Chapter 2: Probability for Statistical Computing

We will **briefly** review some definitions and concepts in probability and statistics that will be helpful for the remainder of the class.

Just like we reviewed computational tools (`R` and packages), we will now do the same for probability and statistics.

Note: This is not meant to be comprehensive. I am assuming you already know this and maybe have forgotten a few things.

i.e. you may need to look some things up on your own.

https://xkcd.com/892/

Alternative text: "Hell, my eighth grade science class managed to conclusively reject it just based on a classroom experiment. It's pretty sad to hear about million-dollar research teams who can't even manage that."

1 Random Variables and Probability

Definition 1.1 A *random variable* is a function that maps sets of all possible outcomes of an experiment (sample space Ω) to $\mathbb{R}$.

"real numbers
$(-\infty, \infty)$

Example 1.1

experiment Toss 2 dice.

$\Omega = \{(i,j) : i = 1,..,6;\ j = 1,..,6\}$

r.v. X = sum of the dice.

Example 1.2

experiment: Randomly select 25 deer & test for CWD

r.v. X_i: $\{0 \text{ or } 1\}$ observe $X_1, ..., X_{25}$

$\Omega = \{+, - \text{ CWD}\}$

r.v. $p = \sum_{i=1}^{25} X_i / 25$ is also a r.v.!

Example 1.3

experiment: Deck of cards, draw one card

r.v. X: 1 if clubs, 0 otherwise.

Ω = {values of all 52 cards in a deck}
= {AC, 2C, 3C, ..., KC,
AS, 2S, ..., KS,
AD, 2D, ..., KD,
AH, 2H, ..., KH}

Ex. 1.4
Today's high temp $= X_i$

Types of random variables –

Discrete take values in a countable set.

Ex. 1.1
X_i from Ex 1.2, X from X 1.3

Continuous take values in an uncountable set (like $\mathbb{R}$)

Ex 1.4 $X_i \in \mathbb{R}$

p from Ex 1.2, $p \in [0, 1]$

1.1 Distribution and Density Functions

Definition 1.2 The *probability mass function (pmf)* of a random variable X is f_X defined by

Notation: sometimes when the r.v. is obvious I will omit the subscript and write $f(x)$.

$$f_X(x) = P(X = x) \quad \text{for any } x \in \mathbb{R}$$

where $P(\cdot)$ denotes the probability of its argument.

There are a few requirements of a **valid** pmf

1. $f(x) \geq 0$ for all $x \in \mathbb{R}$.
2. $\sum_x f(x) = 1$
3. We call $\mathcal{X} = \{x : f(x) > 0\}$ the "support" of X. Not a requirement

Example 1.4 Let Ω = all possible values of a roll of a single die $= \{1, \ldots, 6\}$ and X be the outcome of a single roll of one die $\in \{1, \ldots, 6\}$.

Fair die

$f(1) = P(X=1) = \frac{1}{6}$

$\vdots$

$f(6) = \frac{1}{6}$

$\sum_{x \in \mathcal{X}} f(x) = \sum_{x=1}^{6} \frac{1}{6} = 1$ ✓ valid pmf.

A pmf is defined for **discrete variables**, but what about **continuous**? Continuous variables do not have positive probability mass at any single point.

Definition 1.3 The *probability density function (pdf)* of a random variable X is f_X defined by

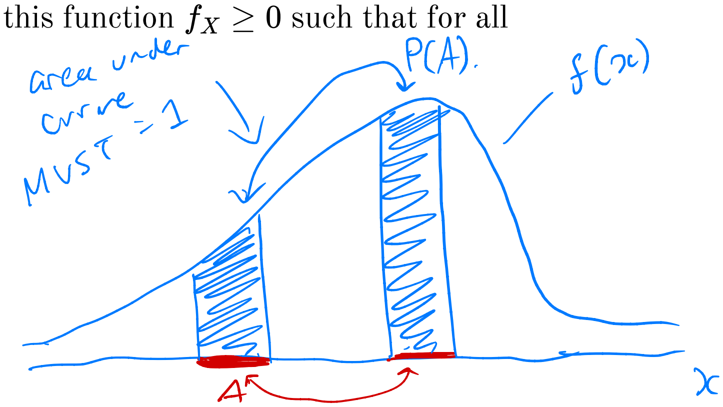
$A \subset \mathbb{R}$

$$P(X \in A) = \int_{x \in A} f_X(x) dx.$$

X is a continuous random variable if there exists this function $f_X \geq 0$ such that for all $x \in \mathbb{R}$, this probability exists.

For f_X to be a valid pdf,

1. $f(x) \geq 0 \quad \forall x \in \mathbb{R}$
2. $\int_{\mathbb{R}} f(x) dx = 1$



Again $\mathcal{X} = \{x : f(x) > 0\}$ is the "support" of X

There are many named pdfs and ~~cdfs~~ pmfs that you have seen in other class, e.g.

Bernoulli (pmf), Poisson (pmf), Gamma (pdf), Normal (pdf), Beta, exponential, hypergeometric.

Example 1.5 Let

$$f(x) = \begin{cases} c(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

← support

Find c and then find $P(X > 1)$

↳ such that $f(x)$ is a valid pdf.

$$1 = \int_0^2 c(4x - 2x^2)\,dx = c\left[2x^2 - \frac{2x^3}{3}\right]_0^2 = c\left[\frac{8}{3}\right] \Rightarrow c = \frac{3}{8}$$

"normalizing constant"

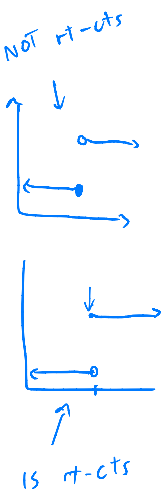
$$P(X > 1) = \int_1^\infty f(x)\,dx = \int_1^2 \frac{3}{8}(4x - 2x^2)\,dx = \frac{3}{8}\left[2x^2 - \frac{2x^3}{3}\right]_1^2 = \frac{1}{2}$$

Definition 1.4 The *cumulative distribution function (cdf)* for a random variable X is F_X defined by

cdf function of $x \in \mathbb{R}$

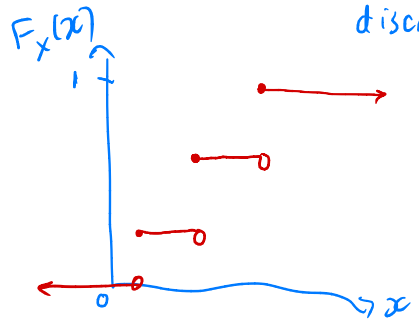
$$F_X(x) = P(X \le x), \quad x \in \mathbb{R}.$$

↳ same def'n for both cts and discrete r.v.'s



The cdf has the following properties

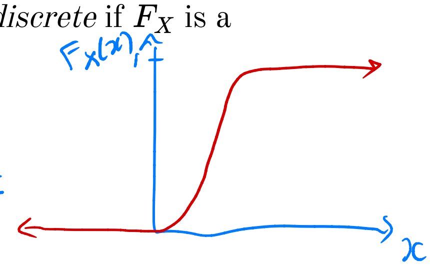
1. F_X is non-decreasing
2. F_X is right-continuous
3. $\lim_{x \to -\infty} F_X(x) = 0$ $\quad \lim_{x \to \infty} F_X(x) = 1$



"step function"

X is discrete

A random variable X is *continuous* if F_X is a continuous function and *discrete* if F_X is a step function.



Example 1.6 Find the cdf for the previous example.

$F_X(x) = P(X \le x)$

for $x \in (0,2)$, $P(X \le x) = \int_0^x \frac{3}{8}(4y - 2y^2)\,dy = \left[\frac{3}{8}\left(2y^2 - \frac{2y^3}{3}\right)\right]_0^x = \frac{3}{4}x^2\left(1 - \frac{x}{3}\right)$

So, $F_X(x) = \begin{cases} 0 & x \le 0 \\ \frac{3}{4}x^2\left(1 - \frac{x}{3}\right) & x \in (0,2) \\ 1 & x \ge 2 \end{cases}$

Note $f(x) = F'(x) = \frac{dF(x)}{dx}$ in the continuous case.

pdf $\qquad$ derivative of cdf

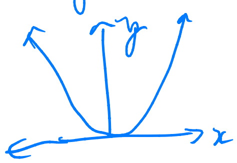
due to the Fundamental Th'm of Calculus.

Recall an indicator function is defined as

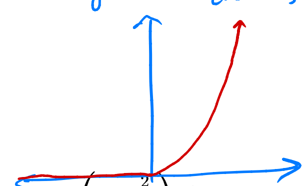
$$1_{\{A\}} = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{otherwise} \end{cases}.$$

condition

Example 1.7 $y = x^2$



$y = x^2 1\{x>0\}$

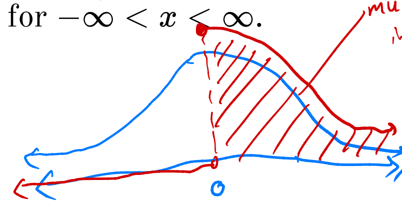


your TURN

Example 1.8 If $X \sim N(0,1)$, the pdf is $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ for $-\infty < x < \infty$.

If $f(x) = \frac{c}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \boxed{1_{\{x>0\}}}$, what is c?

must integrate to 1.



We have symmetry of $N(0,1)$ around 0.

$$\Rightarrow \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \frac{1}{2}$$

$$1 = c \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \frac{c}{2} \Rightarrow c = 2.$$

need this to hold.

1.2 Two Continuous Random Variables

2 r.v.'s that we care about together and are both cts.

Definition 1.5 The *joint pdf* of the continuous vector (X, Y) is defined as

$$P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy$$

joint pdf

for any set $A \subset \mathbb{R}^2$.

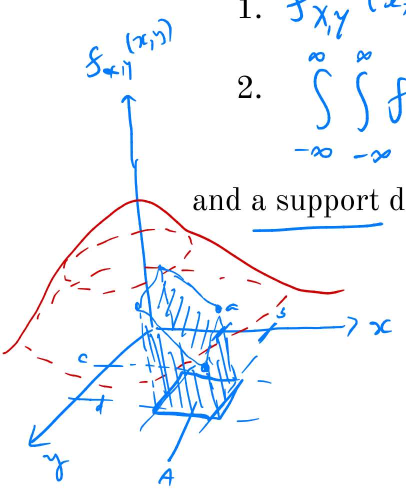
Joint pdfs have the following properties

1. $f_{X,Y}(x,y) \geq 0 \quad \forall\, x, y \in \mathbb{R}^2$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x,y)\,dx\,dy = 1$

Note: We can also have jointly discrete r.v.'s w/ corresponding joint pmfs where

$$\sum\sum f_{X,Y}(x,y) = 1.$$

and a support defined to be $\{(x,y) : f_{X,Y}(x,y) > 0\}$. $= \mathcal{X}$



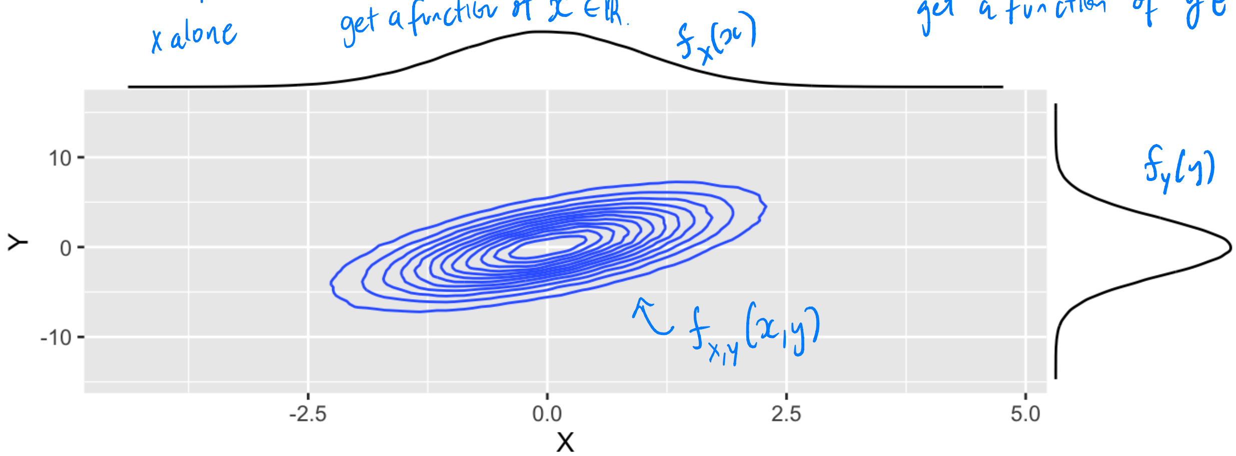
example A could be rectangle $x \in [a,b]$, $y \in [c,d]$

Example 1.9

The *marginal densities* of X and Y are given by

$$f_X(x) = \int_{\infty}^{\infty} f_{X,Y}(x,y)dy \qquad \text{and} \qquad f_Y(y) = \int_{\infty}^{\infty} f_{X,Y}(x,y)dx;$$

x alone   get a function of $x \in \mathbb{R}$.   get a function of $y \in \mathbb{R}$.



Example 1.10 (From Devore (2008) Example 5.3, pg. 187) A bank operates both a drive-up facility and a walk-up window. On a randomly selected day, let X be the proportion of time that the drive-up facility is in use and Y is the proportion of time that the walk-up window is in use. X Y

random vector

The the set of possible values for (X, Y) is the square $D = \{(x, y) : 0 \le x \le 1, 0 \le y \le 1\}$. Suppose the joint pdf is given by

$$f_{X,Y}(x,y) = \begin{cases} \frac{6}{5}(x+y^2) & x \in [0,1], y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

support
$\{(x,y) : 0 \le x \le 1, 0 \le y \le 1,$
not both $x, y = 0\}$

Evaluate the probability that both the drive-up and the walk-up windows are used a quarter of the time or less.

$$P(0 \le X \le \tfrac{1}{4}, 0 \le Y \le \tfrac{1}{4}) = \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x+y^2)\,dx\,dy$$

$$= \int_0^{1/4} \frac{6}{5}\left[\frac{x^2}{2} + xy^2\right]_{x=0}^{1/4} dy$$

$$= \int_0^{1/4} \frac{6}{5}\left(\frac{1}{32} + \frac{y^2}{4}\right) dy$$

$$= \left[\frac{6}{5}\left(\frac{y}{32} + \frac{y^3}{12}\right)\right]_0^{1/4} = \frac{6}{5}\left(\frac{1}{32}\cdot\frac{1}{4} + \frac{1}{12}\left(\frac{1}{4}\right)^3\right) = \frac{7}{640}$$

$$= 0.0109$$

Find the marginal densities for X and Y.

$$f_X(x) = \int_0^1 \frac{6}{5}(x+y^2)\,dy = \frac{6}{5}\left[xy + \frac{y^3}{3}\right]_{y=0}^{1} = \begin{cases} \frac{6}{5}\left(x+\frac{1}{3}\right) & \text{for } x \in [0,1] \\ 0 & \text{o.w.} \end{cases}$$

$f_Y(y)$

Leave up to you.

Compute the probability that the drive-up facility is used a quarter of the time or less.

$$P\left(X \le \frac{1}{4}\right) = \int_0^{1/4} f_X(x)\,dx = \int_0^{1/4} \frac{6}{5}\left(x+\frac{1}{3}\right)dx$$

$$= \frac{6}{5}\left[\frac{x^2}{2} + \frac{x}{3}\right]_0^{1/4}$$

$$= \frac{11}{80} = 0.1375$$

2 Expected Value and Variance

Definition 2.1 The *expected value* (average or mean) of a random variable X with pdf or pmf f_X is defined as

$$E[X] = \begin{cases} \sum\limits_{x \in \mathcal{X}} x f_X(x_i) & X \text{ is discrete} \\ \int\limits_{x \in \mathcal{X}} x f_X(x) dx & X \text{ is continuous.} \end{cases}$$

Where $\mathcal{X} = \{x : f_X(x) > 0\}$ is the support of X.

This is a weighted average of all possible values $\mathcal{X}$ by the probability distribution.

Example 2.1 Let $X \sim \text{Bernoulli}(p)$. Find $E[X]$.

X is the r.v., pmf is $f(x)$, support is $\{0,1\}$,

$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{o.w.} \end{cases} \Rightarrow f(x) = \begin{cases} p & \text{when } x=1 \\ 1-p & \text{when } x=0 \end{cases}$ or $f(x) = p^x(1-p)^{1-x}$ for $x \in \{0,1\}$.

parameter is p.

$$EX = \sum_{x \in \{0,1\}} x f(x) = 0(1-p) + 1(p) = p.$$

Example 2.2 Let $X \sim \text{Exp}(\lambda)$. Find $E[X]$.

cts

$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0. \end{cases}$

HW 1

integration by parts

$$EX = \int_0^\infty x \lambda e^{-\lambda x} dx = \dots = \frac{1}{\lambda}$$

Definition 2.2 Let $g(X)$ be a function of a continuous random variable X with pdf f_X. Then,

$$E[g(X)] = \int_{x \in \mathcal{X}} g(x) f_X(x) dx.$$

sometimes this is hard to compute analytically

$\Rightarrow$ we will need stat computing to estimate (Ch. 5)

Definition 2.3 The *variance* (a measure of spread) is defined as

$$\begin{aligned} Var[X] &= E\left[(X - E[X])^2\right] \\ &= E[X^2] - (E[X])^2 \end{aligned}$$

computational formula

$g(x) = x^2$

Example 2.3 Let X be the number of cylinders in a car engine. The following is the pmf function for the size of car engines.

x	4.0	6.0	8.0
f	0.5	0.3	0.2

i.e. $P(X=4) = 0.5$, etc.

Who might care about this?

Find

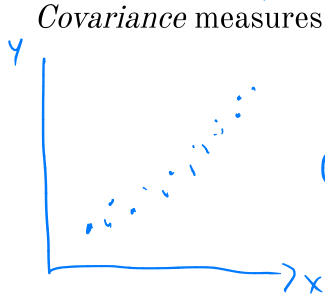
$E[X] = \sum_{x \in \mathcal{X}} x f(x) = 4 \cdot 0.5 + 6 \cdot 0.3 + 8 \cdot 0.2 = 5.4$

$Var[X] = E(X^2) - (EX)^2$

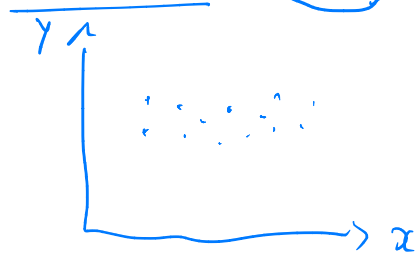
$E(X^2) = \sum_{x \in \mathcal{X}} x^2 f(x) = 4^2(0.5) + 6^2(0.3) + 8^2(0.2) = 31.6$

$\Rightarrow Var\, X = 31.6 - (5.4)^2 = 2.44$ easier to interpret is $sd(X) = \sqrt{Var(X)} = 1.56$

Covariance measures how two random variables vary together (their linear relationship).

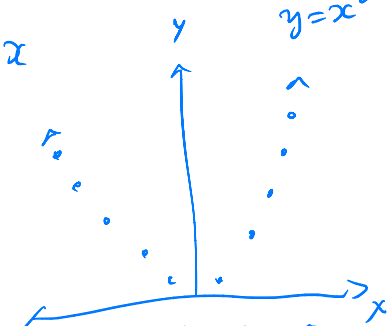


$Cov(X,Y) \geq 0$



"random noise"

$Cov[X,Y] \approx 0$



$Cov(X,Y) = 0$

Definition 2.4 The *covariance* of X and Y is defined by

$$\begin{aligned} Cov[X,Y] &= E\left[(X - E[X])(Y - E[Y])\right] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

↑ computational form

and the *correlation* of X and Y is defined as

$$\rho(X,Y) = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}}.$$

$\rho \in [-1, 1]$

Two variables X and Y are *uncorrelated* if $\rho(X,Y) = 0$.

3 Independence and Conditional Probability

In classical probability, the *conditional probability* of an event A given that event B has occured is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

$P(A|B)P(B) = P(A \cap B)$

Definition 3.1 Two events A and B are *independent* if $P(A|B) = P(A)$. The converse is also true, so

$$A \text{ and } B \text{ are independent} \Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$$

Theorem 3.1 (Bayes' Theorem) Let A and B be events. Then,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

3.1 Random variables

The same ideas hold for random variables. If X and Y have joint pdf $f_{X,Y}(x, y)$, then the conditional density of X given $Y = y$ is

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Thus, two random variables X and Y are independent if and only if

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

Also, if X and Y are independent, then

indep.

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_X(x)\cancel{f_Y(y)}}{\cancel{f_Y(y)}} = f_X(x).$$

def'n

4 Properties of Expected Value and Variance

Suppose that X and Y are random variables, and a and b are constants. Then the following hold:

1. $E[aX + b] =$ $a\,EX + b$

2. $E[X + Y] =$ $EX + EY$

3. If X and Y are independent, then $E[XY] =$ $EX\,EY$ only true with independence.

4. $Var[b] =$ 0

5. $Var[aX + b] =$ $a^2\,Var\,X$

6. If X and Y are independent, $Var[X + Y] =$ $Var\,X + Var\,Y$

5 Random Samples

Definition 5.1 Random variables $\{X_1, \ldots, X_n\}$ are defined as a *random sample* from f_X if $X_1, \ldots, X_n \overset{iid}{\sim} f_X$.

independent and identically distributed.

Example 5.1

$X_1, \ldots, X_n \overset{iid}{\sim} N(0, \sigma^2)$

random sample

vs.

$X_1 \sim N(\mu_1, \sigma^2)$

$X_2 \sim N(\mu_2, \sigma^2)$

these may be independent, but not identically distributed.

not a random sample (unless indep + $\mu_1 = \mu_2$).

Theorem 5.1 If $X_1, \ldots, X_n \overset{iid}{\sim} f_X$, then

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i).$$

joint pdf

product of marginals, easier to deal with.

Example 5.2 Let $X_1, \ldots, X_n$ be iid. Derive the expected value and variance of the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$$E(\bar{X}_n) \overset{\text{def'n of mean}}{=} E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \overset{\text{property 1}}{=} \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) \overset{\text{property 2}}{=} \frac{1}{n}\sum_{i=1}^{n} EX_i \overset{x_1,\ldots,x_n \text{ iid } EX_1=\cdots=EX_n}{=} \frac{1}{n}\sum_{i=1}^{n} EX_1 = \frac{1}{\cancel{n}}\cancel{n}EX_1$$

$$= EX_1$$

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var X_i = \frac{1}{n^2}\cdot n\, Var X_1 = \frac{1}{n} Var X_1$$

property 5

X_i independent + property 6

x_i iid $\Rightarrow$ $Var X_1 = \cdots = Var X_n$

6 R Tips

From here on in the course we will be dealing with a lot of **randomness**. In other words, running our code will return a **random** result.

> But what about reproducibility??

When we generate "random" numbers in R, we are actually generating numbers that *look* random, but are *pseudo-random* (not really random). The vast majority of computer languages operate this way.

This means all is not lost for reproducibility!

```
set.seed(400)
```

set the starting point.

Before running our code, we can fix the starting point (seed) of the pseudorandom number generator so that we can reproduce results.

Speaking of generating numbers, we can generate numbers (also evaluate densities, distribution functions, and quantile functions) from named distributions in R.

what we want

```
rnorm(100)
dnorm(x)
pnorm(x)
qnorm(y)
```

random number

evaluates the density

evaluates cdf

quantiles.

what distribution