08/16/19

Lecture 9

AMS 131

Expectation, Variance, Covariance, Correlation

Tay-Sachs case Y = # of T-S babies out of 5, parents carriers

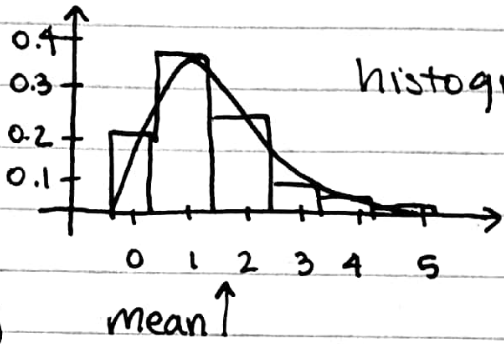
$Y \sim$ Binomial (n, p) $n = 5$, $p = \frac{1}{4}$

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	...	$= 0.2637$
3		$= 0.0879$
4		$= 0.0146$
5	...	$= 0.0010$

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

How many T-S babies should these parents expect to have? (center of dist. of Y)

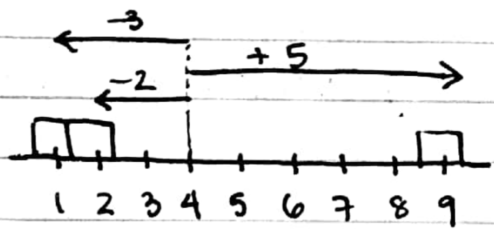
The most likely outcome is 1 T-S baby (the mode of the dist. of Y).



histogram of Y

median: 50% of probability above and below this point

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



"center of mass": balance point
mean = 4, distance to each point balances, $\sum F_i = 0$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \rightarrow \begin{bmatrix} y_1 - c \\ \cdots \\ y_n - c \end{bmatrix} \quad \sum_{i=1}^{n} y_i - c = 0 = \left(\sum_{i=1}^{n} y_i\right) - nc = 0$$

deviations

want the sum to $= 0$ $\quad c$ = center

$$c = \frac{1}{n}\sum_{i=1}^{n} y_i \triangleq \bar{y}$$

sample mean of the dist. of Y (sample)

Here, each value of Y occurred only once

If some values are more probable than others, we should be able to generalize - weighted mean

A R.V. is bounded if all of its possible values are finite

AMS 131

Lecture 9 (cont.)

Let Y be a bounded discrete R.V. with a PMF $f_Y(y) = P(Y=y)$. The mean or expected value or expectation of Y:

$$E(Y) \triangleq \sum_{\text{all } y} y\, P(Y=y) = \sum_{\text{all } y} y f_Y(y)$$

$$E(Y) = (0)(0.2373) + (1)(0.3955) + \ldots + (5)(0.0010) = 1.2500$$

$$= \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y} = \sum_{y=1}^{n} y \binom{n}{y} p^y (1-p)^{n-y}$$ since 0 doesn't matter

↓

$Y \sim \text{Binomial}(n,p)$ for $n > 1$, $E(Y) = np$ $\quad n \geq 1, 0 < p < 1$

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

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

$= 0 \cdot (1-p) + 1(p) = np$ where $n=1$ $= p$

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

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

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

X discrete R.V. with PMF $f_X(x)$, if both $\sum_{x \geq 0} x f_X(x)$ and $\sum_{x<0} x f_X(x)$ are infinite, $E(X)$ is undefined or does not exist. If at least one sum is finite $E(X) = \sum_{\text{all } x} x f_X(x)$ exists (can be infinite)

To create a discrete R.V. whose mean doesn't exist

$\sum_{\text{all } x} f_X(x) = 1$ (finite) but also $\sum_{\text{some } x} x f_X(x) =$ infinite

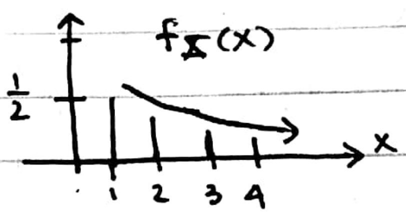
Lecture 9 (cont.) AMS 131

The harmonic series: $\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}\dots\right)=\sum_{x=1}^{\infty}\frac{1}{x}=+\infty$ <u>diverges</u>

Cannot create a R.V. X with PMF $P(X=x)=\frac{c}{x}$
because the probabilities would sum to $+\infty$, but:
$P(X=x)=\frac{c}{x^2}$ or $P(X=x)=\frac{c}{x(x+1)}$ works

$\sum_{x=1}^{\infty}\frac{1}{x^2}=\frac{\pi^2}{6}$ and $\sum_{x=1}^{\infty}\frac{1}{x(x+1)}=1$

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



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

$$f_X(x)=\begin{cases}\frac{1}{2|x|(|x|+1)} & x=\pm1,\pm2\dots\\ 0 & \text{else}\end{cases}\qquad \sum_{\text{all }x}f_X(x)=1$$

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

Expectation for continuous R.V.s
X bounded, continuous with PDF
$$f_X(x)\rightarrow E(X)\triangleq\int_{-\infty}^{\infty}x f_X(x)\,dx$$

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

so $E(X)=\int_{-\infty}^{\infty}\lambda e^{-\lambda x}\,dx=\frac{1}{\lambda}$ integrate by parts

also $f_X(x)=\begin{cases}\frac{1}{\eta}e^{-\frac{x}{\eta}} & x>0\\ 0 & \text{else}\end{cases}$ $\quad X\sim$ Exponential (η) $(\eta>0)$ ✓eta
$\eta=\frac{1}{\lambda}$ $\quad E(X)=\eta$

AMS 131

Lecture 9 (cont.)

If continuous R.V. Y is unbounded with PDF $f_Y(y)$

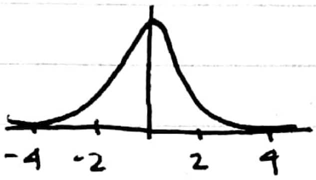
$$\int_{-\infty}^{0} y f_Y(y)\,dy \text{ and } \int_{0}^{\infty} y f_Y(y)\,dy$$ if both are infinite, $E(Y)$ is undefined/ DNE

If at least one integral is finite:

$$E(Y) = \int_{\mathbb{R}} y f_Y(y)\,dy$$ exists, but may be infinite

Cauchy distribution : $f_Y(y) = \dfrac{1}{\pi(1+y^2)}$ $\quad -\infty < y < \infty$

$$\int_{-\infty}^{\infty} \frac{1}{\pi(1+y^2)}\,dy = 1$$



similar to normal, but longer tails

$$\int_{-\infty}^{0} \frac{y}{\pi(1+y^2)}\,dy = -\infty \qquad \int_{0}^{\infty} \frac{1}{\pi(1+y^2)}\,dy = +\infty$$ so $E(Y)$ DNE

the tails go to 0 extremely slow, for large y

$$\frac{y}{1+y^2} \to \frac{1}{y} \text{ and } \int_{c}^{\infty} \frac{1}{y}\,dy = +\infty$$ continous harmonic series, $c > 0$

Expectation of a function of a R.V.

X continuous R.V. PDF $f_X(x)$ $\quad Y \triangleq h(X)$

$E[h(X)] = h[E(X)]$ would be nice, not generally true

1. work out PDF $f_Y(y)$, then

$$E(Y) = \int_{\mathbb{R}} y f_Y(y)\,dy$$ → if this exists

2. $$E(Y) = \int_{\mathbb{R}} h(x) f_X(x)\,dx$$ faster ← Law of the Unconcious Statistician (LOTUS)

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

Lecture 9 (cont.) AMS 131

$X \sim \text{Exponential}(\lambda)$ $\lambda > 0$ $E(X) = \frac{1}{\lambda}$ $Y = X^2$

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

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

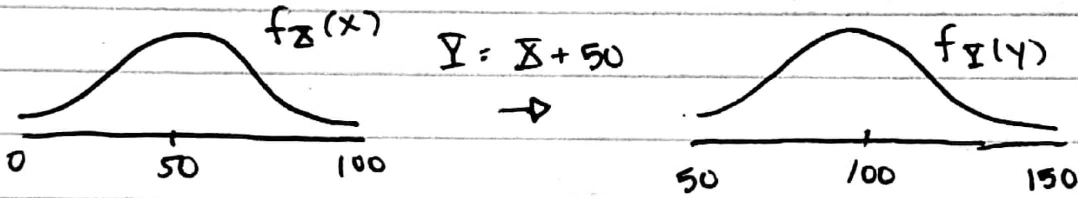
$\frac{2}{\lambda^2} \neq \frac{1}{\lambda^2}$

The only functions $Y = h(X)$ for which $E[h(X)] = h[E(X)]$ are linear: $h(x) = a + bx$.

Properties of $E(Y)$

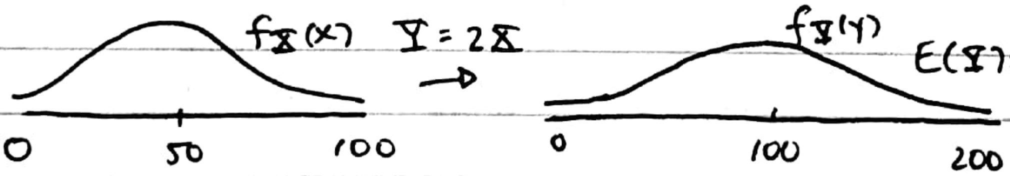
1. If $Y = aX + b$, then $E(Y) = aE(X) + b$ assuming $E(X)$ exists.

2. If you find a constant 'a' with $P(X \geq a) = 1$ then $E(X) \geq a$. If 'b' exists with $P(X \leq b) = 1$ then $E(X) \leq b$



$E(Y) = E(X) + 50$

$E(X + c) = E(X) + c$



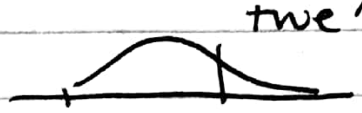
$E(Y) = E(2X)$
$= 2E(X)$

$E(cX) = c \cdot E(X)$

$E(aX + b) = aE(X) + b$

if $Y = aX + b = h(X) \rightarrow E[h(X)] = h[E(X)]$

$E(X) = 3$ not true X



$a = 7$ $E(X) \geq a$

$E(X_1 + X_2) = E(X_1) + E(X_2)$

$E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n)$

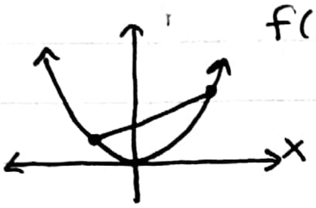
3. If $X_1, \ldots, X_n$, each with finite $E(X_i)$, then

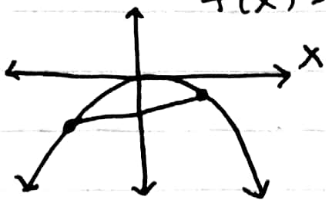
$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

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

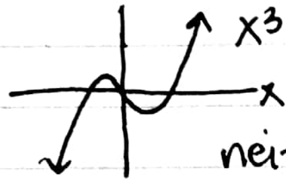
AMS 131 Lecture 9 (cont.)

$f(x) = x^2$ $\qquad$ $f(x) = -x^2$





line always above / below function



x^3

neither

$\frac{d^2}{dx^2} f(x) > 0$ $\qquad$ $\frac{d^2}{dx^2} f(x) < 0$

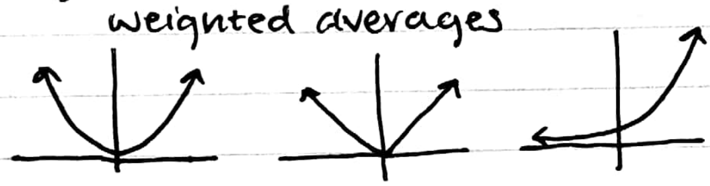
A function $g: \mathbb{R}^n \to \mathbb{R}$, meaning $g(\underset{\sim}{x}) = z$, is convex

- not imaginary value z $\longrightarrow$ real #s $(x_1, \dots, x_n)$

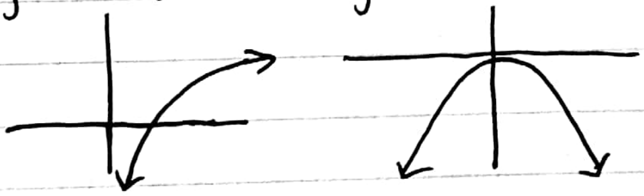
if for every $0 < \alpha < 1$ and every $\underset{\sim}{x}$ and $\underset{\sim}{y}$,

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

weighted averages



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



downwards

random vector: $\underset{\sim}{X} = (X_1, \dots, X_n)$ $\quad E(\underset{\sim}{X}) \triangleq [E(X_1), \dots, E(X_n)]$

$\underset{\sim}{X}$ random vector, finite

g convex : $E(\underset{\sim}{X}) \to E[g(\underset{\sim}{X})] \ge g[E(\underset{\sim}{X})]$ $\quad$ Jensen's

g concave : $\quad E[g(\underset{\sim}{X})] \le g[E(\underset{\sim}{X})]$ $\quad$ inequality

ex: $X_1, \dots, X_n \overset{IID}{\sim}$ Bernoulli (p)

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

'p', 'n' times

$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = n \cdot p =$ mean of Binomial (n,p)

$\uparrow$ Binomial(n,p)

Expectation of a product when X_i are independent

$X_1, \ldots, X_n$ n R.V. with finite $E(X_i)$ $\longrightarrow E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$

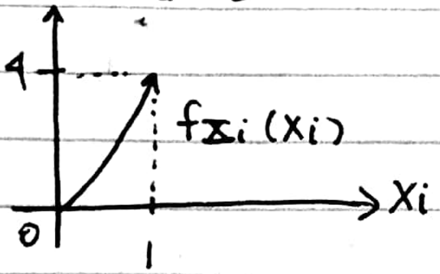
contrasted with: $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$ whether independent or not.

X_1 = proportion of bad stuff removed in first filter

X_2 = proportion removed in second filter, after first filter

X_1, X_2 independent, IID

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



Y = proportion of bad stuff after 2 filters

$= (1 - X_1)(1 - X_2)$

independence (product)

$E(Y) = E[(1-X_1)(1-X_2)] = E[(1-X_1)] \cdot E[(1-X_2)]$

$X_1, X_2 \longleftrightarrow (1-X_1), (1-X_2)$

independent independent

identical distribution

$E(1-X_1) = E(1-X_2) \overset{\triangle}{=} \mu$, then $E(Y) = \mu^2$

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

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

a) Suppose X is a <u>discrete</u> R.V. with possible values $0, 1, 2 \ldots$ then

$$E(X) = \sum_{x=0}^{\infty} P(X \geq x)$$

b.) If X is a <u>continuous</u> R.V. with possible values $(0, \infty)$ then

$$E(X) = \int_0^{\infty} [1 - F_X(x)]\, dx$$

CDF

AMS 131

Lecture 9 (cont.)

a.) Throw a dart until get a bullseye (success)

X = # of throws until first success

FFS   F = failure   success probability is constant,

$X=3$ S = success   throws are independent

$E(X)$ should be inverse to p, higher probability of hitting means you should hit it sooner.

geometric distribution   at least one throw required

$P(X \geq 1) = 1$, for $n > 1$

$$\begin{pmatrix}\text{at least } n \\ \text{tosses required}\end{pmatrix} = \begin{pmatrix}\text{none of the first } (n-1) \\ \text{throws succeeded}\end{pmatrix}$$

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

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

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

$p = 0.01$, succeed on 100TH throw

inverse relationship ✓

$$\begin{bmatrix}1\\2\\9\end{bmatrix} \xrightarrow[\text{(mean)}]{\text{sub. 4}} \begin{bmatrix}-3\\-2\\+5\end{bmatrix} \xrightarrow{\text{abs.}} \begin{bmatrix}+3\\+2\\+5\end{bmatrix}$$

mean = 3.33

MAD, mean absolute deviation

$$\begin{bmatrix}Y_1\\\cdots\\Y_n\end{bmatrix} \xrightarrow{-\bar{Y}} \begin{bmatrix}Y_1-\bar{Y}\\\cdots\\Y_n-\bar{Y}\end{bmatrix} \rightarrow \frac{1}{n}\sum_{i=1}^{n} |(Y_i - \bar{Y})|$$

mean = 0   deviations from mean

$$\xrightarrow{X^2} \begin{bmatrix}(-3)^2\\(-2)^2\\(+5)^2\end{bmatrix} = \begin{bmatrix}9\\4\\25\end{bmatrix}$$

mean = 12.7.

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2} = \text{standard deviation}$$

$= 3.56$

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \text{variance } [\text{data}^2]$$

same scale as data

$$\begin{bmatrix}1\\2\\\cdots\\50\\51\\\cdots\\100\end{bmatrix} = 101 \times 50 = 5050$$

sum (1, 100)

Scanned by CamScanner

Lecture 9 (cont.) AMS 131

X discrete R.V., Uniform$\{1,2,9\}$ $E(X)=4=\mu$ (mean)

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

$E(X-\mu)=0$, so: Laplace 1785

$E|X-\mu| \triangleq$ mean/average absolute deviation (AAD/MAD) not used much

$E(X-\mu)^2 \triangleq$ variance of R.V. X Gauss 1800

X. R.V. with finite mean $E(X)=\mu$

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

if $E(X)=\pm\infty$ or DNE, then $V(X)$ DNE

If X is in \$, $V(X)=\$^2$ wrong units K. Pearson 1890

Standard deviation of $X \triangleq \sqrt{V(X)} \triangleq SD(X)$

$$V(X)=E[(X-\mu)^2]=E[X^2-2\mu X+\mu^2]$$
$$=E(X^2)-2\mu E(X)+\mu^2$$
$$=E(X^2)-\mu^2=E(X^2)-[E(X)]^2$$

$V(X)=$ (expectation of X^2) $-$ (square of expectation of X)

$\begin{bmatrix}1\\2\\9\end{bmatrix}$ $\mu=4$ $X \sim$ Uniform$\{1,2,9\}$

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

$SD(X)=\sqrt{12.7}=3.6$ describes length of arrows

$V(X)=E(X-\mu)^2 \geq 0$, equality iff $X=\mu$, $p=1$

$a<\infty$ $V(X)<\infty$ with bounded. X.

$b<\infty$ finite

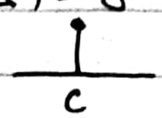
$V(X)=E(X^2)-[E(X)]^2 \geq 0$ so $E(X^2)\geq[E(X)]^2$

convex

$g(x)=x^2$

when convex $E[g(X)] \geq g[E(X)]$

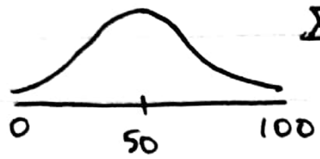
$V(X)=0 \overset{\text{iff}}{\longleftrightarrow} P(X=c)=1$ for some constant c



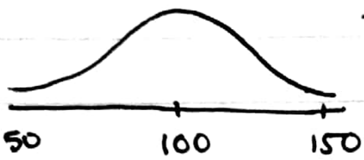
: trivial R.V.

AMS 131 Lecture 9 (cont.)

mean $= E(X) = \mu_X$, $V(X) \triangleq \sigma_X^2$, $SD(X) \triangleq \sigma_X$



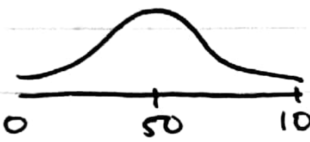
X SD σ_X shifting does not change scale / spread

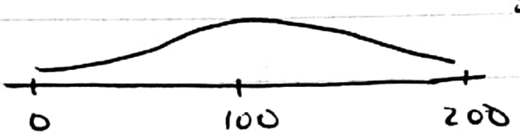


$X + 50 = Y$ $SD(Y) = \sigma_Y = \sigma_X$

$$\left.\begin{aligned} V(X+c) &= V(X) \\ SD(X+c) &= SD(X) \end{aligned}\right\} \text{does not change}$$

$Y = 2X$





$$2SD(X) = SD(Y)$$

$$V(cX) = c^2 V(X)$$

$$SD(cX) = |c|\, SD(X)$$

$$V(aX+b) = V(aX) = a^2 V(X) = a^2 SD(X)^2$$

$$SD(aX+b) = |a|\, SD(X)$$

$a=1$, $V(X+c) = V(X)$ $SD(X+c) = SD(X)$

$b=0$, $V(aX) = a^2 V(X)$ $SD(aX) = |a|\, SD(X)$

if $X_1, \dots, X_n$ are independent R.V. with finite means $V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$

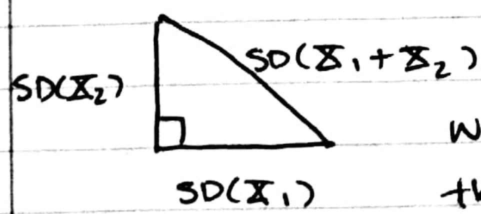
$$V(X_1 + X_2) = V(X_1) + V(X_2)$$

$$SD(X_1 + X_2) \neq SD(X_1) + SD(X_2)$$

$$SD(X_1 + X_2) = \sqrt{V(X_1 + X_2)}$$

indep. $= \sqrt{V(X_1) + V(X_2)}$

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



we use variance (even though the units are wrong) because of the additive property, SD is not