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

chain rule

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

(14 Aug 17)

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

$Y = X^2$ find PDF of Y

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

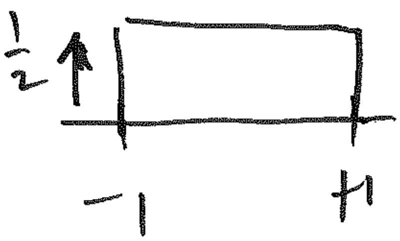
~~for $0<y<1$~~

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

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

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

$$= \sqrt{y}$$

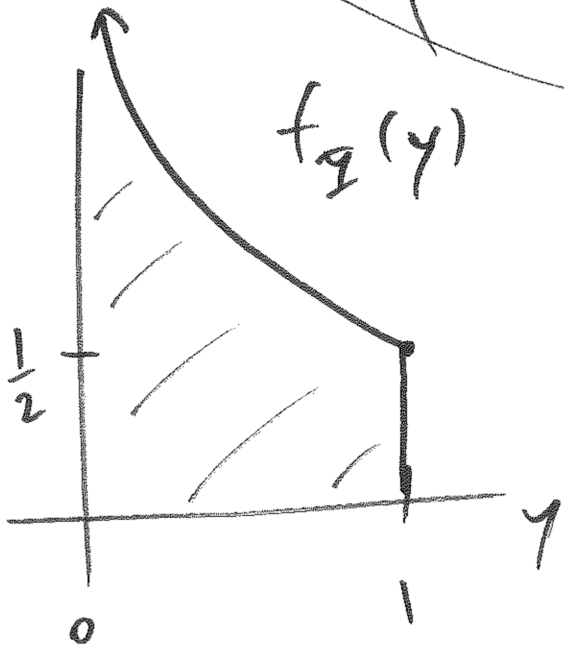


$$f_X(x) = \begin{cases} \frac{1}{2} & -1 \le x \le 1 \\ 0 & \text{else} \end{cases}$$

② Thus

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

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



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, $Y = F_X(X)$

$\rightarrow Y \sim$ 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

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

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,

$\boxed{\text{Algorithm}}$ $U_1, \dots, U_n \stackrel{IID}{\sim} \text{Uniform}(0,1)$

$F_X^{-1}(U_1), \dots, F_X^{-1}(U_n) \stackrel{IID}{\sim} F_X$

~~Earlier~~ Example ~~revisited~~ If $X \sim \text{Exponential}(\lambda)$ $(\lambda > 0)$, its PDF is

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

Earlier we saw that (p. 91) $F_X(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$

and $F_X^{-1}(p) = \dfrac{-\log(1-p)}{\lambda}$ $(0 < p < 1)$

(R demo)

Now you can see immediately that if $U \sim \text{Uniform}(0,1)$ so is $(1-U)$, so to generate IID Exponential (λ) rv you just compute $-\frac{1}{\lambda} \log U_i$, $U_i \stackrel{IID}{\sim} \text{Uniform}(0,1)$.

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.

The method used above for working out (148)

the distribution of $Y = \frac{1}{X}$ can be

generalized, as follows. Some functions $h(X)$

are *nice*, in that they are both *differentiable*

and *one-to-one* (*invertible*) 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 *monotonically increasing* or

decreasing, and h is also continuous,

so it transforms the interval (a, b) to

another open interval $h[(a, b)] = (\alpha, \beta)$

called the *image* of (a, b) under h.

Since h is invertible, it makes sense

to talk about $y = h(x) \leftrightarrow x = h^{-1}(y)$.

Theorem: X continuous rv with PDF $f_X(x)$

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

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 Y is
$$f_Y(y) = \begin{cases} f_X[h^{-1}(y)] \left| \dfrac{dh^{-1}(y)}{dy} \right| & \text{for } \alpha < y < \beta \\ 0 & \text{else} \end{cases}$$
(chain rule)

Easy short-hand way to remember this:

"Multiply" both sides by $|dy|$ to get $f_Y(y)|dy| = f_X(x)|dx|$

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

Earlier example, revisited

$Y = h(X) = \dfrac{1}{X}$: 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 $f_Y(y) = \dfrac{f_X(\frac{1}{y})}{y^2}$ as before

---

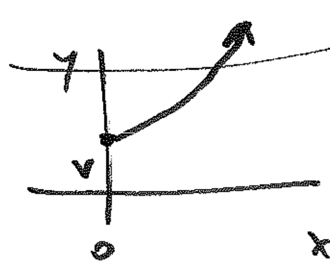
Example: At time 0, population of V organisms introduced into large tank of water with nutrients; X = rate of growth. Under one model that's realistic in some circumstances, at time t the predicted population size would be $Y = V e^{Xt}$ (exponential growth).

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

$y = h(x) = Ve^{xt}$

$x = 0 \to y = V$

$x = 1 \to y = Ve^{t}$ image



strictly increasing

$$\frac{y}{v} = e^{xt} \rightarrow \log\left(\frac{y}{v}\right) = xt \rightarrow x = h^{-1}(y) = \frac{1}{t}\log\left(\frac{y}{v}\right)$$ (49)

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

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

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
$$\left\{ \begin{array}{l} Y_1 = h_1(X_1, \ldots, X_n) \\ \vdots \qquad\qquad \vdots \\ Y_m = h_m(X_1, \ldots, X_n) \end{array} \right\} \quad (\underline{\underline{m \geq 1}})$$

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

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

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 $f_{\underset{\sim}{X}}(\underset{\sim}{x})$.

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

univariate (real) | $A_y = \{\underset{\sim}{x} : h(\underset{\sim}{x}) = y\}$

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

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

$X_i \overset{IID}{\sim}$ CDF F_{X_i}, PDF f_{X_i}

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

is defined to be

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

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

these are examples of the order statistics of $(X_1, \dots, X_n)$

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

$\Updownarrow$ iff

$$= P\left(X_1 \le t, X_2 \le t, \dots, X_n \le t\right)$$

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

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

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

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

Similarly

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

$\Updownarrow$ iff

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

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

$\overset{\text{IID}}{=} 1 - \left[1 - F_{X_i}(t)\right]^n$

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

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

Generalizing the earlier differentiable & 1-1 result

Multivariate transformations (154)

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

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

support of $(X_1, \dots, X_n)$

Define new rvs:

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

(note some arb # of Xs)

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

image of $h_1, \dots, h_n$

Inverse transform:

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

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

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

in which

J is the determinant of the matrix

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

and $|\cdot|$ is absolute value

J is called the <u>Jacobian</u> 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

$$(Y_1, Y_2) \triangleq \left(\frac{X_1}{X_2},\ X_1\cdot X_2\right)$$

$y_1 = h_1(x_1,x_2) = \dfrac{x_1}{x_2}$

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

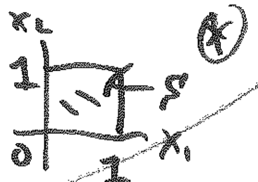
Inverse transform:

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) $\sqrt{\dfrac{y_2}{y_1}} < 1$

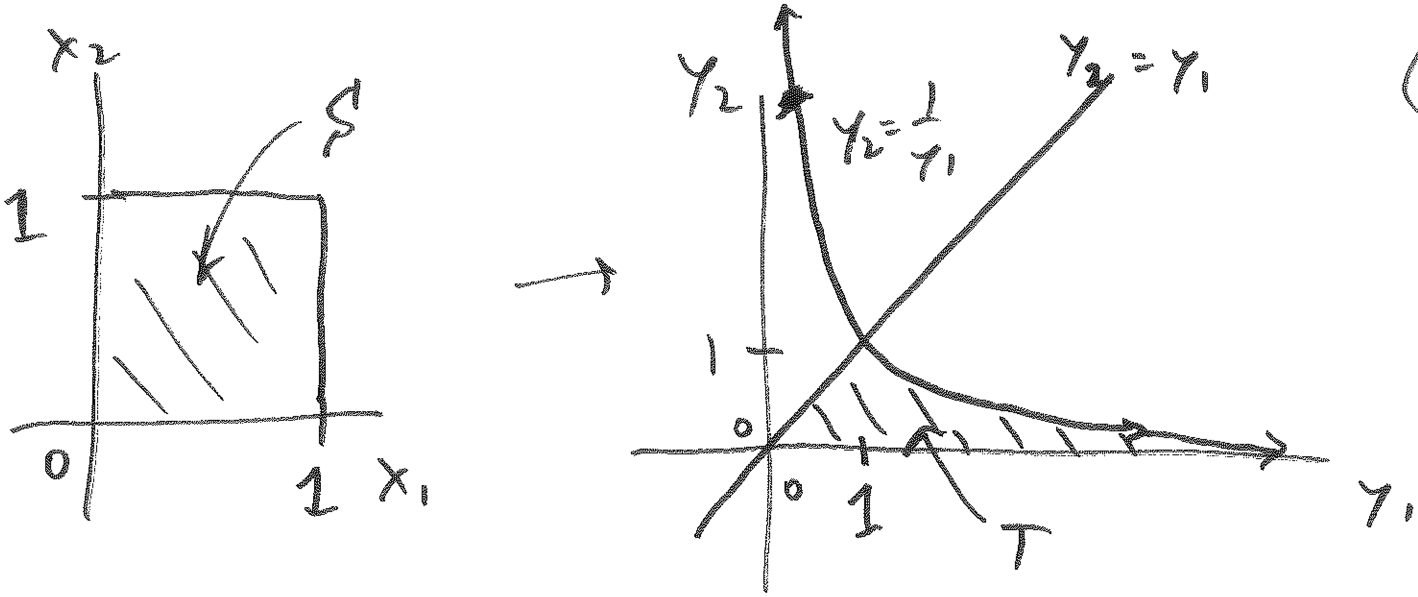
(a) equivalent to $\begin{pmatrix} y_1 > 0 \\ y_2 > 0 \end{pmatrix}$ or $\begin{pmatrix} y_1 < 0 \\ y_2 < 0 \end{pmatrix}$

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

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

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

$$\frac{\partial}{\partial y_2} h_2^{-1} = \frac{1}{2}\sqrt{\frac{1}{y_1 y_2}}$$

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

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

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

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

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

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

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

$$= \left\{ \begin{array}{ll} 2\,\frac{y_2}{y_1} & \text{for } (y_1, y_2) \in T \\ 0 & \text{else} \end{array} \right\}.$$

A useful trick

start with (X_1, X_2) joint dist.; suppose you're interested only in the dist. of univariate $Y_1 = h_1(X_1, X_2)$. (160)

Then one way to compute this dist. is with the following 3 steps.

Step 1: Find another rv $Y_2 = h_2(X_1, X_2)$ 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 (2x2) Jacobian matrix is only 1 and its determinant is therefore 0.

~~Earlier~~

Example, continued

(X_1, X_2) have joint (continuous) PDF

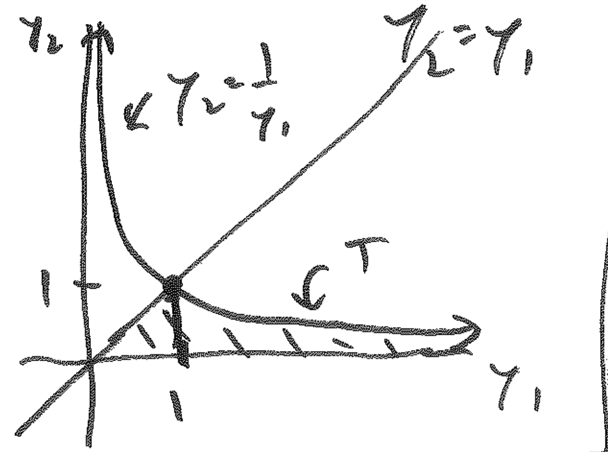
$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 4x_1x_2 & 0<x_1<1 \\ & 0<x_2<1 \\ 0 & \text{else} \end{cases}$$

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

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

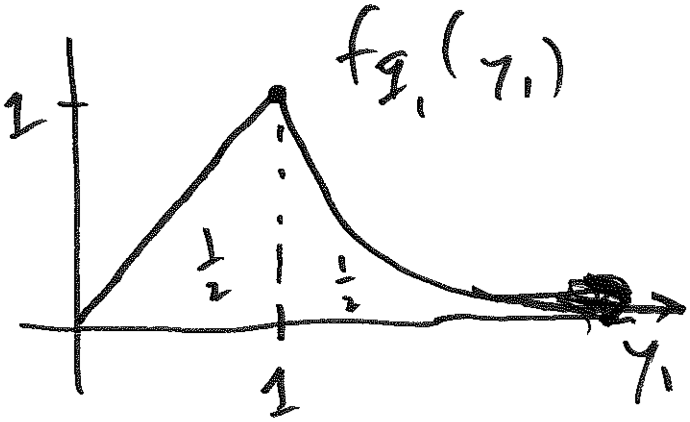
where $T = \left\{(y_1,y_2) : y_1 > 0, y_2 < \min\left(y_1, \frac{1}{y_1}\right)\right\}$.

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

${}_1\underset{\sim}{X}_n = (X_1, \ldots, X_n)$ continuous with joint PDF $f_{X_1 \cdots X_n}(x_1, \ldots, x_n)$;

${}_1\underset{\sim}{Y}_n = (Y_1, \ldots, Y_n)$ is a linear transformation of $\underset{\sim}{X}$: ${}_n\underset{\sim}{Y}^T_1 = {}_nA_n \cdot {}_n\underset{\sim}{X}^T_1$ transpose

where ${}_nA_n$ is an invertible (full-rank) matrix.

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

$$f_{\underset{\sim}{Y}}({}_n y^T_1) = \frac{f_{\underset{\sim}{X}}({}_nA^{-1}_n\, y^T_1)}{|\det A|}$$

Example

$Y_1 = X_1 + X_2$

$Y_2 = X_1 - X_2$

$${}_2A_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

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

$|\det A| = 2$

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

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

Expectation, Variance, Covariance, Correlation

Disc. 4

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

Earlier we worked out the discrete dist. of the rv

$Y = \left(\begin{array}{l}\text{\# of T-S babies in family} \\ \text{of 5, both parents carriers}\end{array}\right)$

we showed that $(Y) \sim$ Binomial (n, p) with $\begin{array}{l} n=5 \\ p=\frac{1}{4}\end{array}$

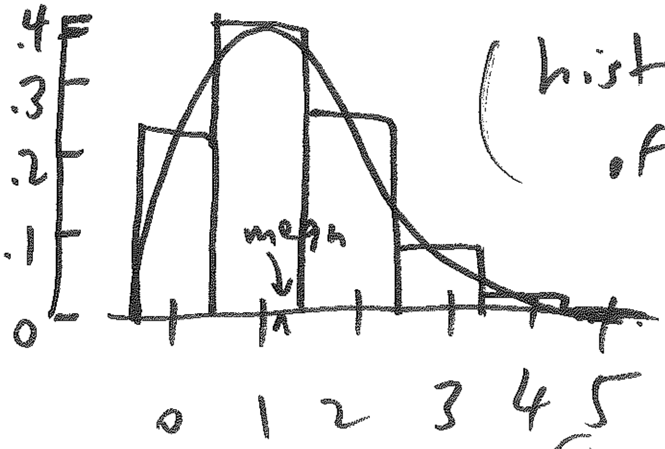
y	$P(Y=y)$
0	$\binom{5}{0}\left(\frac{1}{4}\right)^0 \frac{3}{4}^5 = 0.2373$
1	$\binom{5}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^4 = 0.3955$
2	$\binom{5}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3 = 0.2637$
3	$\binom{5}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2 = 0.0879$
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 expect to have?

(center of dist. of Y?)

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



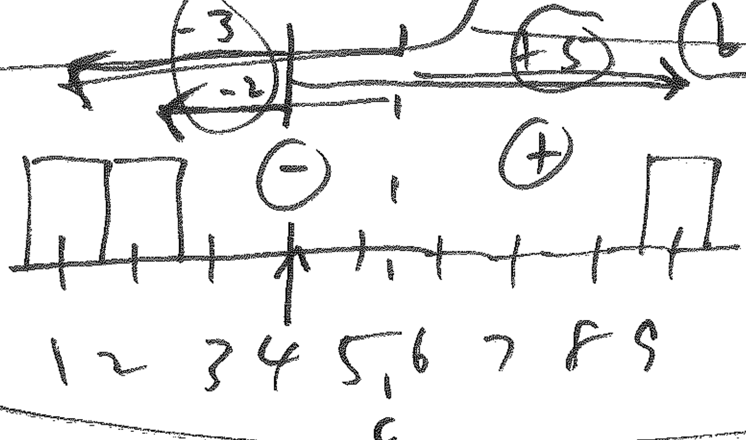
(histogram of Y)

$\boxed{A_2}$ (physics idea)

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

balance point

toy example



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

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)

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

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

y

here $\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$

mean $\bar{y} = 4$

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 <u>bounded</u> if all of its possible values are finite.

Def. Let Y be a bounded discrete rv with PF in $f_Y(y) = P(Y = y)$. The <u>mean</u> or <u>expected value</u> or <u>expectation</u> 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

$$E(Y) = (0)(.2373) + (1)(.3955) + \cdots + (5)(.0010) = 1.2500$$

suspiciously round #

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

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

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 $y=0$)

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

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

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

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

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

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

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

this $= 1$ because binomial probabilities add up to 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$(n, p) \to E(Y) = n \cdot p$.

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

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

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 <u>undefined</u> (or <u>does not exist</u>);

if at least one sum is finite, then

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

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

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

$P(X = x) = \frac{c}{x}$, $x = 1, 2, \ldots$, 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 $f_X(x) = \left\{ \begin{array}{cc} \frac{1}{x(x+1)} & x = 1, 2, \ldots \\ 0 & \text{else} \end{array} \right\}$.

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

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 $f_{\underline{X}}(x) \to E(\underline{X}) \triangleq \int_{-\infty}^{\infty} x f_{\underline{X}}(x)\,dx$.

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

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

So $E(\underline{X}) = \int_0^{\infty} \lambda x e^{-\lambda x}\,dx \overset{\text{integrate by parts}}{=} \frac{1}{\lambda}$.

For this reason, many people parameterize the exponential distribution differently:

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

$\to f_{\underline{X}}(x) = \begin{cases} \frac{1}{\eta} e^{-\frac{x}{\eta}} & x > 0 \\ 0 & \text{else} \end{cases}$

with this parameterization you can see that $E(\underline{X}) = \eta$ (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 $\int_{-\infty}^{0} y\, f_Y(y)\, dy$ and $\int_{0}^{\infty} y\, f_Y(y)\, dy$. 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).

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 published 800 research articles & 5 textbooks in a 52-year period (15 articles/year), but actually first studied carefully by Poisson in 1824).

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

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 (177)

the law of the unconscious statistician,

because method 2 looks like a definition

but is actually a (difficult) theorem (in full generality) (measure theory: pushforward measure, ...)

(16 Aug 17)

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

$Y = X^2$ | $E(Y) = \int_0^\infty x^2 \lambda e^{-\lambda x} = \frac{2}{\lambda^2}$ (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