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

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

$(h_j : \mathbb{R}^n \to \mathbb{R})$

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

$$\left\{\begin{array}{l} y_1 = h_1(x_1,\dots,x_n) \\ \vdots \\ y_m = h_m(x_1,\dots,x_n) \end{array}\right\}.$$

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

151

Simple example of this result

(X_1, X_2) joint continuous PDF

$f_{X_1 X_2}(x_1, x_2)$, $Y = a_1 X_1 + a_2 X_2 + b$

with $a_1 \neq 0$ $\longrightarrow$ Y continuous

with PDF $f_Y(y) = \displaystyle\int_{-\infty}^{\infty} f_{X_1 X_2}\left(\frac{y - b - a_2 x_2}{a_1}, x_2 \right) \frac{dx_2}{|a_1|}$

Important Special case

The simplest thing you can do with two or more rvs is to add them.

This is also important in statistics, where the sample mean $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ plays a key role.

In the result above, take $(a_1, a_2, b) = (1, 1, 0)$ to get

$Y = X_1 + X_2$.

Dist. of Y is called the convolution of the dists. of X_1 and X_2

By the above result

$$f_{Y}(y) = \int_{-\infty}^{\infty} f_{X_1}(y-z)\, f_{X_2}(z)\, dz$$

$$= \int_{-\infty}^{\infty} f_{X_1}(z)\, f_{X_2}(y-z)\, dz .$$

A more complicated example

$X_i \overset{\text{IID}}{\sim}$ CDF $F_{X_\bullet}$, <u>PDF</u> $f_{X_\bullet}$

$(i = 1, \ldots, n)$ (continuous)

is defined to be

↓

$$Y_{(1)} \triangleq \min(X_1, \ldots, X_n)$$

$$Y_{(n)} \triangleq \max(X_1, \ldots, X_n)$$

these are examples of the <u>order</u> statistics of $(X_1, \ldots, X_n)$

~~Take-Home test 2~~ problem 4

$$F_{Y_{(n)}}(t) = P(Y_{(n)} \le t)$$

$\Updownarrow$ iff (check)

$$= P(X_1 \le t, X_2 \le t, \ldots, X_n \le t)$$

$$\overset{\text{I ID}}{=} P(X_1 \le t) \cdots P(X_n \le t)$$

$$\overset{\text{I ID}}{=} \left[F_{X_\bullet}(t)\right]^n$$

so $Y_{(n)}$ has PDF $f_{Y_{(n)}}(t) = \frac{d}{dt}\left[F_{X_\bullet}(t)\right]^n$

$$= n\left[F_{X_\bullet}(t)\right]^{n-1} f_{X_\bullet}(t)$$

Similarly

$$F_{Y_{(1)}}(t) = P(Y_{(1)} \le t) = 1 - P(Y_{(1)} > t)$$

$\Updownarrow$ iff (check)

$$= 1 - P(X_1 > t, \ldots, X_n > t)$$

$$\overset{\text{(I)ID}}{=} 1 - P(X_1 > t) \cdots P(X_n > t)$$

$$\overset{\text{I(ID)}}{=} 1 - \left[1 - F_{X_\bullet}(t)\right]^n$$

So $Y_{(1)}$ has PDF $f_{Y_{(1)}}(t) = \frac{d}{dt} F_{Y_{(1)}}(t)$

$$= n\left[1 - F_{X_\bullet}(t)\right]^{n-1} f_{X_\bullet}(t)$$

Generalizing the earlier differentiable & 1-1 result

Multivariate transformations

$X_1, \ldots, X_n$ continuous joint dist with joint PDF $f_{\underset{\sim}{X}}(\underset{\sim}{x})$

Suppose that there is a subset S of $\mathbb{R}^n$ with $P[(X_1, \ldots, X_n) \in S] = 1$.

support of $(X_1, \ldots, X_n)$ under $f_{\underset{\sim}{X}}$

Define new rvs:

$$Y_1 = h_1(X_1, \ldots, X_n)$$
$$\vdots \qquad \vdots$$
$$Y_n = h_n(X_1, \ldots, X_n)$$

(note same # of Xs)

Assume that the n functions $h_1, \ldots, h_n$ define a 1-1 differentiable transformation of S onto some subset T of $\mathbb{R}^n$.

image of $h_1, \ldots, h_n$

Inverse transform:

$$x_1 = h_1^{-1}(y_1, \ldots, y_n)$$
$$\vdots \qquad \vdots$$
$$x_n = h_n^{-1}(y_1, \ldots, y_n)$$

Then the joint PDF $f_{\underset{\sim}{Y}}(y)$ is

$$f_{\underset{\sim}{Y}}(y) = \begin{cases} f_{\underset{\sim}{X}}\left[h_1^{-1}(\underset{\sim}{y}), \dots, h_n^{-1}(\underset{\sim}{y})\right] |J| & \text{for } (y_1, \dots, y_n) \in T \\ 0 & \text{else} \end{cases}$$

in which

J is the determinant of the matrix

$$\begin{bmatrix} \frac{\partial h_1^{-1}}{\partial y_1} & \dots & \frac{\partial h_1^{-1}}{\partial y_n} \\ \vdots & & \\ \frac{\partial h_n^{-1}}{\partial y_1} & \dots & \frac{\partial h_n^{-1}}{\partial y_n} \end{bmatrix}$$

and $|\cdot|$ is absolute value

(chain rule generalization)

J is called the Jacobian of the transformation from $\underset{\sim}{X}$ to $\underset{\sim}{Y}$.

named after the German mathematician Carl Gustav Jacob Jacobi (1804 - 1851)

(died of smallpox at age 46)

J acts like a generalization of the derivative of the inverse in the earlier result.

Example (X_1, X_2) joint (continuous) PDF $f_{X_1,X_2}(x_1,x_2) = \begin{cases} 4x_1x_2 & \text{for } 0<x_1<1,\ 0<x_2<1 \\ 0 & \text{else} \end{cases}$

(check: $\int_0^1\int_0^1 4x_1x_2\,dx_1\,dx_2$

$= \int_0^1 4x_2\left(\int_0^1 x_1\,dx_1\right)dx_2 = 4\int_0^1 x_2\left(\frac{x_1^2}{2}\Big|_0^1\right)dx_2$

$= 2\int_0^1 x_2\,dx_2 = 2\,\frac{x_2^2}{2}\Big|_0^1 = 1$)

Let's work out the joint PDF of

$(Y_1, Y_2) \triangleq \left(\frac{X_1}{X_2}, X_1\cdot X_2\right)$, $y_1 = h_1(x_1,x_2) = \frac{x_1}{x_2}$

$y_2 = h_2(x_1,x_2) = x_1x_2$

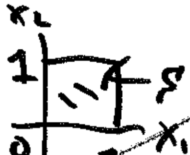
Inverse transform:

157

solve $\left\{ \begin{array}{l} \dfrac{x_1}{x_2} = y_1 \\ x_1 x_2 = y_2 \end{array} \right\}$ for (x_1, x_2): $x_1 = h_1^{-1}(y_1, y_2) = \sqrt{y_1 y_2}$

$x_2 = h_2^{-1}(y_1, y_2) = \sqrt{\dfrac{y_2}{y_1}}$

image: how does $(0 < x_1 < 1,\ 0 < x_2 < 1)$ transform?



(*)

(*) defines 4 inequalities:

$\left\{ \begin{array}{l} x_1 > 0,\ x_1 < 1, \\ x_2 > 0,\ x_2 < 1 \end{array} \right\}$

so (a) $\sqrt{y_1 y_2} > 0$, (b) $\sqrt{y_1 y_2} < 1$

(c) $\sqrt{\dfrac{y_2}{y_1}} > 0$, (d) $\sqrt{\dfrac{y_2}{y_1}} < 1$

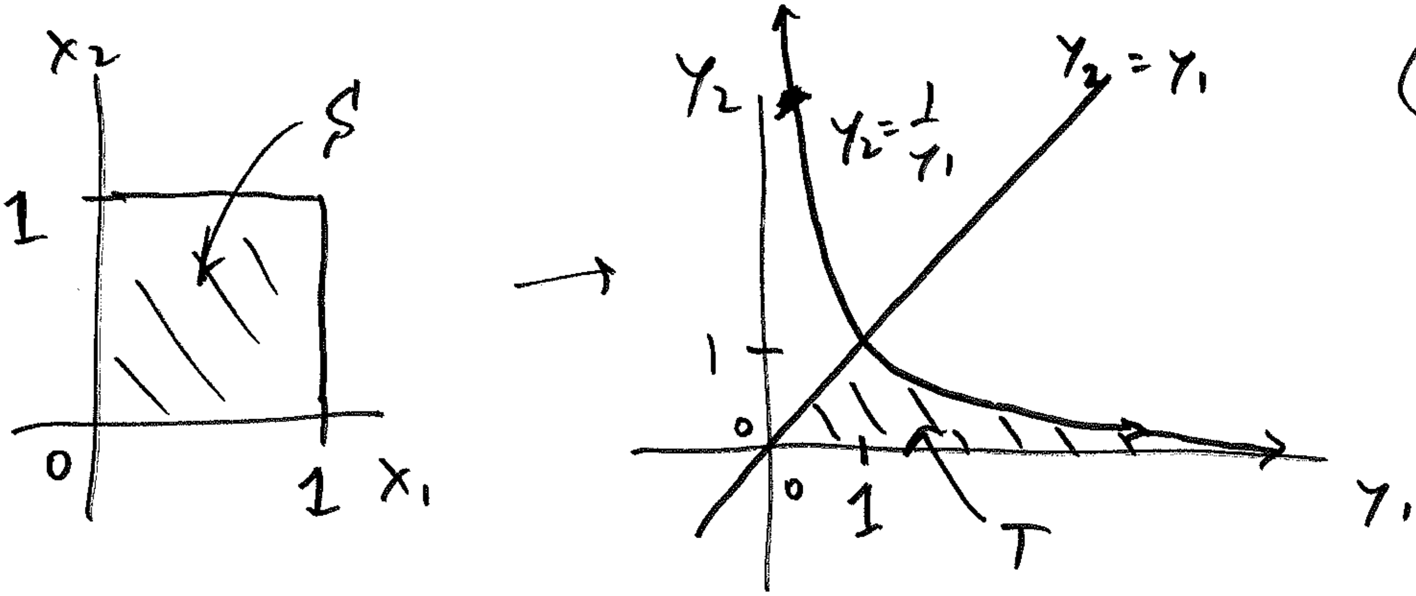
(a) equivalent to $\begin{pmatrix} y_1 > 0 \\ y_2 > 0 \end{pmatrix}$ or $\begin{pmatrix} y_1 < 0 \\ y_2 < 0 \end{pmatrix}$

but $y_1 = \dfrac{x_1}{x_2} > 0$ so it must be $\begin{pmatrix} y_1 > 0 \\ y_2 > 0 \end{pmatrix}$

(c) leads to the same thing

(b) says $y_2 < \dfrac{1}{y_1}$

(d) says $y_2 < y_1$



$$h_1^{-1}(y_1, y_2) = \sqrt{y_1 y_2}$$

$$h_2^{-1}(y_1, y_2) = \sqrt{\frac{y_2}{y_1}}$$

so $\dfrac{\partial}{\partial y_1} h_1^{-1} = \dfrac{1}{2}\sqrt{\dfrac{y_2}{y_1}}$

$$\frac{\partial}{\partial y_2} h_1^{-1} = \frac{1}{2}\sqrt{\frac{y_1}{y_2}}$$

$$\frac{\partial}{\partial y_1} h_2^{-1} = -\frac{1}{2}\left(\frac{y_2}{y_1^3}\right)^{\frac{1}{2}} \qquad \frac{\partial}{\partial y_2} h_2^{-1} = \frac{1}{2}\sqrt{\frac{1}{y_1 y_2}}$$

so $J = \det \begin{bmatrix} \frac{1}{2}\left(\frac{y_2}{y_1}\right)^{1/2} & \frac{1}{2}\left(\frac{y_1}{y_2}\right)^{1/2} \\ -\frac{1}{2}\left(\frac{y_2}{y_1^3}\right)^{1/2} & \frac{1}{2}\left(\frac{1}{y_1 y_2}\right)^{1/2} \end{bmatrix} = \dfrac{1}{2y_1}$

recall
$\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$

and (since $y_1 > 0$) $|J| = \dfrac{1}{2y_1}$

To finish the calculation, in the PDF of $\underset{\sim}{X}$, $f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \begin{cases} 4x_1 x_2 & \begin{pmatrix} 0 < x_1 < 1 \\ 0 < x_2 < 1 \end{pmatrix} \\ 0 & \text{else} \end{cases}$,

substitute $x_1 = \sqrt{y_1 y_2}$, $x_2 = \sqrt{\frac{y_2}{y_1}}$ and bring in the Jacobian:

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = f_{\underset{\sim}{X}}\left[h_1^{-1}(\underset{\sim}{y}), h_2^{-1}(\underset{\sim}{y}) \right] |J|$$

$$= 4 \left(\sqrt{y_1 y_2} \right) \left(\sqrt{\frac{y_2}{y_1}} \right) \frac{1}{2 y_1}$$

$$= \begin{cases} 2 \dfrac{y_2}{y_1} & \text{for } (y_1, y_2) \in T \\ 0 & \text{else} \end{cases}.$$

A useful trick

Start with (X_1, X_2) joint dist.; suppose you're interested only in the dist. of $Y_1 = h_1(X_1, X_2)$ (univariate). Then one way to compute this dist. is with the following 3 steps.

Step 1: Find another r.v. $Y_2 = h_2(X_1, X_2)$ such that the transformation $(X_1, X_2) \to (Y_1, Y_2)$ is 1-1 with a differentiable inverse transformation & the calculations are straightforward.

Step 2: Work out the joint dist. of (Y_1, Y_2).

Step 3: Integrate Y_2 out of the joint dist. (i.e., marginalize over Y_2) to get the marginal dist. of Y_1.

Example of a Y_2 that wouldn't work

$$Y_1 = 2X_1$$

$$Y_2 = 3X_1 = \frac{3}{2}Y_1$$

Here Y_2 is linearly dependent on Y_1, so the rank of the (2×2) Jacobian matrix is only 1 and its determinant is therefore 0.

~~Earlier~~ Example, continued

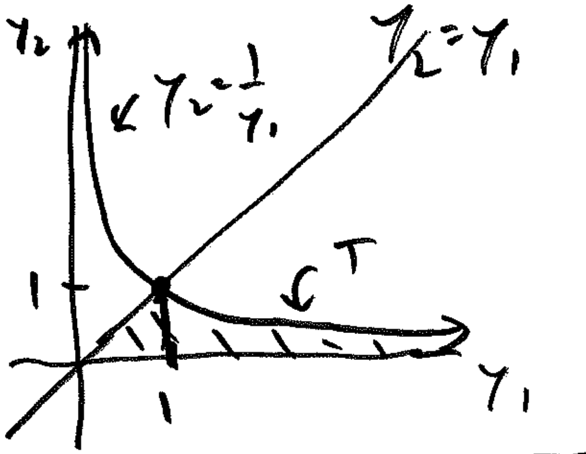
(X_1, X_2) have joint (continuous) PDF
$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 4x_1x_2 & 0<x_1<1,\ 0<x_2<1 \\ 0 & \text{else} \end{cases}$$

Earlier we found that with $(Y_1, Y_2) = \left(\frac{X_1}{X_2}, X_1 \cdot X_2\right)$ the transformed PDF was

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} 2\frac{y_2}{y_1} & \text{for } (y_1,y_2)\in T \\ 0 & \text{else} \end{cases}$$

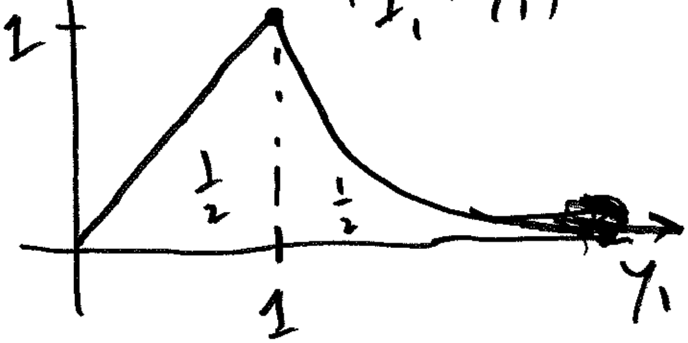
where $T = \left\{(y_1,y_2) : y_1>0,\ y_2 < \min\left(y_1, \frac{1}{y_1}\right)\right\}$.

Suppose you were only really interested in the (marginal) dist. of $Y_1 = \frac{X_1}{X_2}$; then all you have to do is integrate Y_2 out of the joint dist.



For $y_1 > 0$, the allowable region for y_2 is in two parts: for $0 < y_1 < 1$, $0 < y_2 < y_1$ and for $y_1 > 1$, $0 < y_2 < \frac{1}{y_1}$

So $$f_{Y_1}(y_1) = \begin{cases} \int_0^{y_1} 2\left(\frac{y_2}{y_1}\right) dy_2 = y_1 & \text{for } 0 < y_1 < 1 \\ \int_0^{1/y_1} 2\left(\frac{y_2}{y_1}\right) dy_2 = y_1^{-3} & \text{for } y_1 > 1 \end{cases}$$



weird PDF: not differentiable at $y_1 = 1$

useful consequence of Jacobian story

$\underset{1 \sim n}{X} = (X_1, \dots, X_n)$ continuous with joint PDF $f_{X_1 \cdots X_n}(x_1, \dots, x_n)$;

$\underset{1 \sim n}{Y} = (Y_1, \dots, Y_n)$ is a linear transformation of $\underset{\sim}{X}$: $\underset{n \sim 1}{Y^T} = \underset{n\,n}{A} \cdot \underset{n\,1}{X^T}$ transpose

where $\underset{n\,n}{A}$ is an invertible (full-rank) matrix.

Then the PDF of $\underset{\sim}{Y}$ is

$$f_{\underset{\sim}{Y}}(\underset{n\,1}{y^T}) = \frac{f_{\underset{\sim}{X}}(\underset{n\,n}{A^{-1}} \underset{n\,1}{y^T})}{|\det A|}$$

Example

$Y_1 = X_1 + X_2$

$Y_2 = X_1 - X_2$

$$\underset{2\,2}{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$\det A = -2$ $= ad - bc$

$|\det A| = 2$

(recall that)

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \to A^{-1} = \frac{\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}}{ad - bc}$$

$$A^{-1} = \frac{1}{-2} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} = \frac{1}{2} A$$

Expectation, Variance, Covariance, Correlation

DS ch. 4

Example: Tay-Sachs (T-S) disease (continued)

Earlier we worked out the discrete dist. of the rv Y = (# of T-S babies in family of 5, both parents carriers)

We showed that $(Y) \sim$ Binomial (n, p) with $n = 5$, $p = \frac{1}{4}$

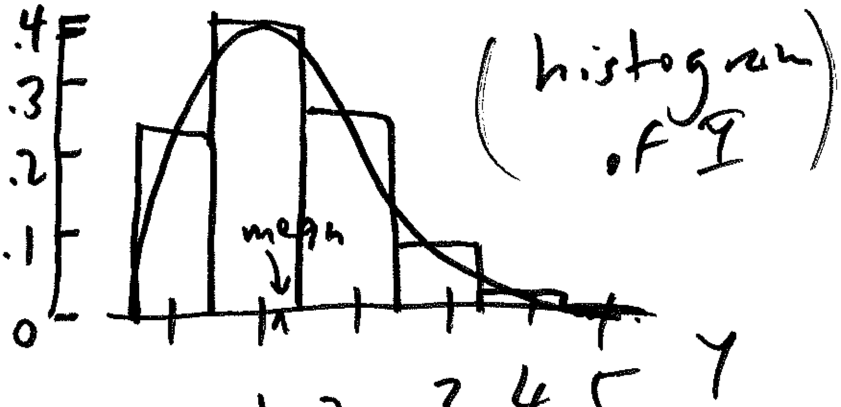
y	$P(Y = y)$
0	$\binom{5}{0}\left(\frac{1}{4}\right)^0 \frac{3}{4}^5 = 0.2373$
1	$\binom{5}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^4 = 0.3955$
2	$\binom{5}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3 = 0.2637$
3	$\binom{5}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2 = 0.0879$
4	$\binom{5}{4}\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^1 = 0.0146$
5	$\binom{5}{5}\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right)^0 = 0.0010$
	1.0000

$$P(Y = y) = \begin{cases} \binom{n}{y} p^y (1-p)^{n-y} & y = 0, 1, \ldots, n \\ 0 & \text{else} \end{cases}$$

$\boxed{Q:}$ About how many T-S babies should these parents <u>expect</u> to have?

(<u>center</u> of dist. of Y?)

$\boxed{A_1}$ Most likely outcome is 1 T-S baby (mode of the dist. of Y)

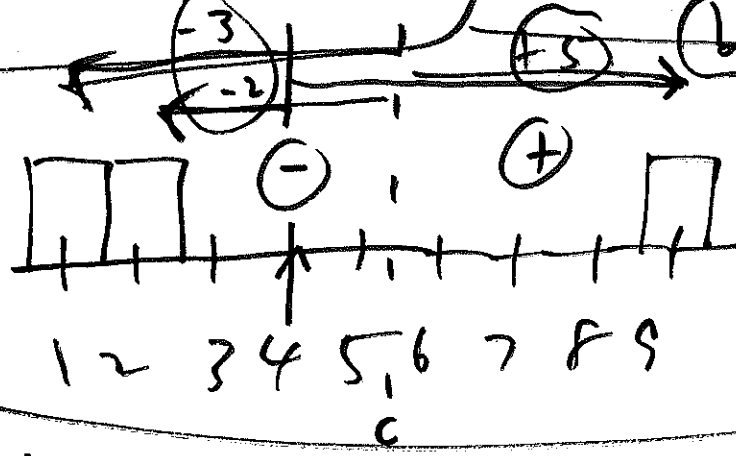


$\boxed{A_2}$ (physics idea)

Let's work out the center of mass of the distribution

balance point

toy example



$$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \quad \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

outlier

Let's find the place c where the histogram balances: where (the sum of forces exerted by the histogram bars to the left of c) equals (the sum of forces to the right):

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \rightarrow \begin{bmatrix} y_1 - c \\ \vdots \\ y_n - c \end{bmatrix}$$

want sum = 0

$$\sum_{i=1}^{n} (y_i - c) = 0 =$$

$$\left(\sum_{i=1}^{n} y_i \right) - n c = 0$$

A_3 median of the dist. of Y (here that's also 1)

here $\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$ mean $\bar{y} = 4$

$$\sum_{i=1}^{n} y_i - nc = 0 \leftrightarrow$$

$$c = \frac{1}{n} \sum_{i=1}^{n} y_i \triangleq \bar{y} = \text{the sample mean of the (sample) data set } \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Here each value of Y occurred only once:

$$\bar{y} = \sum_{i=1}^{n} \left(\frac{1}{n}\right) y_i$$

If some values are more probable than others, the generalization of $\left(\frac{1}{n}\right)$ weight on each y value would be to weight each y by its probability $P(Y = y)$.

Def. A rv is __bounded__ if all of its possible values are finite.

Def. Let Y be a bounded discrete rv with PMF $f_Y(y) = P(Y = y)$. The __mean__ or __expected value__ or __expectation__ of Y, is $E(Y) \triangleq \sum_{\text{all } y} y \, P(Y = y) = \sum_{\text{all } y} y \, f_Y(y)$.

T-S example	$E(Y) = (0)(.2373) + (1)(.3955) + \dots + (5)(.0010) = 1.2500000$

↑ suspiciously round #

Symbolically if $Y \sim \text{Binomial}(n, p)$

then $E(Y) = \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y}$

$\boxed{\text{wa}}$

this assumes that $n > 1$; proof for $n = 1$ is on the next page

$$= \sum_{y=1}^{n} y \binom{n}{y} p^y (1-p)^{n-y}$$

(since summand is 0 for $y = 0$)

$\boxed{\text{Wolfram alpha}}$

$$= \sum_{y=1}^{n} y \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}$$

cancel y against $y \cdot (y-1)!$

$$= \sum_{y=1}^{n} \cancel{y} \frac{n \cdot (n-1)!}{\cancel{y}(y-1)! \, [n-1-(y-1)]!} p \cdot p^{y-1} (1-p)^{n-y}$$

$$= np \sum_{y=1}^{n} \underbrace{\frac{(n-1)!}{(y-1)!(n-y)!}}_{\binom{n-1}{y-1}} p^{y-1} (1-p)^{n-1-(y-1)}$$

$$= np \sum_{y=1}^{n} \binom{n-1}{y-1} p^{y-1} (1-p)^{n-1-(y-1)}$$

$$= np \boxed{\sum_{i=0}^{n-1} \binom{n-1}{i} p^{i} (1-p)^{n-1-i}} \quad \begin{pmatrix} \text{substitute} \\ i = y-1 \end{pmatrix}$$

Binomial$(n-1, p)$ dist.

this $= 1$ because binomial probabilities add up to $\underline{1}$

So: if

$Y \sim$ Binomial(n, p)

for $n > 1$, $E(Y) = np$

when $n = 1$, Binomial$(1, p) =$ Bernoulli(p).

In this case $E(Y) = 0 \cdot P(Y=0) + 1 \cdot P(Y=1)$

$= 0 \cdot (1-p) + 1 \cdot p = p$

$= np$ with $n = 1$

So: for all $n \geq 1$ (integer) and $0 < p < 1$, $Y \sim$ Binomial$(n, p) \rightarrow E(Y) = n \cdot p$.

T-S example $(n=5,\ p=\frac{1}{4})$ $E(Y) = \frac{5}{4} = 1.25$

If discrete X is unbounded, the expectation of X may not exist, either because

$$\sum_{x<0} x\, f_X(x) = -\infty \left(\text{and/or } \sum_{x \geq 0} x\, f_X(x) = +\infty \right)$$

or the distribution "puts too much mass near $\pm\infty$"

Def. X discrete rv with PMF $f_X(x)$; consider $\sum_{x<0} x\, f_X(x)$ and $\sum_{x \geq 0} x\, f_X(x)$. If both sums are infinite, $E(X)$ is undefined (or does not exist); if at least one sum is finite, then

$$E(X) = \sum_{\text{all } x} x\, f_X(x) \text{ exists } \left(\begin{array}{c} \text{it} \\ \text{may} \\ \text{still} \\ \text{be} \\ \text{infinite} \end{array} \right)$$

To create a discrete rv whose mean doesn't exist, you have to play a careful game, because $\sum_{\text{all } x} f_X(x)$ has to be finite (it has to equal 1) but $\sum_{\text{some } x} x\, f_X(x)$ has to be infinite.

$\boxed{\text{Example}}$ The *harmonic series* $\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \dots\right) = \sum_{\substack{x=1 \\ \text{(integers)}}}^{\infty} \frac{1}{x}$ was known to the ancient Greeks, because the wavelengths of the overtones of a vibrating string are $\frac{1}{2}, \frac{1}{3}, \dots$ of the fundamental wavelength of the string. The fact that $\sum_{x=1}^{\infty} \frac{1}{x} = +\infty$ (i.e., the *harmonic series diverges*) was first shown in the 1300s (!) by the French philosopher Nicole Oresme (~1320 – 1382).

It's clear from this divergence that you can't create a rv X with PMF

$$p(X=x) = \frac{c}{x}, \quad x = 1, 2, \dots,$$ because the probability would sum to $+\infty$.

But $p(X=x) = \frac{c}{x^2}$ or $p(X=x) = \frac{c}{x(x+1)}$ turns out to work; for example, $\sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6}$ (!) and, even more conveniently, $\sum_{x=1}^{\infty} \frac{1}{x(x+1)} = 1$.

DS use this to construct two pathological discrete distributions, to show what can go wrong with the idea of expectation.

Example 1

$$f_X(x) = \begin{cases} \frac{1}{x(x+1)} & x = 1, 2, \dots \\ 0 & \text{else} \end{cases}.$$

$$E(X) = \sum_{x=1}^{\infty} x \cdot \frac{1}{x(x+1)} = \sum_{x=1}^{\infty} \frac{1}{x+1} = +\infty$$

so $E(X)$ exists, it's just infinite.

Example 2

$$f_X(x) = \begin{cases} \dfrac{1}{2|x|(|x|+1)} & x = \pm 1, \pm 2, \ldots \\ 0 & \text{else} \end{cases}$$

we already know that $\sum_{\text{all } x} f_X(x) = 1$, so X is a well-defined rv; but $\sum_{x=-1}^{-\infty} x \cdot \frac{1}{2|x|(|x|+1)} = -\infty$

and $\sum_{x=1}^{\infty} x \frac{1}{2x(x+1)} = +\infty$, so $E(X)$ does not exist.

we won't work with pathological rv, mostly.

Expectation for continuous rvs

Def. X bounded continuous rv

with PDF $f_X(x) \rightarrow E(X) \triangleq \int_{-\infty}^{\infty} x f_X(x)\,dx$.

Example $X \sim \text{Exponential}(\lambda)$ $(\lambda > 0)$:

recall that $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{else} \end{cases}$

So $E(X) = \int_{0}^{\infty} \lambda x e^{-\lambda x}\,dx \overset{\text{integrate by parts}}{=} \frac{1}{\lambda}$.

For this reason, many people <u>parameterize</u> the exponential distribution differently:

Alternative definition $X \sim \text{Exponential}(\overset{\text{eta}}{\eta})$ $(\eta > 0)$

$\rightarrow f_X(x) = \begin{cases} \frac{1}{\eta} e^{-\frac{x}{\eta}} & x > 0 \\ 0 & \text{else} \end{cases}$

with this parameterization you can see that $E(X) = \eta$ (easier to interpret).

Nevertheless, to avoid confusion with DS, I'll stick with $\lambda e^{-\lambda x}$. If continuous rv Y is unbounded, a bit of care is once again required to define $E(Y)$. Def.

Y continuous rv with PDF $f_Y(y)$; consider

$$\int_{-\infty}^{0} y\, f_Y(y)\, dy \quad \text{and} \quad \int_{0}^{\infty} y\, f_Y(y)\, dy.$$

If both integrals are infinite, $E(Y)$ is undefined (or does not exist); if at least one of these integrals is finite, $E(Y) = \int_{\mathbb{R}} y\, f_Y(y)\, dy$ exists (but it may still be infinite).

$\boxed{\text{Example}}$ A dist. that does arise in practical statistical applications is the Cauchy distribution (attributed to Augustin-Louis Cauchy (1789–1857), a French mathematician who wrote 800 published research articles & 5 textbooks in a 52-year period (15 articles/year), but actually first studied carefully by Poisson in 1824).

$$\boxed{f_Y(y) = \frac{1}{\pi(1+y^2)} \quad (-\infty < y < \infty)}$$

is the (standard) Cauchy distribution.

It does integrate to 1, but $\int_0^\infty \frac{y}{\pi(1+y^2)}\, dy = +\infty$

and $\int_{-\infty}^0 \frac{y}{\pi(1+y^2)}\, dy = -\infty$, so $E(Y)$ does not exist,

because its tails go to 0 extremely slowly.

This is because for large y, $\frac{y}{1+y^2} \doteq \frac{1}{y}$

and $\int_c^\infty \frac{1}{y}\,dy = +\infty$, the continuous

(any $c>0$) analogue of the harmonic series.

Expectation of a function of a rv

$\cancel{X}$ continuous rv with PDF $f_X(x)$, $Y \triangleq h(X)$.

method 1

work out PDF $f_Y(y)$;

then $E(Y) = \int_{\mathbb{R}} y\, f_Y(y)\, dy$.

(if this exists)

method 2 (faster)

$$E(Y) = \int_{\mathbb{R}} h(x)\, f_X(x)\, dx.$$

Discrete version:

$$E[h(X)] = \sum_{\text{all } x} h(x)\, f_X(x).$$

discrete

DS (and some other people) call Method 2 the law of the unconscious statistician (LOTUS), because Method 2 looks like a definition but is actually a (difficult) theorem (in full generality) (16 Aug 17) (measure theory: pushforward measure, ...)

Example: $X \sim \text{Exponential}(\lambda)$, $(\lambda > 0)$, $E(X) = \frac{1}{\lambda}$

$Y = X^2$: $E(Y) = \int_0^\infty x^2 \lambda e^{-\lambda x} \, dx = \frac{2}{\lambda^2}$ (integrate by parts twice)

Notice that

$$E(X^2) \neq [E(X)]^2$$

$$\frac{2}{\lambda^2} \neq \left(\frac{1}{\lambda}\right)^2$$

The only functions $Y = h(X)$ for which $E[h(X)] = h[E(X)]$ are linear: $h(x) = a + bx$, as we'll see later

Properties of $E(Y)$

① If $Y = aX + b$ then

$$E(Y) = a\,E(X) + b \quad \left(\begin{array}{c}\text{assuming}\\ E(X)\\ \text{exists}\end{array}\right).$$

② If you can find a constant a with $P(X \geq a) = 1$ then (naturally enough) $E(X) \geq a$; if b exists with $P(X \leq b) = 1$ then $E(X) \leq b$.

③ If $X_1, \dots, X_n$ are n rvs, each with finite $E(X_i)$, then $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$,

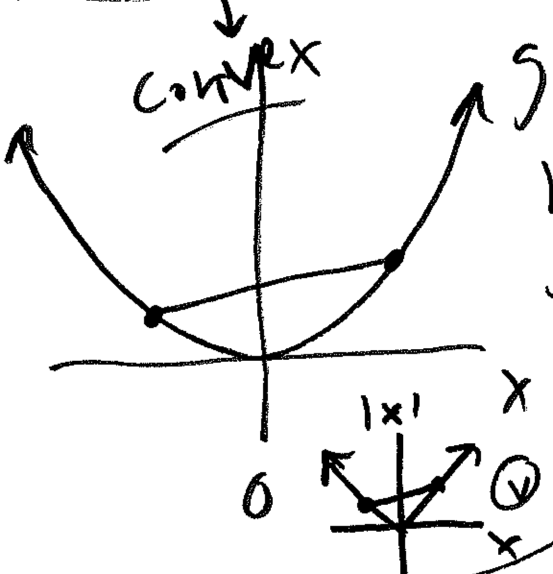
④ and $E\left[\sum_{i=1}^{n} (a_i X_i + b)\right] = \left(\sum_{i=1}^{n} a_i E(X_i)\right) + b$.

for all constants $(a_1, \dots, a_n)$ and b.

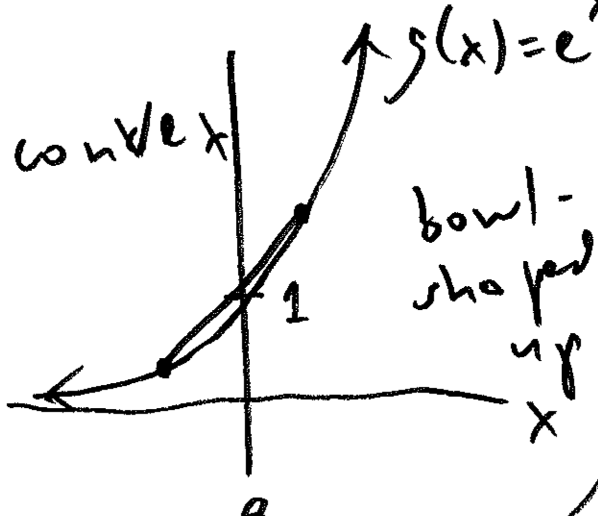
Def. A function $g: \mathbb{R}^n \to \mathbb{R}$ (this means that $g(\underset{\sim}{x}) = z$) is convex

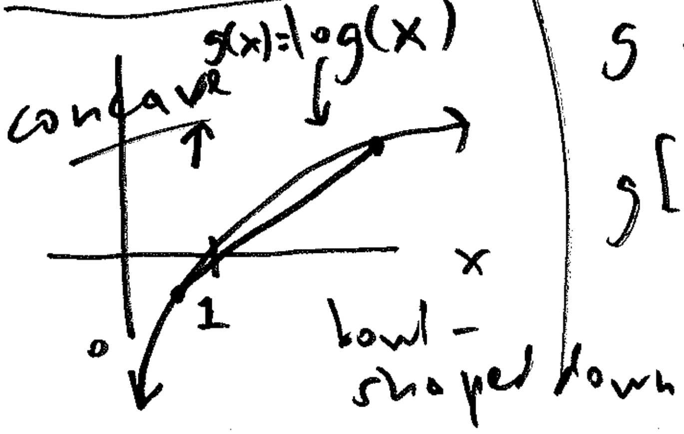
$\underset{\sim}{x} = (x_1, \dots, x_n)$ real #s if for every $0 < \alpha < 1$ and every

$\underset{\sim}{x}$ and $\underset{\sim}{y}$, $g[\alpha \underset{\sim}{x} + (1-\alpha) \underset{\sim}{y}] \leq \alpha\, g(\underset{\sim}{x}) + (1-\alpha)\, g(\underset{\sim}{y})$



Graphical version of this: pick any two points on the function & connect them with a line segment; the function is convex if the line segment entirely lies above the function except at the endpoints.





g is concave if

$$g[\alpha \underset{\sim}{x} + (1-\alpha) \underset{\sim}{y}] \geq \alpha\, g(\underset{\sim}{x}) + (1-\alpha)\, g(\underset{\sim}{y})$$

Def. The expectation of a random vector $\underset{\sim}{X} = (X_1, \ldots, X_n)$ is $E(\underset{\sim}{X}) \triangleq [\underbrace{E(X_1), \ldots, E(X_n)}_{n}]$.

(5) (a) g convex, $\underset{\sim}{X}$ random vector with finite $E(\underset{\sim}{X}) \rightarrow E[g(\underset{\sim}{X})] \geq g[E(\underset{\sim}{X})]$.

(b) g concave $\rightarrow E(g(\underset{\sim}{X})) \leq g[E(\underset{\sim}{X})]$.

Jensen's Inequality

(attributed to Johan Jensen (1859 - 1925), Danish mathematician & engineer)

(14 May 19)

Application of (3)

Suppose that $X_1, \ldots, X_n \overset{IID}{\sim}$ Bernoulli(p).

Then $E(X_i) = 0 \cdot \underset{P(X=0)}{(1-p)} + 1 \cdot \underset{P(X=1)}{p} = p$ and

$$E\Big(\underbrace{\sum_{i=1}^{n} X_i}_{\text{Binomial}(n,p)}\Big) = \sum_{i=1}^{n} E(X_i) = np = \text{mean of Binomial}(n,p)$$