useful consequence of Jacobian story

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

$\underset{1\sim n}{\underline{Y}} = (Y_1, \ldots, Y_n)$ 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 $\underset{\sim}{Y}$ 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

$Y_1 = X_1 + X_2$

$Y_2 = X_1 - X_2$

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

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

$|\det A| = 2$

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

(recall that)

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

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 $n = 5$, $p = \frac{1}{4}$

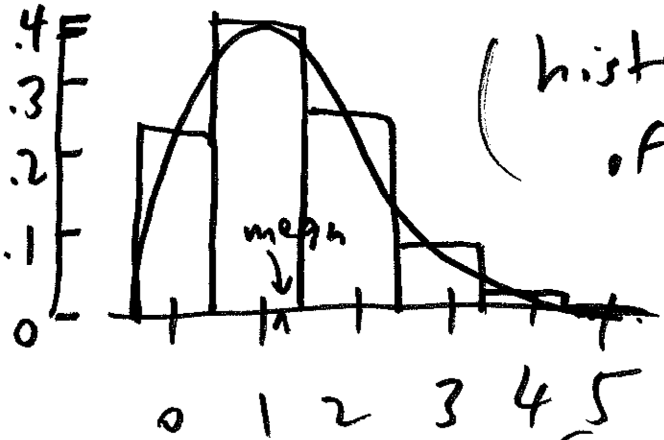
y	$P(Y = y)$
0	$\binom{5}{0}\left(\frac{1}{4}\right)^0 \frac{3}{4}^5 = 0.2373$
1	$\binom{5}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^4 = 0.3955$
2	$\binom{5}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3 = 0.2637$
3	$\binom{5}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2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

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

$\boxed{Q:}$ 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 baby (mode of the dist. of Y)

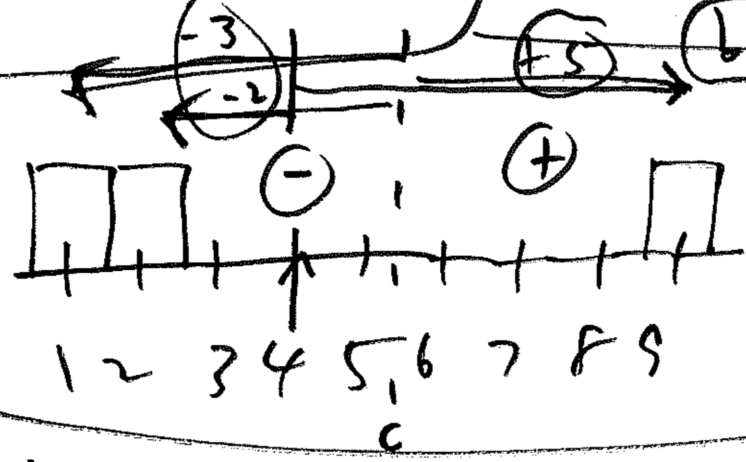


$\boxed{A_2}$ (physics idea)

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

balance point

toy example



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

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 $= 0$

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

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

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

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

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

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

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 PF (or PMF) $f_Y(y) = P(Y = y)$. The **mean** or **expected value** or **expectation** of Y, is $E(Y) \triangleq \sum_{\text{all } y} y\, P(Y=y) = \sum_{\text{all } y} y f_Y(y)$. (167)

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

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

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

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

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

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

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

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

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

$$= np \underbrace{\boxed{\sum_{i=0}^{n-1} \binom{n-1}{i} p^{i} (1-p)^{n-1-i}}}_{\text{Binomial}(n-1,p) \text{ dist.}} \qquad \begin{pmatrix}\text{substitute} \\ i = y-1\end{pmatrix}$$

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

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

$= np$ with $n = 1$

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

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

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

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

$\boxed{\text{Def.}}$ 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

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

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 ($\sim$1320 – 1382).

It's clear from this divergence that (171)
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

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

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

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 parameterize the exponential distribution differently:

Alternative definition

$X \sim \text{Exponential}(\underset{\text{eta}}{\eta})$ $(\eta > 0)$

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

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 $\int_{-\infty}^{0} y\, f_Y(y)\, dy$ and $\int_{0}^{\infty} y\, f_Y(y)\, dy$. If both integrals are infinite, $E(Y)$ is <u>undefined</u> (or <u>does not exist</u>); 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 & 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	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 (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) $X \sim \text{Exponential}(\lambda)$, $(\lambda > 0)$, $E(X) = \frac{1}{\lambda}$

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

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

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

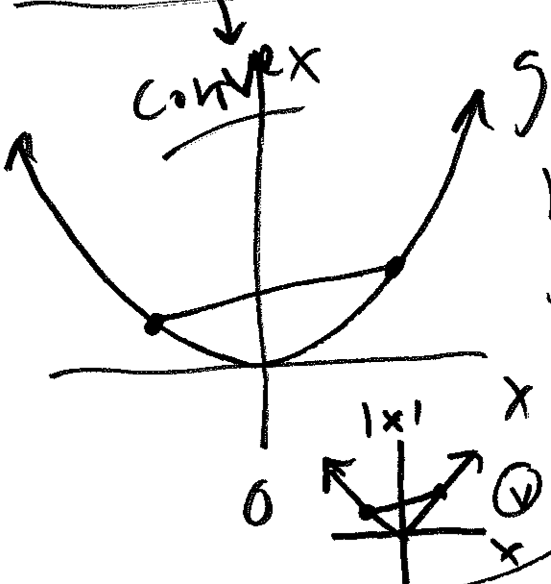
for all constants $(a_1, \dots, a_n)$ and b.

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

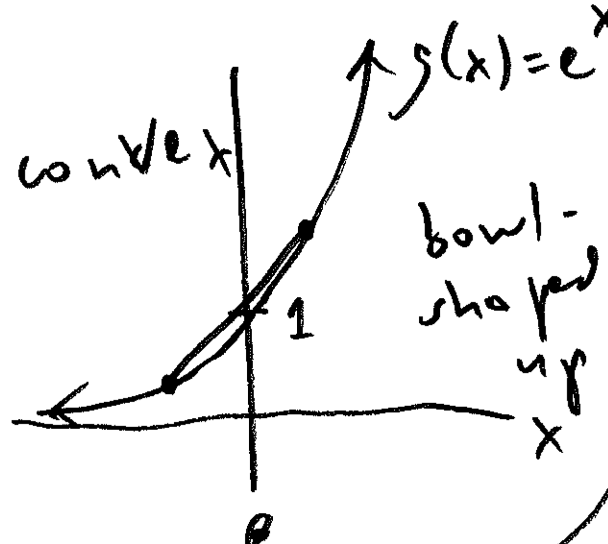
real #s

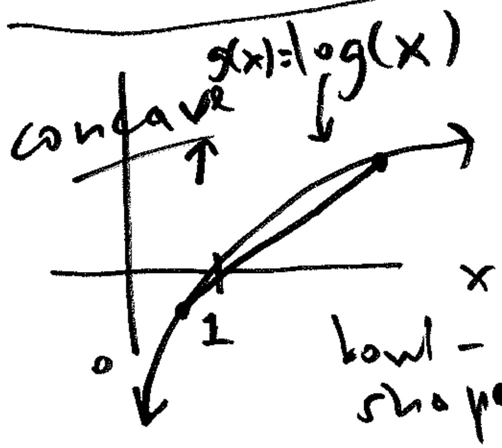
$(x_1, \dots, x_n)$

if for every $0 < \alpha < 1$ and every $\underset{\sim}{x}$ and $\underset{\sim}{y}$, $g[\alpha \underset{\sim}{x} + (1-\alpha)\underset{\sim}{y}] < \alpha\, g(\underset{\sim}{x}) + (1-\alpha)\, g(\underset{\sim}{y})$



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

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

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

③ (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)

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

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

Expectation of a product when the X_i are <u>independent</u>

$X_1, \dots, X_n$ n independent rv, each with finite $E(X_i)$ $\rightarrow$

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

Contrast this with a <u>sum</u>: $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$ whether the X_i are independent or not; $E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$ only when the X_i are <u>independent</u>.

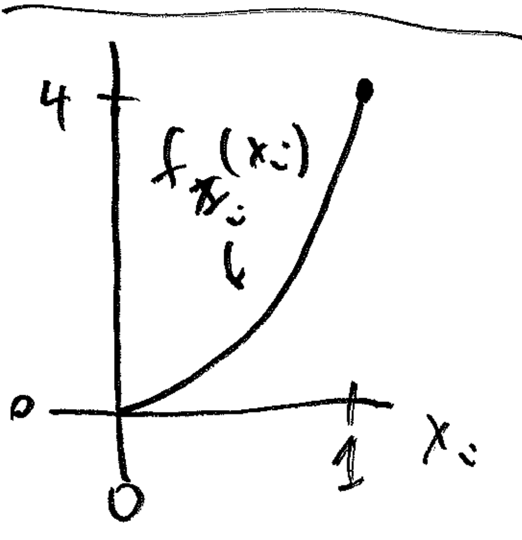
Example You have a (Brita) water filter that you use to improve the taste of Santa Cruz water. How much better would the filter do its job if you filtered the water twice instead of once?

X_1 = proportion of bad stuff removed in the 1st filtering

X_2 = proportion removed in 2nd filtering of what was left from 1st filtering

Reasonable to assume that X_1, X_2 are independent; suppose they're IID with common PDF

$$f_{X_i}(x_i) = \begin{cases} 4x^3 & 0 < x < 1 \\ 0 & \text{else} \end{cases}$$



(sensible shape)

Set Y = proportion of original bad stuff remaining after 2 filtrations $= (1-X_1)(1-X_2)$

Then $E(Y) = E[(1-X_1)(1-X_2)] \overset{\text{independence}}{=} E((1-X_1)) \cdot E((1-X_2))$

X_1, X_2 independent $\leftrightarrow (1-X_1), (1-X_2)$ independent too

$E(1-X_1) \overset{\text{identical distribution}}{=} E(1-X_2) \triangleq \mu$;

then $E(Y) = \mu^2$.

$\mu = E(1-X_i) = \int_0^1 (1-x_i)\, 4x_i^3\, dx_i = 0.2,$ (183)

so 80% of bad stuff expected to be removed in 1^{st} filtering; $E(Y) = \mu^2 = 0.04$, so expect only 4% of bad stuff to remain after 2 filterings.

(6) (a) Suppose X is a discrete rv with possible values 0, 1, 2, ...; then $E(X) = \sum_{n=0}^{\infty} P(X \geq n)$.

(b) If X is a continuous rv with possible values $(0, \infty)$, and CDF $F_X(x)$, then $E(X) = \int_0^{\infty} [1 - F_X(x)]\, dx$.

Example of 6 (a)

I throw a dart at a dart board repeatedly, trying to get a bullseye (success). X = # of throw on which I first succeed.

(Ex. throws $FFS \to X = 3$

F = failure

S = success)

Suppose that my success probability is constant across the throws and equals p, & throws are independent.

Then $E(X)$ should be inversely related to p:

the worse I am, the longer I expect the (1st bullseye) process to take; $E(X) = ?$

At least 1 throw always required so $P(X \geq 1) = 1$; for $n > 1$

$\left(\begin{array}{c}\text{at least } n \\ \text{tosses required}\end{array}\right) \leftrightarrow \left(\begin{array}{c}\text{none of the first } (n-1) \\ \text{throws succeeded}\end{array}\right)$

so $P(X \geq n) = (1-p)^{n-1}$ and

geometric series

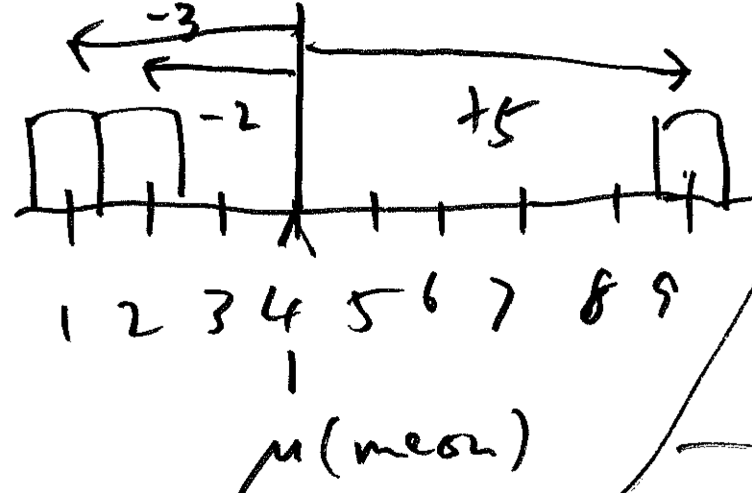
$$E(X) = \sum_{n=1}^{\infty} (1-p)^{n-1} = 1 + (1-p) + (1-p)^2 + \ldots$$

$$= \frac{1}{1-(1-p)} = \frac{1}{p}$$

(inverse relation ✓)

If I'm terrible (e.g. $p = .01$) I expect to succeed on the $\frac{1}{.01} = 100^{\text{th}}$ throw.

Variance and standard deviation



$[\$1, \$2, \$9]$

mean $\$4 = \mu$

X discrete rv, Uniform $\{1, 2, 9\}$; $E(X) = 4 = \mu$

$\boxed{Q:}$ How spread out is the dist. of X around its mean μ?

$(X - \mu) \sim$ Uniform $\{-3, -2, +5\}$

deviation from μ

Could try calculating $E(X - \mu)$, but this is 0 for any rv X, because of cancellation of (+) and (−) deviations; two different easy fixes: $E|X - \mu| \triangleq$ mean average absolute deviation (AAD) (MAD)

(Laplace)

Gauss

or $E(X - \mu)^2 \triangleq$ variance of rv X.

AAD not used much; variance used constantly.

Def X rv with finite mean $E(X) = \mu$;

variance of X = $V(X) \triangleq E\left[(X-\mu)^2\right]$.

DS use $Var(X)$ | If $E(X) = \pm\infty$ or $E(X)$ doesn't exist, $V(X)$ doesn't exist.

One problem with variance | The units are wrong: if X is in \$, $V(X)$ is in $\$^2$.

Easy fix: standard deviation of X $\triangleq \sqrt{V(X)} \triangleq SD(X)$.

Consequences of these definitions

① $V(X) = E\left[(X-\mu)^2\right]$

$= E\left(X^2 - 2\mu X + \mu^2\right)$

$= E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \mu^2$

$= E(X^2) - \mu^2 = E(X^2) - \left[E(X)\right]^2$.

this is a different way to compute the variance so $V(X) = \begin{pmatrix}\text{expectation}\\ \text{of } X^2\end{pmatrix} - \begin{pmatrix}\text{square of}\\ \text{expectation}\\ \text{of } X\end{pmatrix}$.

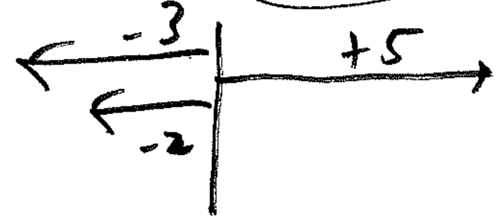
Toy example $\begin{bmatrix}1\\2\\9\end{bmatrix}$ mean $\mu = 4$

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

$$E(X-\mu)^2 = \frac{1}{3}(1-4)^2 + \frac{1}{3}(2-4)^2 + \frac{1}{3}(9-4)^2 \doteq 12.7 = V(X).$$

so $SD(X) \doteq \sqrt{12.7} \doteq 3.6$

This is a reasonable summary of the length of the arrows



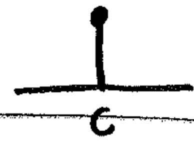
(2) For any rv X, $V(X) \geq 0$; if X is bounded, $V(X)$ exists & is finite.

This is a consequence * of Jensen's Inequality:

$g(x) = x^2$ is convex so $E(X^2) \geq [E(X)]^2$,

i.e. $V(X) = E(X^2) - [E(X)]^2 \geq 0$.

③ $V(X) = 0 \overset{\text{iff}}{\longleftrightarrow} P(X = c) = 1$ for some constant c



(this is a <u>trivial</u> rv)

Notation In the same way that, by convention, $E(X) = \mu_X$, $V(X) \stackrel{\Delta}{=} \sigma_X^2$ and $SD(X) \stackrel{\Delta}{=} \sigma_X$

(lower-case sigma)

④ X rv, $Y = aX + b$

$\rightarrow$ $V(Y) = a^2 V(X) = a^2 \sigma_X^2$ and $SD(Y) = |a| \sigma_X$.

(for any constants a, b)

Special cases $a = 1$: $V(X + c) = V(X)$

$SD(X + c) = SD(X)$

$V(aX) = a^2 V(X)$

$(b = 0)$ $SD(aX) = |a| SD(X)$

⑤ If $X_1, \ldots, X_n$ are <u>independent</u> rv with finite means, $V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$.

This is why the concept of variance has endured even though the units of the variance are wrong: for independent rvs, variance is additive, whereas SD is not.

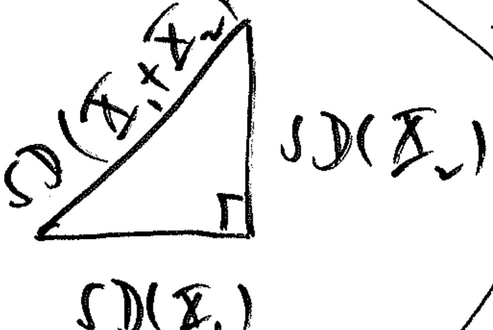
correct units

name of SD: Karl Pearson (1890)

Special case of 5:

X_1, X_2 independent $\rightarrow$ $V(X_1 + X_2) = V(X_1) + V(X_2)$

$= [SD(X_1)]^2 + [SD(X_2)]^2$



So $SD(X_1 + X_2) = \sqrt{[SD(X_1)]^2 + [SD(X_2)]^2}$

i.e., SD grows like the hypotenuse of a right triangle.

Immediately, $\max\left\{ \begin{array}{l} SD(X_1), \\ SD(X_2) \end{array} \right\} < \underset{\text{(indep)}}{SD(X_1 + X_2)} < SD(X_1) + SD(X_2)$