

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

(0 < p < 1)

(R demo)

Now you can see immediately

(45)

that if $U \sim \text{Uniform}(0,1)$ so is $(1-U)$,
so to generate IID Exponential⁽²⁾ rr $\gamma \sim$

just compute $-\frac{1}{\lambda} \log U_i$, $U_i \stackrel{\text{IID}}{\sim} \text{Uniform}(0,1)$

(8Aug16)

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
(computer-based)
to conduct a simulation study driven
by pseudo-random numbers.

The method used above for working out
the distribution of $\bar{X} = \frac{1}{n} \sum X_i$ can be
generalized, as follows.

functions $h(\bar{x})$
are nice, in that they are both differentiable

and one-to-one (invertible)

real-valued

Calculus
reminder

If $h(x)$ 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 $y = h(x) \leftrightarrow x = h^{-1}(y)$. (147)

Prereq: X continuous rv with PDF $f_X(x)$
could be infinite

and for which $P(a < X < b) = 1$; $\Sigma = h(X)$,
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 $\alpha < y < \beta$ $\xrightarrow{\text{chain rule}}$ PDF

$$\text{if } \Sigma \text{ is } f_\Sigma(y) = \begin{cases} f_X[h^{-1}(y)] \left| \frac{dh^{-1}(y)}{dy} \right| & \text{for } \alpha < y < \beta \\ 0 & \text{else} \end{cases}$$

Every short-hand
way to remember this: "multiply" both sides

$$y = h(x) \quad | \quad \text{by } |dy| \text{ to get } f_\Sigma(y) |dy| = f_X(x) |dx|$$

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

$$\Sigma = h(\bar{X}) = \frac{1}{n} \sum_{i=1}^n \Sigma_i$$

Earlier
example,
revisited

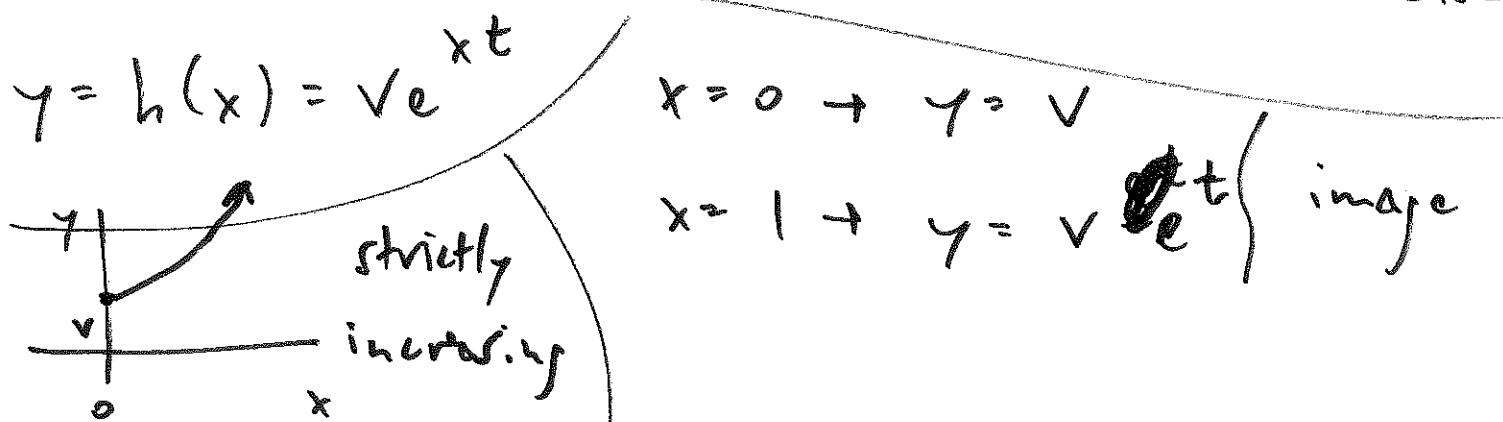
time in the bank queue

Here $y = h(x) = \frac{1}{x}$ so $x = h^{-1}(y) = \frac{1}{y}$ (148)

and $\frac{1}{y} \frac{1}{1} = -\frac{1}{y^2}$; thus $f_{\Sigma}(y) = \frac{f_{\Sigma}\left(\frac{1}{y}\right)}{y^2}$ or before

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

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



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

$$\frac{d}{dy} \frac{1}{t} \log(\frac{t}{v}) = \frac{1}{t} (\frac{t}{v})^{-1} \cdot \frac{1}{v} = \frac{1}{tv} \quad \text{Thus}$$

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

Functions
of 2 or
more rvs

Case 1:
discrete
with joint RF $f_{\Sigma}(x)$;

n rvs $\underline{\xi}_1, \dots, \underline{\xi}_n$
discrete joint dist.

$$\text{define } \left\{ \begin{array}{l} \underline{\xi}_1 = h_1(\underline{\xi}_1, \dots, \underline{\xi}_n) \\ \vdots \\ \underline{\xi}_m = h_m(\underline{\xi}_1, \dots, \underline{\xi}_n) \end{array} \right\} \quad (m \geq 1)$$

Given values $\underline{y} = (y_1, \dots, y_n)$ of $f(\underline{\xi}_1, \dots, \underline{\xi}_n)$ 150

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 $f_{\underline{\xi}}(\underline{x})$ is given by

$$f_{\underline{\xi}}(\underline{x}) = \sum_{(x_1, \dots, x_n) \in A} f_{\underline{\xi}}(\underline{x})$$

Case 2: n rvs $\underline{\xi}_1, \dots, \underline{\xi}_n$, continuous
continuous, $(n=1)$ joint dist with joint PDF $f_{\underline{\xi}}(\underline{x})$.

$\underline{\xi} = h(\underline{\xi})$ For each y define
univariate $A_y = \{\underline{x} : h(\underline{x}) = y\}$

Then PDF of $\underline{\xi}$ is $f_{\underline{\xi}}(y) = \int_{A_y} \cdots \int f_{\underline{\xi}}(\underline{x}) d\underline{x}$.

Single example of \mathbf{P}_1 , result (\bar{X}_1, \bar{X}_2) joint continuous PDF (51)

$$f_{\bar{X}_1, \bar{X}_2}(x_1, x_2), \quad \bar{Y} = g_1 \bar{X}_1 + g_2 \bar{X}_2 + b$$

with $g_1 \neq 0 \rightarrow \bar{Y}$ continuous

with PDF $f_{\bar{Y}}(y) = \int_{-\infty}^{\infty} f_{\bar{X}_1, \bar{X}_2}\left(\frac{y-b-g_2 x_2}{g_1}, x_2\right) \frac{dx_2}{|g_1|}$

Important special case The simplest thing you can do with two 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 $\bar{Y} = \bar{X}_1 + \bar{X}_2$ with $(g_1, g_2, b) = (1, 1, 0)$ to get

Dist. of \bar{Y} is called the convolution of the dists. of \bar{X}_1 and \bar{X}_2

(152)

By the
above
result

$$f_{\Sigma}(y) = \int_{-\infty}^{\infty} f_{\Sigma_1}(y-z) f_{\Sigma_2}(z) dz$$

A more
complicated

$$= \int_{-\infty}^{\infty} f_{\Sigma_1}(z) f_{\Sigma_2}(y-z) dz.$$

example

$$\Sigma_i \stackrel{\text{IID}}{\sim} \text{CDF } F_{\Sigma_i}, \text{ PDF } f_{\Sigma_i} \quad (i=1, \dots, n) \quad (\text{continuous})$$

$$\Sigma_{(1)} \triangleq \min(\Sigma_1, \dots, \Sigma_n) \quad \left. \begin{array}{l} \text{These are examples} \\ \text{of the order} \end{array} \right\}$$

$$\Sigma_{(n)} \triangleq \max(\Sigma_1, \dots, \Sigma_n) \quad \left. \begin{array}{l} \text{statistics of} \\ (\Sigma_1, \dots, \Sigma_n) \end{array} \right\}$$

$$F_{\Sigma_{(n)}}(t) = P(\Sigma_{(n)} \leq t)$$

iff

$$= P(\Sigma_1 \leq t, \Sigma_2 \leq t, \dots, \Sigma_n \leq t)$$

$\stackrel{\text{IID}}{=}$

$$= P(\Sigma_1 \leq t) \cdots P(\Sigma_n \leq t)$$

$\stackrel{\text{IID}}{=}$

$$[F_{\Sigma_i}(t)]^n$$

$$\text{So } \mathbb{I}_{(n)} \text{ has PDF } f_{\mathbb{I}_{(n)}}(t) = \frac{d}{dt} \left[F_{\mathbb{X}_i}(t) \right]^n \quad (183)$$

Similarly

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

$$F_{\mathbb{I}_{(n)}}(t) = P(\mathbb{I}_{(n)} \leq t) = 1 - P(\mathbb{I}_{(n)} > t)$$

↓ if

$$= 1 - P(\mathbb{X}_1 > t, \dots, \mathbb{X}_n > t)$$

(ID)

$$= 1 - P(\mathbb{X}_1 > t) \dots P(\mathbb{X}_n > t)$$

I(D)

$$= 1 - \left[1 - F_{\mathbb{X}_i}(t) \right]^n$$

So $\mathbb{I}_{(n)}$ has IDF

$$f_{\mathbb{I}_{(n)}}(t) = \frac{d}{dt} F_{\mathbb{I}_{(n)}}(t)$$

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

Generalizing
the earlier
differentiable
& 1-1
result

Multivariate transformations (54)
 X_1, \dots, X_n continuous joint
dist with joint PDF $f_{\underline{X}}(\underline{x})$

Suppose Σ is a subset of \mathbb{R}^n with

$$P\{(\underline{X}_1, \dots, \underline{X}_n) \in \Sigma\} = 1. \quad \boxed{\text{Define new rvs:}}$$

$$Y_1 = h_1(\underline{X}_1, \dots, \underline{X}_n)$$

$$\vdots \quad \vdots$$

$$Y_n = h_n(\underline{X}_1, \dots, \underline{X}_n)$$

(note
some
gr. #
of Σ)

Assume that the n
functions h_1, \dots, h_n

define a 1-1
differentiable

transformation of Σ onto
some subset T of \mathbb{R}^n .

Inverse
transform:

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

$$\vdots \quad \vdots$$

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

Then the joint PDF $f_{\tilde{\underline{x}}}(\underline{z})$ is (155)

$$f_{\tilde{\underline{x}}}(\underline{z}) = \left\{ \begin{array}{l} f_{\tilde{\underline{x}}}[\tilde{h}_1^{-1}(\underline{z}), \dots, \tilde{h}_n^{-1}(\underline{z})] | J| \\ \text{for } (\gamma_1, \dots, \gamma_n) \in T \\ \text{else} \end{array} \right\}$$

in which

J is the determinant of the matrix

$$\begin{bmatrix} \frac{\partial \tilde{h}_1^{-1}}{\partial \gamma_1} & \dots & \frac{\partial \tilde{h}_1^{-1}}{\partial \gamma_n} \\ \vdots & & \vdots \\ \frac{\partial \tilde{h}_n^{-1}}{\partial \gamma_1} & \dots & \frac{\partial \tilde{h}_n^{-1}}{\partial \gamma_n} \end{bmatrix} \quad \text{and } |J| \text{ is absolute value}$$

J is called the Jacobian of the transformation from $\tilde{\underline{x}}$ to \underline{z} .
named after the German mathematician

Carl Gustav Jacob Jacobi (1804 - 1851)

(1st of small pgs)
at ye 46

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

Example (\bar{X}_1, \bar{X}_2) joint

(continuous) PDF $f_{\bar{X}_1, \bar{X}_2}(x_1, x_2) = \begin{cases} 4x_1 x_2 & \text{for } 0 < x_1 < 1 \\ & \quad 0 < x_2 < 1 \\ 0 & \text{else} \end{cases}$

(check: $\int_0^1 \int_0^1 4x_1 x_2 dx_1 dx_2$)

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

$$= 2 \int_0^1 x_2 dx_2 = 2 \left. \frac{x_2^2}{2} \right|_0^1 = 1$$

let's work out the joint PDF of

$$(\bar{Y}_1, \bar{Y}_2) \triangleq \left(\frac{\bar{X}_1}{\bar{X}_2}, \bar{X}_1 \cdot \bar{X}_2 \right) \quad \begin{aligned} Y_1 &= h_1(x_1, x_2) \\ &= \frac{x_1}{x_2} \end{aligned}$$

$$Y_2 = h_2(x_1, x_2) = x_1 x_2$$

Inverse transform:

solve $\begin{cases} \frac{x_1}{x_2} = y_1 \\ x_1 x_2 = y_2 \end{cases}$ for (x_1, x_2) :

$$x_1 = h_1^{-1}(y_1, y_2)$$

$$= \sqrt{y_1 y_2}$$

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

$$= \sqrt{\frac{y_2}{y_1}}$$

defining

4 inequalities:

image: how does

$$(0 < x_1 < 1, 0 < x_2 < 1)$$

transform?

④

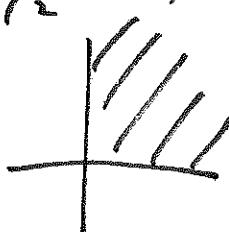
$$\begin{cases} x_1 > 0, x_1 < 1, \\ x_2 > 0, x_2 < 1 \end{cases}$$

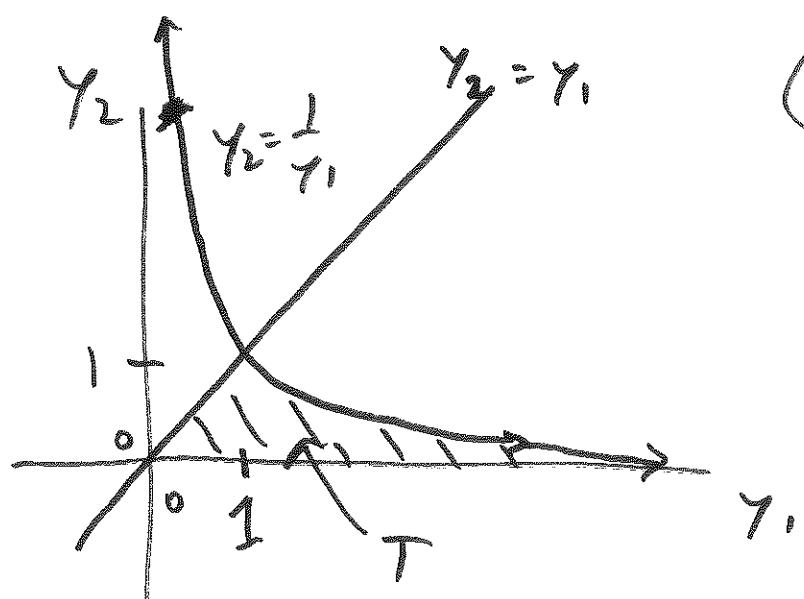
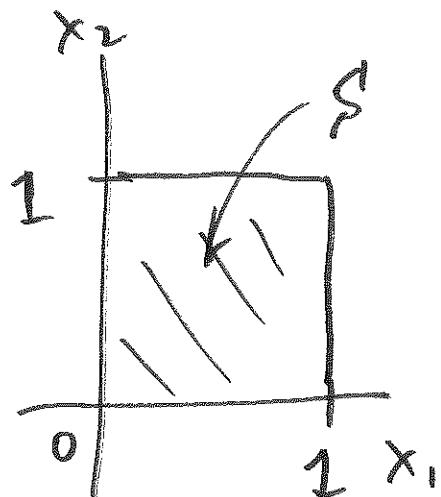
So $\begin{cases} (a) \sqrt{y_1 y_2} > 0, \\ (b) \sqrt{y_1 y_2} < 1 \end{cases}$ (a) equivalent to $(y_1 > 0)$
 $\begin{cases} (c) \sqrt{\frac{y_2}{y_1}} > 0, \\ (d) \sqrt{\frac{y_2}{y_1}} < 1 \end{cases}$ or $(y_1 < 0)$
 $\quad \quad \quad y_2 < 0$

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

(c) leads to the same thing

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





$$h_1^{-1}(\gamma_1, \gamma_2) = \sqrt{\gamma_1 \gamma_2}$$

$$\text{so } \frac{\partial}{\partial \gamma_1} h_1^{-1} = \frac{1}{2} \sqrt{\frac{\gamma_2}{\gamma_1}}$$

$$h_2^{-1}(\gamma_1, \gamma_2) = \sqrt{\frac{\gamma_2}{\gamma_1}}$$

$$\frac{\partial}{\partial \gamma_2} h_1^{-1} = \frac{1}{2} \sqrt{\frac{\gamma_1}{\gamma_2}}$$

$$\frac{\partial}{\partial \gamma_1} h_2^{-1} = -\frac{1}{2} \left(\frac{\gamma_2}{\gamma_1^3} \right)^{\frac{1}{2}}$$

$$\frac{\partial}{\partial \gamma_2} h_2^{-1} = \frac{1}{2} \sqrt{\frac{1}{\gamma_1 \gamma_2}}$$

$$\text{so } J = \det \begin{bmatrix} \frac{1}{2} \left(\frac{\gamma_2}{\gamma_1} \right)^{\frac{1}{2}} & \frac{1}{2} \left(\frac{\gamma_1}{\gamma_2} \right)^{\frac{1}{2}} \\ -\frac{1}{2} \left(\frac{\gamma_2}{\gamma_1^3} \right)^{\frac{1}{2}} & \frac{1}{2} \left(\frac{1}{\gamma_1 \gamma_2} \right)^{\frac{1}{2}} \end{bmatrix} = \frac{1}{2 \gamma_1}$$

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

and (since $\gamma_1 > 0$) $|J| = \frac{1}{2 \gamma_1}$

To finish the calculation, in the

$$\text{PDF of } \underline{\underline{x}}, \quad f_{\underline{\underline{x}}}(\underline{\underline{x}}) = \begin{cases} 4x_1 x_2 & (\text{if } x_1 < 1 \\ & \text{and } x_2 < 1) \\ 0 & \text{else} \end{cases}$$

substitute $x_1 = \sqrt{\gamma_1 \gamma_2}$, $x_2 = \sqrt{\frac{\gamma_2}{\gamma_1}}$,
and bring in the Jacobian:

$$f_{\underline{\underline{x}}}(\underline{\underline{x}}) = f_{\underline{\underline{x}}}[\tilde{h}_1(\underline{\underline{x}}), \tilde{h}_2(\underline{\underline{x}})] |\mathcal{J}|$$

$$= 4 \left(\sqrt{\gamma_1 \gamma_2} \right) \left(\sqrt{\frac{\gamma_2}{\gamma_1}} \right) \frac{1}{2\gamma_1}$$

$$= \begin{cases} 2 \frac{\gamma_2}{\gamma_1} & \text{for } (\gamma_1, \gamma_2) \in T \\ 0 & \text{else} \end{cases}$$

A useful trick) start with $(\mathbb{X}_1, \mathbb{X}_2)$ joint dist.; suppose you're interested only in the dist. of $\mathbb{Y}_1 = h_1(\mathbb{X}_1, \mathbb{X}_2)$. (160)

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

Step 1: Find

another w/ $\mathbb{Y}_2 = h_2(\mathbb{X}_1, \mathbb{X}_2)$ such that the transformation $(\mathbb{X}_1, \mathbb{X}_2) \rightarrow (\mathbb{Y}_1, \mathbb{Y}_2)$ is 1-1 with a differentiable inverse transformation & the calculations are straightforward.

Step 2 work out the joint dist. of

$(\mathbb{Y}_1, \mathbb{Y}_2)$.

Step 3

Integrate \mathbb{Y}_2 out of the joint dist. (ie., marginalize over \mathbb{Y}_2) to get the marginal dist. of \mathbb{Y}_1 .

Example of

γ_2 not
wouldn't work

$$\mathbf{I}_1 = 2 \mathbf{I}_1$$

$$\mathbf{I}_2 = 3 \mathbf{I}_1 = \frac{3}{2} \mathbf{I}_1$$

Here \mathbf{I}_2 is linearly dependent on \mathbf{I}_1 , so the rank of the (2×2) Jacobian matrix is only 1 and its determinant is therefore 0.

Earlier

Example,

continued

$(\mathbf{X}_1, \mathbf{X}_2)$ have

joint (continuous) $f_{\mathbf{X}_1, \mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2) =$

PDF

Earlier

we found

that with $(\mathbf{I}_1, \mathbf{I}_2) = \left(\frac{\mathbf{X}_1}{\mathbf{X}_2}, \mathbf{X}_1 \cdot \mathbf{X}_2 \right)$

$4 \mathbf{x}_1 \mathbf{x}_2$

$\cdot \mathbf{e}_{\mathbf{x}_1} \mathbf{e}_{\mathbf{x}_2}$

$\cdot \mathbf{e}_{\mathbf{x}_2} \mathbf{e}_{\mathbf{x}_1}$

else

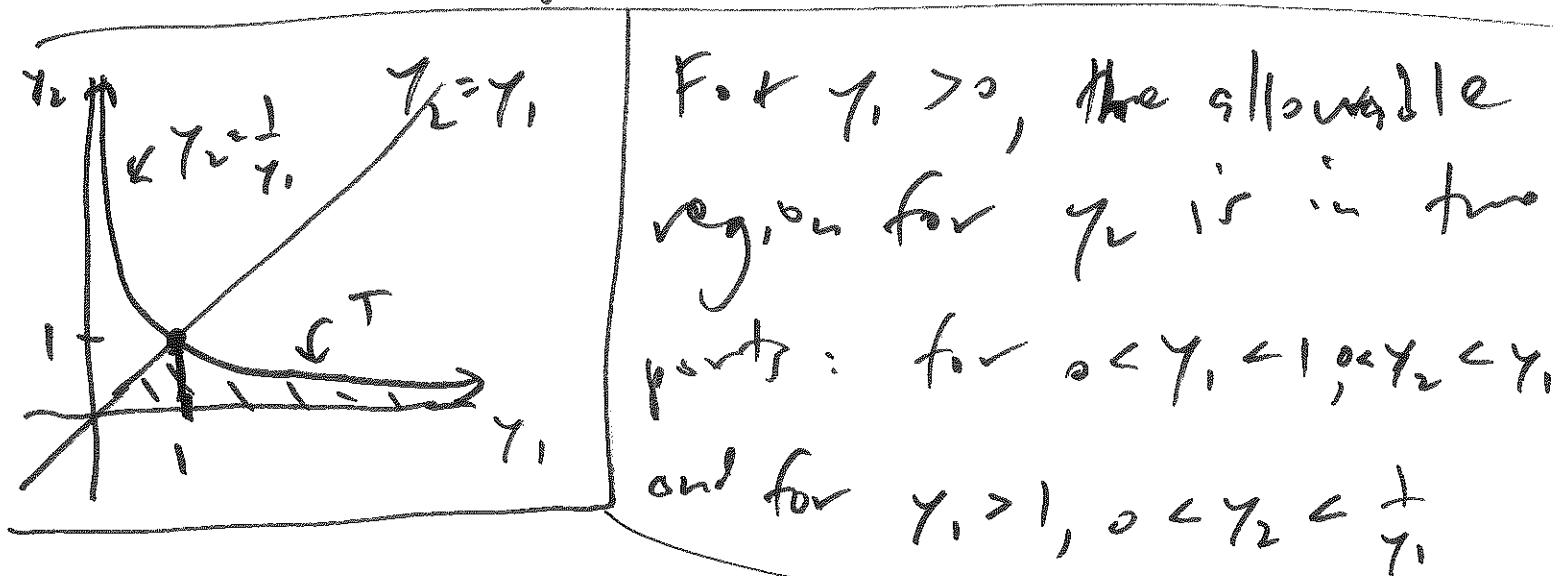
be transformed

PDF was

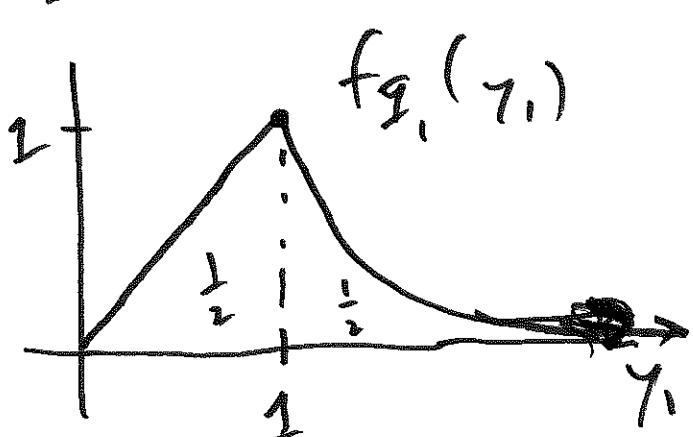
$$f_{\mathbf{I}_1, \mathbf{I}_2}(\mathbf{y}_1, \mathbf{y}_2) = \begin{cases} 2 \frac{\mathbf{y}_2}{\mathbf{y}_1} & \text{for } (\mathbf{y}_1, \mathbf{y}_2) \in T \\ 0 & \text{else} \end{cases}$$

$$\text{where } T = \left\{ (\mathbf{y}_1, \mathbf{y}_2) : \mathbf{y}_1 > 0, \mathbf{y}_2 < \min\left(\mathbf{y}_1, \frac{1}{\mathbf{y}_1}\right) \right\}.$$

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



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



weird PDF: not
~~continuous~~ differentiable
 at $y_1 = 1$

useful consequence
· f Jacobian story

$\underline{X} = (X_1, \dots, X_n)$ continuous
with joint PDF $f_{\underline{X}_1, \dots, \underline{X}_n}(x_1, \dots, x_n)$. (16.3)

$\underline{Z} = (Z_1, \dots, Z_n)$ is a linear transformation of \underline{X} : $\underline{Z}^T = A \cdot \underline{X}^T$,
where A is an invertible (full-rank) transpose matrix.

Then the PDF of \underline{Z} is

$$f_{\underline{Z}}(\underline{z}) = \frac{f_{\underline{X}}(A^{-1}\underline{z})}{|\det A|}$$

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$A^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

$\det A = -2$

$= ad - bc$

$|\det A| = 2$

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

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

Example

$$Z_1 = X_1 + X_2$$

$$Z_2 = X_1 - X_2$$

Expectation,
Variance,
Covariance,
Correlation

we showed

Def. 4

Example: Tag-¹⁸⁴
Jacks (T-J)
disrael (continued)

Earlier we worked out the
discrete dist. of the rv

$\Omega = \{ \# \text{ of T-J's dist. in family}$
 $\text{of } 5, \text{ both parents carriers} \}$

but $(\mathbb{I}) \sim \text{Binomial } (n, p) \text{ with } \begin{cases} n=5 \\ p=\frac{1}{4} \end{cases}$

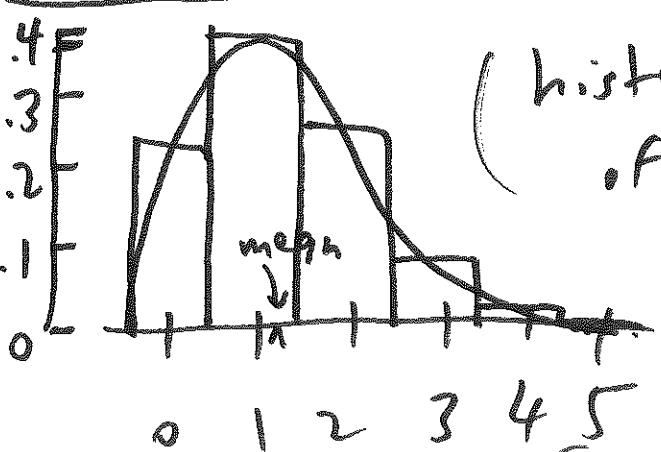
y	$P(\Omega=y)$
0	$\binom{5}{0} \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^5 = 0.2373$
1	$\binom{5}{1} \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^4 = 0.3955$
2	$\binom{5}{2} \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^3 = 0.2637$
3	$\binom{5}{3} \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2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

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

Q: About how
many T-J babies
should these parents
expect to have?

(center of dist.
of \mathbb{I})

A₁ Most likely outcome is 1 T-s tally
 (note of the dist. of \bar{I})



(histogram
of \bar{I})

A₂

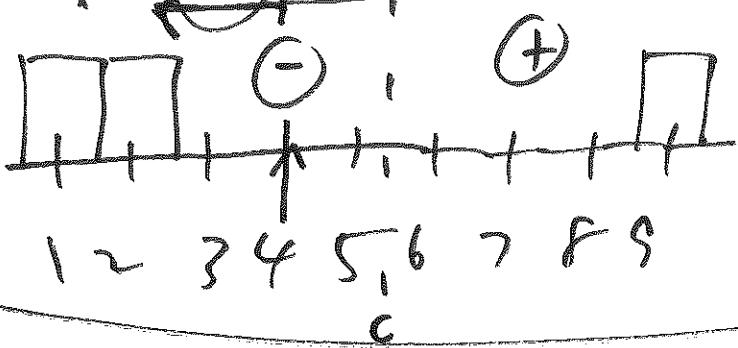
(physics idea)

let's work out the
center of mass

of the distribution

balance point

toy
example



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

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

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

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

(166)

A₃ median of $\sum_{i=1}^n Y_i - nc = 0 \Leftrightarrow$

the dist. of Σ

(here Pct's
also 1)

here $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

now $\bar{Y} = 4$

$$c = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y} = \text{the sample mean}$$

of the (sample) dataset
 $\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$

Here each value of Σ

occurred only once.

$$\bar{Y} = \sum_{i=1}^n \left(\frac{1}{n}\right) Y_i \quad \text{Def.}$$

If some values are more probable than others, the generalization of $\left(\frac{1}{n}\right)$ weight on each

Y value would be each to weight Y by its probability

$$P(\Sigma = y)$$

A n is bounded if all of its possible values are finite. Def.

let Σ be a bounded discrete rv with PF

$$f_{\Sigma}(y) = P(\Sigma = y).$$

mean or expected value or expectation of Σ ,

$$\text{is } E(\mathbb{I}) \stackrel{\Delta}{=} \sum_{y=0}^n y P(\mathbb{I}=y) = \sum_{\text{all } y} y f_{\mathbb{I}}(y)$$

T-s
example

$$E(\mathbb{I}) = (0)(.2373) + (1)(.3955)$$

~~$$+ \dots + (5)(.0012) = 1.2500$$~~

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

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

↑
surprisingly
round
#

since
summand
is 0
for $y=0$)

the
series

but
 $n > 1$;

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

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

proof
for

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

$n-p^2$
is on
the next
page

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

$$= n \gamma \sum_{y=1}^n \binom{n-1}{y-1} p^{y-1} (1-p)^{n-1-(y-1)} \quad (168)$$

$$= np \left[\sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \right] \quad (\text{substitute})$$

\therefore if $I \sim \text{Binomial}(n-1, p)$ dist. because binomial probabilities add up to 1

$$I \sim \text{Binomial}(n, p)$$

$$\text{for } n \geq 1, E(I) = np$$

When $n=1$, $\text{Binomial}(1, p) = \text{Bernoulli}(p)$.

$$\text{In this case } E(I) = 0 \cdot P(I=0) + 1 \cdot P(I=1)$$

$$\begin{aligned} \therefore \text{for all } n \geq 1 \text{ (integer)} \\ &= 0 \cdot (1-p) + 1 \cdot p = p \\ &= np \text{ with } n=1 \end{aligned}$$

and $0 < p < 1$, $I \sim \text{Binomial}(n, p) \rightarrow E(I) = np$.

T-S example $(h=5, p=\frac{1}{4}) E(\bar{X}) = \frac{5}{4} = 1.25$ ✓ 169

If discrete \bar{X} is unbounded, the expectation of \bar{X} may not exist, either because

$$\sum_{x < 0} x f_{\bar{X}}(x) = -\infty \quad (\text{and } \sum_{x > 0} x f_{\bar{X}}(x) = +\infty)$$

or the distribution "puts too much mass

"near $\pm\infty$ "

Def.

\bar{X} discrete rv with

PF $f_{\bar{X}}(x)$; consider $\sum_{x < 0} x f_{\bar{X}}(x)$ and

$\sum_{x > 0} x f_{\bar{X}}(x)$. If both sums are infinite,

$E(\bar{X})$ is undefined (or does not exist);

if at least one sum is finite, then

$E(\bar{X}) = \sum_{\text{all } x} x f_{\bar{X}}(x)$ exists (it may still be infinite)

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

Example

The harmonic

$$\text{series } \left(1 + \frac{1}{2} + \frac{1}{3} + \dots\right) = \sum_{x=1}^{\infty} \frac{1}{x} \text{ was known}$$

to the ancient Greeks, because the wavelengths of the overtones of a vibrating string are $\frac{1}{2}, \frac{1}{3}, \dots$ if 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 philosopher Nicole Oresme ($\sim 1320 - 1382$).

It's clear from this divergence that (17)
you can't create a rv \bar{X} with PF

$P(\bar{X}=x) = \frac{c}{x}, x=1, 2, \dots$, because the
probability ^{would} sum to ∞ . But $P(\bar{X}=x) = \frac{c}{x^2}$

or $P(\bar{X}=x) = \frac{c}{x(x+1)}$ turns out to work;

for example, $\sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6}$ (!) and, even

more conveniently, $\sum_{x=1}^{\infty} \frac{1}{x(x+1)} = 1$.

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

Example 1 | $f_{\bar{X}}(x) = \begin{cases} \frac{1}{x(x+1)} & x=1, 2, \dots \\ 0 & \text{else} \end{cases}$.

$$E(\bar{X}) = \sum_{x=1}^{\infty} x \cdot \frac{1}{x(x+1)} = \sum_{x=1}^{\infty} \frac{1}{x+1} = +\infty \quad (172)$$

So $E(\bar{X})$ exists, it's just infinite.

Example 2

$$f_{\bar{X}}(x) = \begin{cases} \frac{1}{2|x|(|x|+1)} & x = \pm 1, \pm 2, \dots \\ 0 & \text{else} \end{cases}$$

We already knew that $\sum_{\text{all } x} f_{\bar{X}}(x) = 1$, so \bar{X} is well-defined rv; but

$$\sum_{x=-1}^{\infty} x \cdot \frac{1}{2|x|(|x|+1)} = -\infty$$

and $\sum_{x=1}^{\infty} x \cdot \frac{1}{2x(x+1)} = +\infty$, so $E(\bar{X})$

does not exist.

We won't work with pathological rv, mostly.

Expectation
for continuous
rvs

Def. \bar{X} bounded
continuous rv

with PDF $f_{\bar{X}}(x) \rightarrow E(\bar{X}) \stackrel{(173)}{=} \int_{-\infty}^{\infty} x f_{\bar{X}}(x) dx$

Example) $\bar{X} \sim \text{Exponential}(2)$ ($2 > 0$):

recall that $f_{\bar{X}}(x) = \begin{cases} 2e^{-2x} & \text{for } x > 0 \\ 0 & \text{else} \end{cases}$

$$\text{So } E(\bar{X}) = \int_0^{\infty} 2x e^{-2x} dx \stackrel{\text{integrate by parts}}{=} \frac{1}{2}.$$

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

Alternative definition $\bar{X} \sim \text{Exponential}(\gamma)$ ($\gamma > 0$)

$$+ f_{\bar{X}}(x) = \begin{cases} \frac{1}{\gamma} e^{-\frac{x}{\gamma}} & x > 0 \\ 0 & \text{else} \end{cases}$$

with this parameterization you can see that $E(\bar{X}) = \gamma$ (easier to interpret).

Nevertheless, to avoid confusion with (74)
DS, I'll stick with $\lambda e^{-\lambda x}$. If continuous

rv Σ is unbounded, a lit of care is once again required to define $E(\Sigma)$. Def.

Σ continuous rv with PDF $f_{\Sigma}(y)$; consider

$\int_{-\infty}^{\infty} y f_{\Sigma}(y) dy$ and $\int_0^{\infty} y f_{\Sigma}(y) dy$. If both integrals are infinite, $E(\Sigma)$ is undefined (or does not exist); if at least one of these integrals is finite, $E(\Sigma) = \int_{\mathbb{R}} y f_{\Sigma}(y) dy$ 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))

*published
a French mathematician who wrote 800
85 textbooks
regards articles in a 52-year period ($\frac{15}{\text{year}}$)
but actually first studied carefully by

Poisson in 1824).
$$f_{\Sigma}(y) = \frac{1}{\pi(1+y^2)} (-\infty, \infty)$$

is the (standard) Cauchy distribution.

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

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

because its tails go to ∞ extremely slowly.

This is because for large γ , $\frac{1}{1+\gamma^2} \approx \frac{1}{\gamma}$
and $\int_c^\infty \frac{1}{\gamma} dy = +\infty$, the continuous
analogue of the harmonic series.

(any $c > 0$)
Expectation
of a function
of a rv



\mathbb{E} ^{continuous}
rv with PDF

$$f_{\mathbb{E}}(x), \mathbb{E} \stackrel{\text{def}}{=} h(\mathbb{E})$$

Method 1 Work out PDF $f_{\mathbb{E}}(\gamma)$;

then $E(\mathbb{E}) = \int_R \gamma f_{\mathbb{E}}(\gamma) d\gamma$.

K (if m exists)

Method 2
(faster)

$$E(\mathbb{E}) = \int_R h(x) f_{\mathbb{E}}(x) dx.$$

Discrete
version:

$$E[h(\mathbb{E})] = \sum_{\text{all } x} h(x) f_{\mathbb{E}}(x).$$

↑
discrete

DS (and some other people) call Method 2 (171)
the law of the Unconscious Statistician,
because Method 2 looks like a definition
but is actually a theorem (difficult)
~ (in full generality) measure theory:
pushforward measure, ...)

Example) $\Sigma \sim \text{Exponential}(2), (\lambda > 0)$

$E(\Sigma) = \frac{1}{\lambda}$

$\Sigma = \Sigma^2$

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

Notice that

$E(\Sigma^2) \neq (E(\Sigma))^2$ the only functions

$\frac{2}{\lambda^2} \neq \left(\frac{1}{\lambda}\right)^2$ $\Sigma = h(\Sigma)$ for which $E[h(\Sigma)] = h[E(\Sigma)]$

are linear: $h(x) = a + bx$, as we'll see later

(10 Aug 16)