).

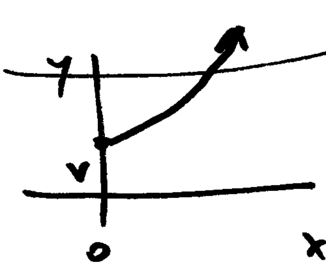
X unknown, modeled with $f_X(x) = \begin{cases} 3(1-x)^2 & 0 < x < 1 \\ 0 & \text{else} \end{cases}$.



$y = h(x) = V e^{xt}$

$x = 0 \to y = V$

$x = 1 \to y = V e^{t}$ (image



strictly increasing

$$\frac{y}{v} = e^{xt} \rightarrow \log\left(\frac{y}{v}\right) = xt \rightarrow x = h^{-1}(y) = \frac{1}{t}\log\left(\frac{y}{v}\right)$$

$$\frac{d}{dy}\,\frac{1}{t}\log\left(\frac{y}{v}\right) = \frac{1}{t}\left(\frac{y}{v}\right)^{-1}\cdot\frac{1}{v} = \frac{1}{ty}$$ thus

$$f_Y(y) = \begin{cases} \dfrac{3\left[1 - \frac{1}{t}\log\left(\frac{y}{v}\right)\right]^2}{ty} & v < y < ve^{t} \\ 0 & \text{else} \end{cases}$$

(9 May 19)

Functions of 2 or more rvs

Case 1: discrete

n rvs $X_1, \ldots, X_n$ discrete joint dist. with joint PMF $f_{\underset{\sim}{X}}(\underset{\sim}{x})$;

define
$$\begin{cases} Y_1 = h_1(X_1, \ldots, X_n) \\ \vdots \qquad\qquad \vdots \\ Y_m = h_m(X_1, \ldots, X_n) \end{cases} \quad (m \geq 1)$$

real-valued $\quad (h_j : \mathbb{R}^n \rightarrow \mathbb{R})$

Given values $\underset{\sim}{y} = (y_1, \dots, y_m)$ of $(Y_1, \dots, Y_m) = \underset{\sim}{Y}$

let A be the set of points $(x_1, \dots, x_n)$

such that $\left\{ \begin{array}{ccc} y_1 & = & h_1(x_1, \dots, x_n) \\ \vdots & & \vdots \\ y_m & = & h_m(x_1, \dots, x_n) \end{array} \right\}$. Then

the joint PMF $f_{\underset{\sim}{Y}}(\underset{\sim}{y})$ is given by

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \sum_{(x_1, \dots, x_n) \in A} f_{\underset{\sim}{X}}(\underset{\sim}{x}).$$

Case 2: continuous, $(m=1)$ | n rvs $X_1, \dots, X_n$, continuous, joint dist with joint PDF $f_{\underset{\sim}{X}}(\underset{\sim}{x})$,

$Y = h(\underset{\sim}{X})$ univariate (real) | For each real y define $A_y = \{ \underset{\sim}{x} : h(\underset{\sim}{x}) = y \}$

Then PDF of Y is $f_Y(y) = \int \cdots \int_{A_y} f_{\underset{\sim}{X}}(\underset{\sim}{x}) \, d\underset{\sim}{x}$.