5/14/19

Lecture 13

New office hours: Tu/Th 7:10-8:40pm @E2 194

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

Let A be the set of points $(x_1, \dots, x_n)$ such that

$$\begin{Bmatrix} y_1 = h_1(x_1, \dots, x_n) \\ \vdots \\ y_m = h_m(x_1, \dots, x_n) \end{Bmatrix}$$

Then the joint pmf of $f_Y(y)$ 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 $f_{\underset{\sim}{X}}(x)$

$Y = h(\underset{\sim}{X})$

$\uparrow$

univariate (real)

For each real y define $A_y = \{x : h(x) = y\}$

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

Ex: (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 $a_1 \neq 0 \to Y$ continuous

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

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 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 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

$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y-z) f_{X_2}(z)\, dz$ by the above result

A more complicated example (THT2 #4)

$X_i \overset{\text{IID}}{\sim}$ CDF F_X, PDF f_X $(i=1,\dots,n)$ (continuous)

$Y_1 \triangleq \min(X_1, \dots, X_n)$ $\quad Y_n \triangleq \max(X_1, \dots, X_n)$

is defined to be

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

$F_{Y_{(n)}}(t) = P(Y_{(n)} \le t)$ ← biggest one is less than or equal to t

iff

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

$\overset{\text{IID}}{=} P(X_1 \le t) \dots P(X_n \le t)$

$\overset{\text{IID}}{=} [F_X(t)]^n$

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

$$= n\left[F_X(t)\right]^{n-1} f_X(t)$$

Similarly

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

$\Updownarrow$ iff

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

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

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

So $Y_{(1)}$ has PDF

$$f_{Y_{(1)}}(t) = \frac{d}{dt} F_{Y_{(1)}}(t) = n\left[1 - F_X(t)\right]^{n-1} f_X(t)$$

Generalizing the earlier differentiable and 1-1 result

Multivariate transformations

$X_1, \dots, X_n$ continuous joint dist with joint PDF $f_X(x)$

Suppose that there is a subset S (support of $(X_1, \dots, X_n)$ under f_X) of $\mathbb{R}^n$ with

$$P[(X_1, \dots, X_n \in S] = 1$$

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

$$\vdots \qquad \vdots$$

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

We're generalizing this:

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

so that we can get the joint PDF $f_{\underset{\sim}{Y}}(\underset{\sim}{y})$:

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

in which J is the determinant of the matrix

$$\begin{bmatrix} \dfrac{dh_1^{-1}}{dy_1} & \cdots & \dfrac{dh_1^{-1}}{dy_n} \\ \vdots & & \vdots \\ \dfrac{dh_n^{-1}}{dy_1} & \cdots & \dfrac{dh_n^{-1}}{dy_n} \end{bmatrix}$$

(chain rule generalization)

$|\cdot|$ is absolute value

J: Jacobian of the transformation from $\underset{\sim}{X}$ to $\underset{\sim}{Y}$

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

Ex: $(\mathbb{X}_1, \mathbb{X}_2)$ joint (continuous) PDF

$$f_{\mathbb{X}_1 \mathbb{X}_2}(x_1, x_2) = \begin{cases} 4x_1x_2 & \text{for } 0<x_1<1 \text{ and } 0<x_2<1 \\ 0 & \text{else} \end{cases}$$

Check: $\int_0^1 \int_0^1 4x_1 x_2 \, dx_1 \, dx_2$

$$= \int_0^1 4x_2 \left(\int_0^1 x_1 \, dx_1 \right) dx_2$$

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

$$= 2 \int_0^1 x_2 \, dx_2$$

$$= 2 \frac{x_2^2}{2} \Big|_0^1$$

$$= 1$$

(useful for THT 2 #2)

Are x_1 and x_2 independent based on intuition?
Yes, the support set for the bivariate random variable does not involve any entanglement between x_1 and x_2
You could also marginalize $4x_1x_2$ to $2x_1 \cdot 2x_2$

Let's work out the joint PDF of

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

$$y = h_1(x_1, x_2) = \frac{x_1}{x_2}$$

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

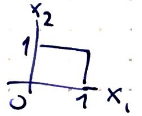
Inverse transform

Solve $\left\{ \begin{array}{l} \frac{x_1}{x_2} = y_1 \\ x_1 x_2 = y_2 \end{array} \right\}$ for (x_1, x_2):

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

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

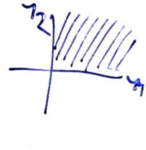
Image: How does $(0 < x_1 < 1,\ 0 < x_2 < 1)$ transform?



$\left\{ \begin{array}{l} x_1 > 0,\ x_1 < 1 \\ x_2 > 0,\ x_2 < 1 \end{array} \right\}$ so (a) $\sqrt{y_1 y_2} > 0$

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

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

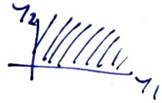


(b) $\sqrt{y_1 y_2} < 1$

says $y_2 < \frac{1}{y_1}$

c) $\sqrt{\frac{y_2}{y_1}} > 0$

Also leads to $\begin{pmatrix} y_1 > 0 \\ y_2 > 0 \end{pmatrix}$

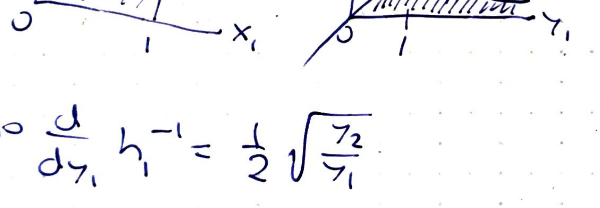


(d) $\sqrt{\frac{y_2}{y_1}} < 1$

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 $\frac{d}{dy_1} h_1^{-1} = \frac{1}{2}\sqrt{\frac{y_2}{y_1}}$

$\frac{d}{dy_2} = h_1^{-1} = \frac{1}{2}\sqrt{\frac{y_1}{y_2}}$

$\frac{d}{dy_1} h_2^{-1} = -\frac{1}{2}\left(\frac{y_2}{y_1^3}\right)^{1/2}$

$$\frac{d}{dy_2} = h_2^{-1} = \frac{1}{2}\sqrt{\frac{1}{y_1 y_2}}$$

With those partials you build a matrix

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

(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 $\underset{\sim}{X}$

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \begin{cases} 4x_1 x_2 & (0 < x_1 < 1 \text{ and } 0 < x_2 < 1) \\ 0 & \text{else} \end{cases}$$

Substitute $x_1 = \sqrt{y_1 y_2}$ $x_2 = \sqrt{\frac{y_2}{y_1}}$

and bring in the Jacobian:

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

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

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

<u>useful trick:</u> Start with (X_1, X_2) joint dist.;

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

<u>Step 1:</u> Find another rv $Y_2 = h_2(X_1, X_2)$ such that the transform $(X_1, X_2) \to (Y_1, Y_2)$ is 1 to 1 with a differentiable inverse transformation and the calculations are straightforward.

<u>Step 2:</u> Work out the joint dist of $(Y_1\ Y_2)$

<u>Step 3:</u> Integrate Y_2 out of the joint distribution (i.e. marginalize over Y_2) to get the marginal dist. of Y_1

Ex of a Y_2 that wouldn't work: $Y_1 = 2X_1$

collapsed from 2 dimensions to one $\Rightarrow$ $Y_2 = 3X_1 = \frac{3}{2}Y_1$

You need two vectors that are orthogonal to each other

Here Y_2 is linearly dependent on Y_1 so the rank of the (2x2) Jacobian matrix is only 1 and its dominant is therefore zero.

<u>Earlier example included:</u> (X_1, X_2) have joint (continuous) PDF $f_{X_1 X_2}(x_1, x_2) = \begin{cases} 4x_1 x_2 & \text{for } 0 < x_1 < 1 \text{ and } 0 < x_2 < 1 \\ 0 & \text{else} \end{cases}$

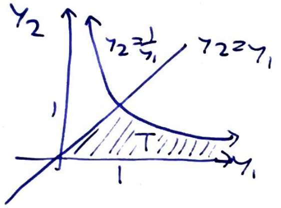
We found that w/ $(Y_1, Y_2) = \left(\dfrac{X_1}{X_2}, X_1 \cdot X_2 \right)$

the transformed PDF was

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

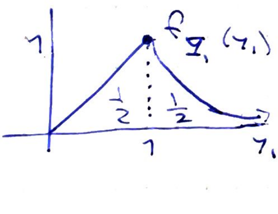
Where $T = \{ (y_1, y_2) : y_1 > 0, y_2 < \min(y_1, \frac{1}{y_1}) \}$

Suppose you were only really interested in the marginal dist of $Y_1 = \dfrac{X_1}{X_2}$; 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 $f_{Y_1}(y_1) = \begin{cases} \displaystyle\int_0^{y_1} 2\left(\frac{y_1}{y_2}\right) dy_2 = y_1 & \text{for } 0 < y_1 < 1 \\ \displaystyle\int_0^{\frac{1}{y_1}} 2\left(\frac{y_2}{y_1}\right) dy_2 = y_1^{-3} & \text{for } y_1 > 1 \end{cases}$



Weird PDF: not differentiable at $y_1 = 1$

Useful Consequence of Jacobian Story

$\underset{1\sim n}{\underline{X}} = (X_1, \ldots, X_n)$ continuous with joint PDF

$f_{X_1 \ldots X_n}(x_1, \ldots, x_n)$

$\underset{1\sim n}{\underline{Y}} = (Y_1, \ldots, Y_n)$ 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 Y is $f_{\underset{\sim}{Y}}(\underset{\sim 1}{y}) = \dfrac{f_X({}_nA_n\, y_1^T)}{|\det A|}$

Ex: $Y_1 = X_1 + X_2$

$Y_2 = X_1 - X_2$ $\quad {}_2A_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

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

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

recall that

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

Expectation, Variance, Covariance, Correlation

Ex: T-S disease

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	$\binom{5}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^5 = 0.2373$
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

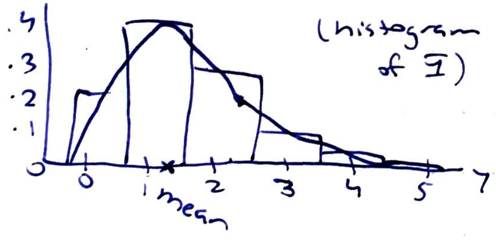
$$P(Y=y) = \begin{cases} \binom{n}{y} p^y (1-p)^{n-y} & \text{for } y = 0, 1, \ldots n \\ 0 & \text{else} \end{cases}$$

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

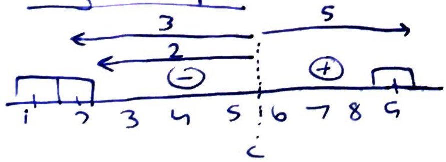
(Center of dist. of Y)?

A_1: Most likely outcome is 1 T-S baby (mode of the dist. of Y)

A_2: (Physics idea) Work out the center of mass of the distribution (balance point)



Toy example



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

outlier

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):

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \to \begin{bmatrix} y_1 - c \\ \vdots \\ y_n - c \end{bmatrix} \qquad \sum_{i=1}^{n} (y_i - c) = 0 = \left(\sum_{i=1}^{n} y_i \right) - nc = 0$$

want sum $= 0$

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

$$\sum_{i=1}^{n} y_i - nc = 0 \longleftrightarrow c = \frac{1}{n} \sum_{i=1}^{n} y_i \triangleq \bar{y} =$$ the sample mean of the sample dataset $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

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 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 ex: $E(Y) = (0)(.2373) + (1)(.3955) + \ldots + (5)(.0010) = 1.2500000$

suspiciously round #

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

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

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

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

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

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

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

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

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

this $= 1$ because binomial probabilities add up to 1

So if $Y \sim \text{Binomial}(n,p)$ for $(n>1)$, $E(Y) = np$

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

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

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

$$= np \text{ with } n=1$$

So for all $n \geq 1$ (integer) and $0 < p < 1$

$X \sim \text{Binomial}(n, p) \rightarrow E(X) = np$

T-S Ex: $(n = 5,\ p = \frac{1}{4})$ $E(X) = \frac{5}{4} = 1.25$ ✓

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

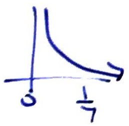
$$\sum_{x<0} x f_X(x) = -\infty \left(\text{and/or} \sum_{x \geq 0} x f_X(x) = +\infty\right)$$

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



$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\, dy$$

continuous rv

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

Ex: The harmonic series $\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \dots\right) = \sum_{x=1}^{\infty} \frac{1}{x}$ 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 $\sum_{x=1}^{\infty} \frac{1}{x} = +\infty$ (ie the harmonic series diverges) was first shown in the 1300s by the French philosopher Nicole Oresme (1300 - 1382)

It's clear from this divergence that you can't create a rv X with pmf $P(X=x) = \frac{c}{x}$, $x = 1, 2, \dots$, because the probabilities 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 ex, $\sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6}$ and even more conveniently,

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

The book uses this to construct two pathological discrete distributions to show what can go wrong with the idea of expectation.

Ex 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.

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

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 rv

Def: X bounded continuous rv with PDF

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

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

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

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 parameterize the exponential distribution differently

Alternative def: $X \sim \text{Exponential}(\xi)$ $(\xi > 0)$

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

With this parameterization you can see that $E(X) = \xi$ (easier to interpret)

If continuous rv Y is unbounded, a bit of care is once again required to define $E(Y)$

⋮

Def: Y continuous rv with PDF $f_Y(y)$

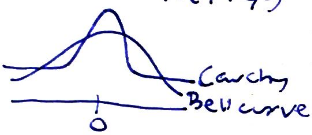
Consider $\int_{-\infty}^{0} y f_Y(y)\,dy$ and $\int_{0}^{\infty} y f_Y(y)\,dy$

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)

Ex: A dist. that does arise in practical statistical applications is the Cauchy distribution

$f_Y(y) = \frac{1}{\pi(1+y^2)}$ $(-\infty < y < \infty)$ is the standard Cauchy dist.



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, $\frac{y}{1+y^2} \approx \frac{1}{y}$ and $\int_{c}^{\infty} \frac{1}{y}\,dy = +\infty$

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

Expectation of a function of a rv

X continuous rv with PDF $f_X(x)$, $Y \triangleq h(X)$

Method 1

Work out PDF $f_Y(y)$; then $E(Y) = \int_{\mathbb{R}} y f_Y(y)\, dy$

Method 2 (faster) if this exists

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

Discrete version:

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

discrete

Ex: $X \sim$ exponential(λ) $(\lambda > 0)$

$E(X) = \frac{1}{\lambda}$ $\quad E(Y) = \int_0^\infty x^2 \lambda e^{-\lambda x}\, dx = \frac{2}{\lambda^2}$

$Y = X^2$

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: $h(x) = a + bx$

Properties of $E(Y)$

1) If $Y = aX + b$ then $E(Y) = aE(X) + b$ (assuming $E(X)$ exists)

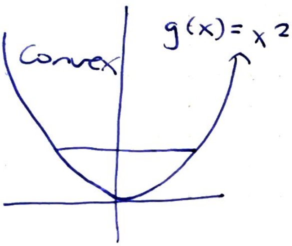
2) If you can find a constant a with $P(X \geq a) = 1$ then (naturally enough) $E(X) \geq a$; if b exists with $P(X \leq b) = 1$ then $E(X) \leq b$

3) If $X_1, \dots, X_n$ are n rvs, each with finite $E(X_i)$, then

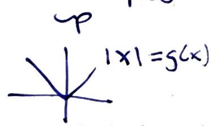
$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i)$$

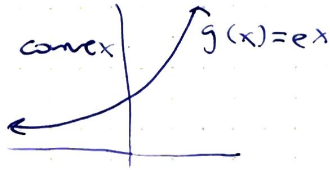
4) $$E\left[\sum_{i=1}^n (a_i X_i + b)\right] = \left(\sum_{i=1}^n a\, E(X_i)\right) + b$$ for all constants $(a_1, \dots, a_n)$ and b

Def: A function $g: \mathbb{R}^n \to \mathbb{R}$ (this means that $g(\underset{\sim}{x}) = z$)
$\uparrow$ $(x_1, \dots, x_n)$ $\uparrow$ real #s
is convex if for every $0 < \alpha < 1$ and every $\underset{\sim}{x}$ and $\underset{\sim}{y}$, $g[\alpha \underset{\sim}{x} + (1-\alpha)\underset{\sim}{y}] \leq \alpha g(\underset{\sim}{x}) + (1-\alpha) g(\underset{\sim}{y})$

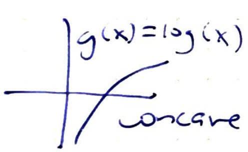


bowl-shaped up





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 lies entirely above the function except at the endpoints



bowl-shaped down

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

Def: The expectation of a random vector $\underset{\sim}{X} = (X_1, \dots, X_n)$ is $E(\underset{\sim}{X}) \triangleq [E(X_1), \dots, E(X_n)]$

$\longleftarrow n \longrightarrow$

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

Application of 3)

suppose that $X_1, \dots, X_n \overset{\text{IID}}{\sim}$ Bernoulli(p)

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

and $E(\underset{\substack{\uparrow \\ \text{binomial}(n,p)}}{\sum_{i=1}^{n} X_i}) = \sum_{i=1}^{n} E(X_i) = np =$ mean of Binomial(n,p)