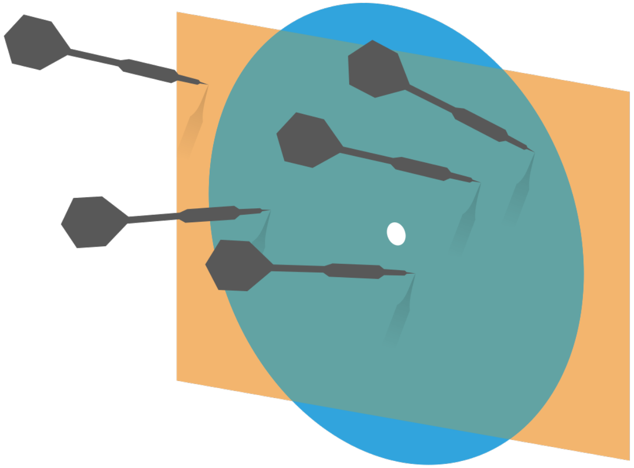
# Probability and Statistics for Computer Science



Credit: wikipedia

"The weak law of large numbers gives us a very valuable way of thinking about expectations." ---Prof. Forsythe

Hongye Liu, Teaching Assistant Prof, CS361, UIUC, 09.22.2020

Midterm exam 1 is on Oct. 8

Please schedule CBTF exam

2 Practice exams are (linked to Problems & Solutions

One practice exam will be given in a week through Gradescope to mimic CBTF exam protocol.

# How many possible colors ?

Hex Color codes uses Hexadecimal (16 per position) format to define colors. With 9 Hex-digits, how many colors can be represented?

$16 \times 16 \times \ldots 16$

$16^9$

$2^{4 \times 9}$

# Last time

Random Variable (R.V.)

Definition

properties

$X(\omega)$

* Expected value

$f(X)$

* Variance & Covariance

$f(X, Y)$

# Objectives

Random variable (R.V.)

* Review of expectations
* Markov's Inequality
  Chebyshev's Inequality
* The weak law of large numbers

# Expected value

- The **expected value** (or **expectation**) of a random variable $X$ is

$$E[X] = \sum_x xP(x)$$

$P(X=x) \geq 0$

The expected value is a weighted sum of **all** the values $X$ can take

# Linearity of Expectation

$$E[aX+b] = aE[X]+b$$

$$E[X+Y] = E[X]+E[Y]$$

$$E\left[\sum_i X_i\right] = \sum_i E[X_i]$$

# Expected value of a function of $X$

$$E[f(X)] = \sum_{x} f(x) \cdot P(x)$$

# Probability distribution

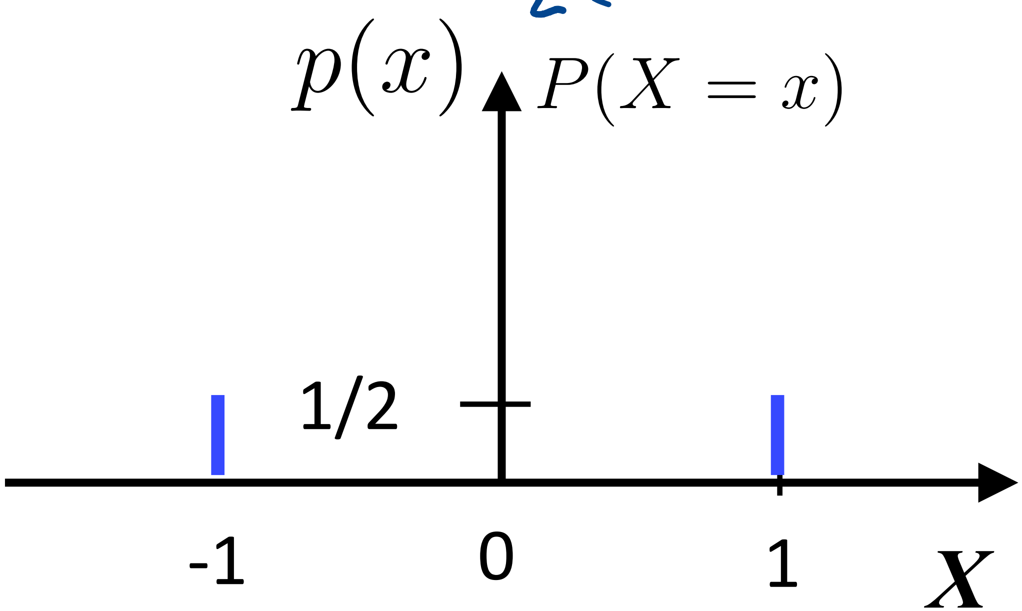
- Given the random variable ***X***, what is

E[2|***X***| +1]?

$E[2|X|+1] = 2E[|X|]+1 = 3$

$\sum (2|X|+1)p(x)$

$E[|X|] = \sum_x |x| \cdot p(x) = 1\times\frac{1}{2} + 1\times\frac{1}{2} = 1$




A. 0

B. 1

C. 2

$\boxed{\text{D.}}$ 3

E. 5

# Expected time of cat

- A cat moves with random constant speed **V**, either 5mile/hr or 20mile/hr with equal probability, what's the expected time for it to travel 50 miles?

$$T = \frac{D}{V} = f(V)$$

$$E[T] = \sum_{V} f(v)P(v) \cancel{\neq \frac{D}{E[V]}}$$

$$= \frac{D}{V_1} \cdot P(V_1) + \frac{D}{V_2} \cdot P(V_2)$$

$$= \frac{50}{5} \times \frac{1}{2} + \frac{50}{20} \times \frac{1}{2} = 6.25$$

Jensen's inequality
for convex func. $g(x)$

$$E[g(x)] \geq g(E[x])$$

---

Can't assume
$$E[g(x)] = g(E[X])!$$

# A neater expression for variance

- Variance of Random Variable X is defined as:

$$var[X] = E[(X - E[X])^2]$$

$var[kX] = \underline{?}$ $k^2 var[X]$

- It's the same as:

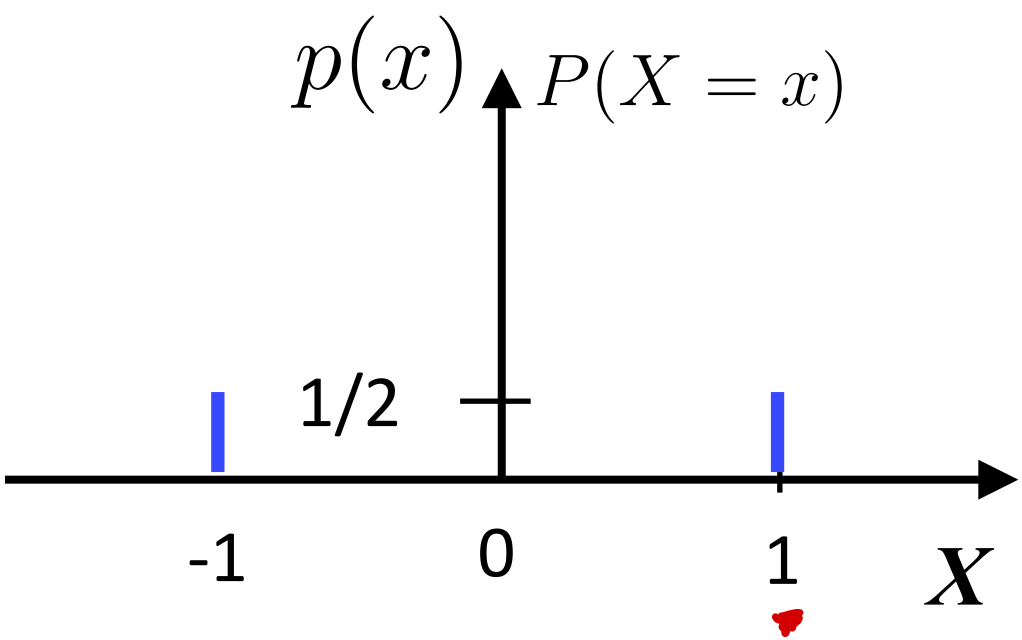
$$\boxed{var[X] = E[X^2] - E[X]^2}$$

# Probability distribution and cumulative distribution

- Given the random variable ***X***, what is

var[2|***X***| +1]?

$= 2^2 \, var[|x|]$




A. 0

B. 1

C. 2

D. 3

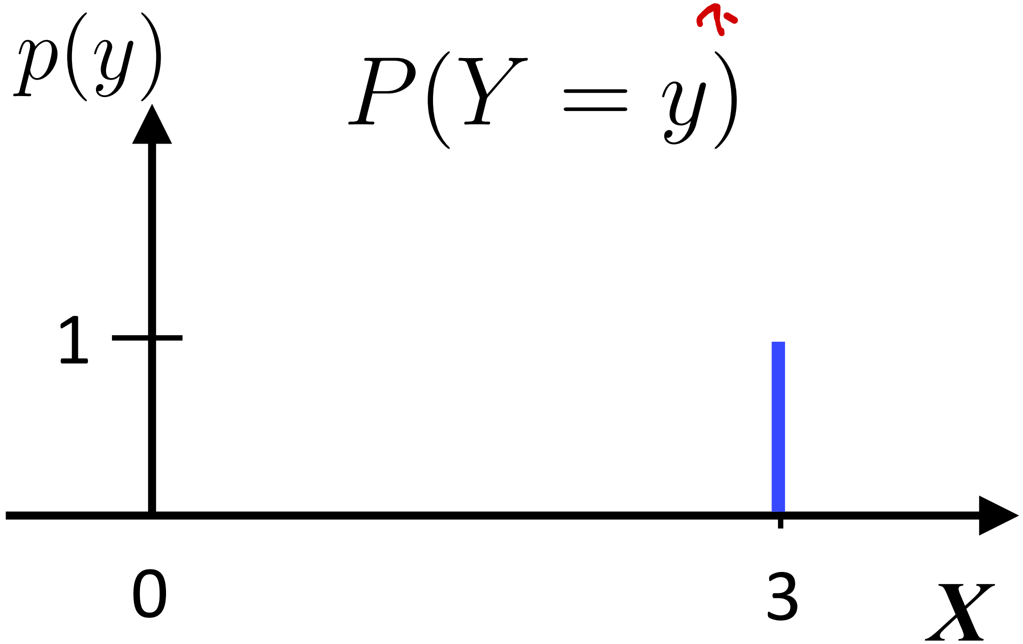
E. -1

# Probability distribution

- Given the random variable ***X***, what is

var[2|***X***| +1]? Let ***Y*** = 2|***X***|+1

$$P(Y = y)$$



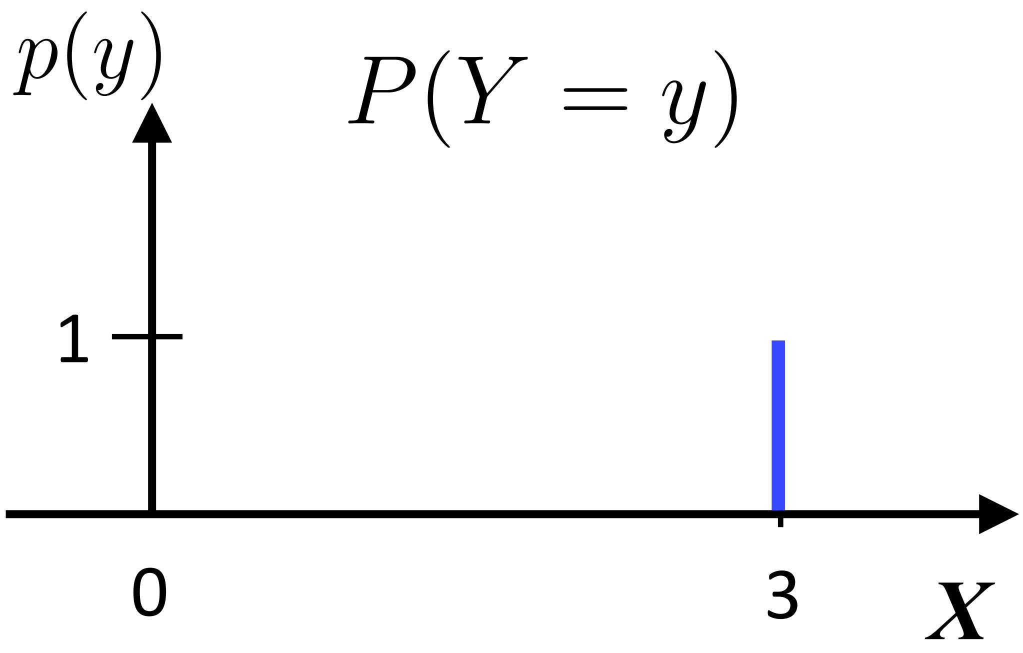

$\mathrm{var}[|X|]$

$= E[|X|^2] - E^2[|X|]$

# Probability distribution

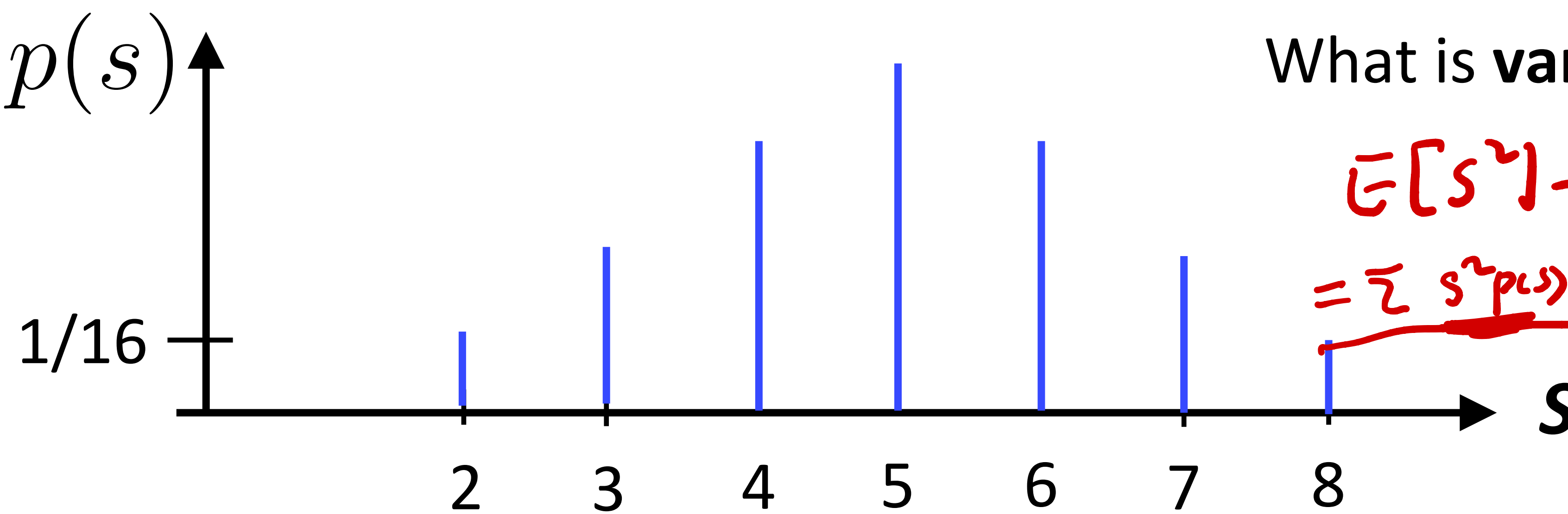
- Given the random variable ***X***, what is

var[2|***X***| +1]? Let ***Y*** = 2|***X***|+1



# Probability distribution

- Give the random variable ***S*** *in the 4-sided die, whose range is* {2,3,4,5,6,7,8}, probability distribution of ***S*.**




What is **var[S]** ?

$= E[S^2] - E^2(S)$

$= \sum s^2 p(s) - E^2(S)$

These are equivalent:

(I) $\operatorname{cov}(X, Y) = 0$; $\operatorname{corr}(X, Y) = 0$

(II) $E[XY] = E[X]E[Y]$

(III) $\operatorname{var}[X+Y] = \operatorname{var}[X] + \operatorname{var}[Y]$

they all mean $X, Y$ are uncorrelated.

$$\operatorname{cov}(X, Y) = E[XY] - E[X]E[Y]$$

$$\operatorname{var}[X+Y] = \operatorname{var}[X] + \operatorname{var}[Y] + 2\operatorname{cov}(x, y)$$

# Properties of independence in terms of expectations

- $$E[XY] = E[X]E[Y]$$ if X, Y are indpt.

Proof: $LHS = \sum_x \sum_y xy P(x,y)$

if X, Y are indpt.

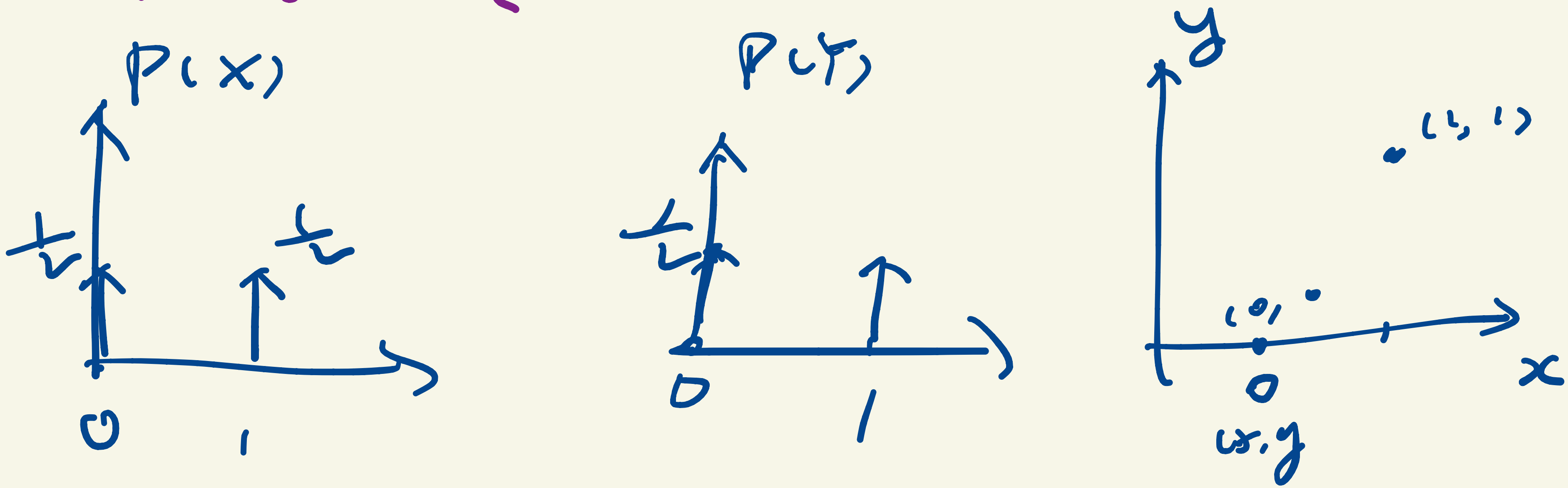
$p(x,y) = P(x) P(y)$ for all $x, y$

$LHS = \sum_x x P(x) \sum_y y P(y)$

$= E[X] E[Y] = RHS$

If $X, Y$ are independent

then
$$\begin{cases} Cov(X,Y)=0;\ corr(X,Y)=0 \\ E[XY]=E[X]E[Y] \\ var[X+Y]=var[X]+var[Y] \end{cases}$$



Work on it offline

# Q: What is this expectation?

- We toss two identical coins A & B independently for three times and 4 times respectively, for each head we earn \$1, we define $X$ is the earning from A and $Y$ is the earning from B. What is E[$XY$]?

$E[X] = ?$

$E[Y] = ?$

A. \$2 B. \$3 C. \$4

Work on it offline

# Uncorrelated vs Independent

- If two random variables are uncorrelated, does this mean they are independent? Investigate the case X takes -1, 0, 1 with equal probability and $Y=X^2$.

$E[XY] = E[X]E[Y]$

but $p(x,y) \neq p(x)p(y)$

# How do you make it with a die?

as if
throwing a biased coin
with probability = 0.75 coming up head?

4-sided die

0.75

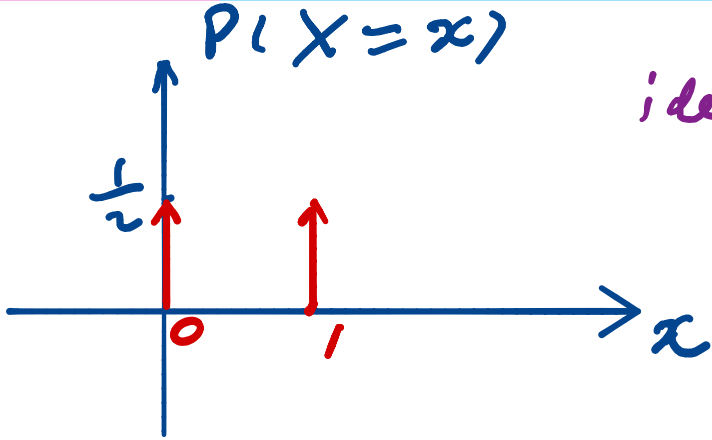
1, 2, 3 → H

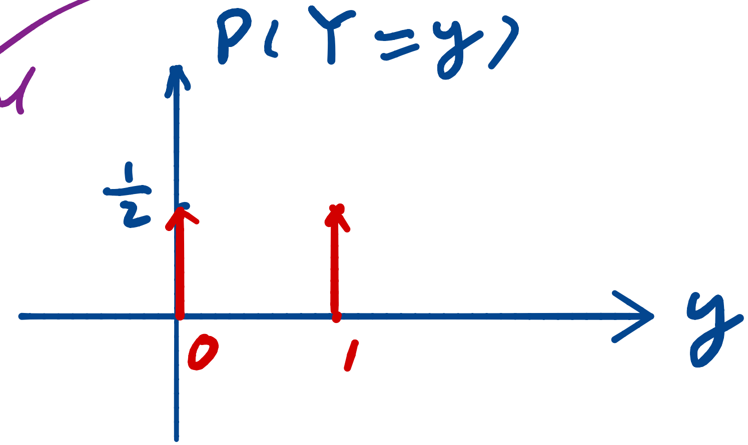
4 → T

$p = 10\%$

# What does it mean that 2 RVs have the same distr.?

identical




$$X(\omega) = \begin{cases} 0 & \text{tail} \\ 1 & \text{head} \end{cases}$$

$$Y(\omega) = \begin{cases} 0 & \text{4-die comes up even} \\ 1 & \text{4-die comes up odd} \end{cases}$$

$$Z(\omega) = \begin{cases} 0 & \text{4-die shows 1 or 2} \\ 1 & \text{. . . 3 or 4} \end{cases}$$

# Three experiments of 2 students

Report the sum of random number each finds after rolling a fair 6-die.

1) each roll once, then add them

2) one of them rolls twice, then add them.

3) one rolls once, then times with 2.

same

same variance

$X+Y$ $\qquad$ $X_1+X_2$ $\qquad$ $2X$

1 M 1st

1 M 2nd

sum of N M-die

distinct numbers

$$\overbrace{1+1 - \cdots 1}^{N} = N$$

$$M \times N = MN$$

$N-1$ N $N+1$ MN

$$MN-(N-1)$$

# Towards the weak law of large numbers

- The weak law says that if we repeat a random experiment many times, the average of the observations will “converge” to the expected value
- For example, if you repeat the profit example, the average earning will “converge” to E[$X$]=20p-10

  $p > 50\%$

- The weak law justifies using simulations (instead of calculation) to estimate the expected values of random variables

  $E[X] \to \sum_x x\, p(x)$

# Markov's inequality

- For any random variable $X$ that *only* takes $x \geq 0$ and constant $a > 0$

$$P(X \geq a) \leq \frac{E[X]}{a}$$

- For example, if $a$ = 10 E[X]

$$P(X \geq 10E[X]) \leq \frac{E[X]}{10E[X]} = 0.1$$

# Proof of Markov's inequality

$X$ only take $x \geq 0$  $a > 0$

$$E[X] = \cancel{\sum_{x \leq 0} x\,p(x)}^{\,0} + \sum_{0<x<a} \overbrace{x\,p(x)}^{\geq 0} + \sum_{x \geq a} x\,p(x)$$

$$\geq \sum_{x \geq a} x\,p(x) \geq \sum_{x \geq a} a\,p(x)$$

$$= a \underbrace{\sum_{x \geq a} p(x)}_{P(X \geq a)}$$

$$= a \cdot P(X \geq a)$$

$$\boxed{P(X \geq a) \leq \frac{E[X]}{a}}$$

# Chebyshev's inequality

- For any random variable $X$ and constant *a* >0

$$P(|X - E[X]| \geq a) \leq \frac{var[X]}{a^2}$$

- If we let a = kσ where σ = std[$X$]

$$P(|X - E[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

- In words, the probability that $X$ is greater than k standard deviation away from the mean is small

# Proof of Chebyshev's inequality

- Given Markov inequality, a>0, x ≥ 0

$$P(X \geq a) \leq \frac{E[X]}{a}$$

it's the same as:

- We can rewrite it as

$Y = |U|$

$y \geq 0$

$\omega > 0$

$$P(|U| \geq w) \leq \frac{E[|U|]}{w}$$

$U = (X - E[X])^2$

$|U| = (X - E[X])^2$

# Proof of Chebyshev's inequality

- If $U = (X - E[X])^2$

$$P(|U| \geq w) \leq \frac{E[|U|]}{w} = \frac{E[(X-E[x])^2]}{w}$$

$$RHS = \frac{var[X]}{w}$$

$w > 0$

$a > 0$

$$p((X-E[x])^2 \geq w) \leq \frac{var[x]}{w}$$

$w = a^2$

$$p(|X-E[x]| \geq a) \leq \frac{var[x]}{a^2}$$

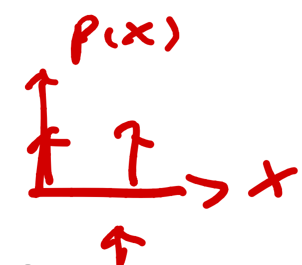
# Now we are closer to the law of large numbers

# Sample mean and IID samples

- We define the sample mean $\overline{\mathbf{X}}$ to be the average of ***N*** random variables $X_1, \ldots, X_N$.

mutual

- If $X_1, \ldots, X_N$ are ***independent*** and have ***identical*** probability function $P(x)$




  then the numbers randomly generated from them are called **IID** samples

- The **sample mean** is a random variable

# Sample mean and IID samples

- Assume we have a set of **IID samples** from ***N*** random variables $X_1, \ldots, X_N$ that have probability function $P(x)$

- We use $\overline{\mathbf{X}}$ to denote the **sample mean** of these **IID samples**

$$\boxed{\overline{\mathbf{X}} = \frac{\sum_{i=1}^{N} X_i}{N}}$$

$E[\bar{X}]$

$\mathrm{var}[\bar{X}]$

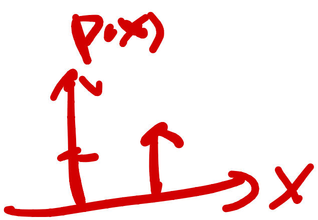
# Expected value of sample mean of IID random variables

- By linearity of expected value

$$E[\overline{\mathbf{X}}] = E[\frac{\sum_{i=1}^{N} X_i}{N}] = \frac{1}{N}\sum_{i=1}^{N} E[X_i]$$

$E[\sum x_i]$

$= \sum_i E[x_i]$




$E[x_1] = E[x_2] = \ldots = E[x_N]$

$E[x]$

$\underline{E[\bar{x}]} = \frac{1}{N} \cdot N \cdot E[x] = \underline{E[x]}$

# Expected value of sample mean of IID random variables

- By linearity of expected value

$$E[\overline{\mathbf{X}}] = E[\frac{\sum_{i=1}^{N} X_i}{N}] = \frac{1}{N}\sum_{i=1}^{N} E[X_i]$$

$E[X] = E[X_1] = E(X_2) \cdots = E[X_n]$

- Given each $X_i$ has identical $P(x)$

$$\boxed{E[\overline{\mathbf{X}}] = \frac{1}{N}\sum_{i=1}^{N} E[X] = E[X]}$$

# Variance of sample mean of IID random variables

- By the scaling property of variance

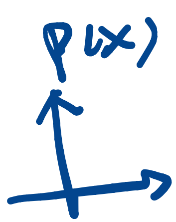
$$var[\overline{\mathbf{X}}] = var[\frac{1}{N}\sum_{i=1}^{N} X_i] = \frac{1}{N^2} var[\sum_{i=1}^{N} X_i]$$

$\{(X_i, X_j)\}$ mutual indpt.

$$var[X_1 + X_2] = var[X_1] + var[X_2]$$

$$var[\sum_i X_i] = \sum_i var[X_i]$$

$X_i$:




$$var[X_1] = var[X_2] \ldots = var[X_N]$$

$$= var[X]$$

$$var[\bar{X}] = \frac{1}{N^2} \sum^{N} var[X] = \frac{1}{N^2} \cdot N \cdot var[X]$$

# Variance of sample mean of IID random variables

- By the scaling property of variance

$$var[\overline{\mathbf{X}}] = var[\frac{1}{N}\sum_{i=1}^{N} X_i] = \frac{1}{N^2} var[\sum_{i=1}^{N} X_i]$$

- And by independence of these IID random variables

$$var[\overline{\mathbf{X}}] = \frac{1}{N^2}\sum_{i=1}^{N} var[X_i]$$

$$var[\bar{x}] = \frac{1}{N^2} \cdot \sum_{i=1}^{N} var[x] = \frac{1}{N^2} \cdot N \cdot var[x]$$

$$= \frac{1}{N} var[x]$$

# Expected value and variance of sample mean of IID random variables

- The expected value of sample mean is the same as the expected value of the distribution

$$E[\overline{\mathbf{X}}] = E[X]$$

- The variance of sample mean is the distribution's variance divided by the sample size ***N***

$$var[\overline{\mathbf{X}}] = \frac{var[X]}{N}$$

# Weak law of large numbers

- Given a random variable $X$ with finite variance, probability distribution function $P(x)$ and the sample mean $\overline{\mathbf{X}}$ of size ***N***.

- For any positive number $\epsilon > 0$

$$\lim_{N \to \infty} P(|\overline{\mathbf{X}} - E[X]| \geq \epsilon) = 0$$

- That is: the value of the mean of **IID** samples is very close with high probability to the expected value of the population when sample size is very large

# Proof of Weak law of large numbers

- Apply Chebyshev's inequality

$$P(|\overline{\mathbf{X}} - E[\overline{\mathbf{X}}]| \geq \epsilon) \leq \frac{var[\overline{\mathbf{X}}]}{\epsilon^2}$$

$$E[\bar{x}] = E[x]$$

$$var[\bar{x}] = \frac{1}{N} var[x]$$

# Proof of Weak law of large numbers

- Apply Chebyshev's inequality

$$P(|\overline{\mathbf{X}} - E[\overline{\mathbf{X}}]| \geq \epsilon) \leq \frac{var[\overline{\mathbf{X}}]}{\epsilon^2}$$

- Substitute $E[\overline{\mathbf{X}}] = E[X]$ and $var[\overline{\mathbf{X}}] = \frac{var[X]}{N}$

$$P(|\overline{\mathbf{X}} - E[\mathbf{X}]| \geq \epsilon) \leq \frac{var[\mathbf{X}]}{N\epsilon^2} \rightarrow 0$$

$N \rightarrow \infty$

# Proof of Weak law of large numbers

- Apply Chebyshev's inequality

$$P(|\overline{\mathbf{X}} - E[\overline{\mathbf{X}}]| \geq \epsilon) \leq \frac{var[\overline{\mathbf{X}}]}{\epsilon^2}$$

- Substitute $E[\overline{\mathbf{X}}] = E[X]$ and $var[\overline{\mathbf{X}}] = \dfrac{var[X]}{N}$

iid X.

$$P(|\overline{\mathbf{X}} - E[\mathbf{X}]| \geq \epsilon) \leq \frac{var[\mathbf{X}]}{N\epsilon^2} \longrightarrow 0$$

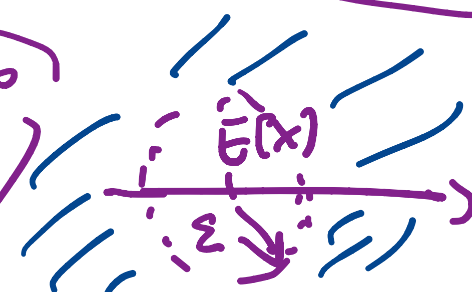
$$\boxed{N} \to \infty$$

N ↑

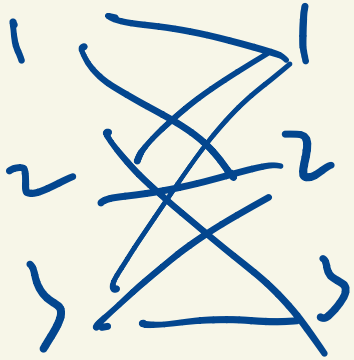
$\varepsilon$

$\varepsilon$

$10^{-10^{10}}$



$N = 2 \qquad m = 3 \qquad \{1, 2, 3\}$




Sum can be

$\{2, 3, 4, 5, 6\}$

$$\underbrace{N \quad N+1 \quad N+2 \cdots N\cdot m}_{N\cdot m-(N-1)}$$

| 1st | 2nd | |
|---|---|---|
| 2 | 2 | $\frac{2+2}{2}$ |
| 3 | 3 | $\frac{3+3}{2}$ |
| 3 | 2 | $\frac{3+2}{2}$ |
| 1 | 1 | $\frac{1+1}{2}$ |
| 1 | 3 | $\frac{1+3}{2}$ |

$\bar{X}$

R.V.

# Proof of Weak law of large numbers

- Apply Chebyshev's inequality

$$P(|\overline{\mathbf{X}} - E[\overline{\mathbf{X}}]| \geq \epsilon) \leq \frac{var[\overline{\mathbf{X}}]}{\epsilon^2}$$

- Substitute $E[\overline{\mathbf{X}}] = E[X]$ and $var[\overline{\mathbf{X}}] = \frac{var[X]}{N}$

$$P(|\overline{\mathbf{X}} - E[\overline{\mathbf{X}}]| \geq \epsilon) \leq \frac{var[\mathbf{X}]}{N\epsilon^2} \xrightarrow[N \to \infty]{} 0$$

$$\boxed{\lim_{N\to\infty} P(|\overline{\mathbf{X}} - E[X]| \geq \epsilon) = 0}$$

# Applications of the Weak law of large numbers

# Applications of the Weak law of large numbers

- The law of large numbers ***justifies using simulations*** (instead of calculation) to estimate the expected values of random variables

$$\lim_{N\to\infty} P(|\overline{\mathbf{X}} - E[X]| \geq \epsilon) = 0$$

- The law of large numbers also ***justifies using histogram*** of large random samples to approximate the probability distribution function $P(x)$, see proof on

Pg. 353 of the textbook by DeGroot, et al.

# Histogram of large random IID samples approximates the probability distribution

- The law of large numbers justifies using histograms to approximate the probability distribution. Given ***N*** IID random variables $X_1$, …, $X_N$
  - According to the law of large numbers

$$\overline{\mathbf{Y}} = \frac{\sum_{i=1}^{N} Y_i}{N} \xrightarrow{N \to \infty} E[Y_i]$$

  - As we know for indicator function

$$E[Y_i] = P(c_1 \leq X_i < c_2) = P(c_1 \leq X < c_2)$$

# Simulation of the sum of two-dice

- http://www.randomservices.org/random/apps/DiceExperiment.html

# Probability using the property of Independence: Airline overbooking

- An airline has a flight with **s** seats. They always sell **t** (**t**>**s**) tickets for this flight. If ticket holders show up independently with probability **p**, what is the probability that the flight is overbooked ?

$$\text{P( overbooked)} = \sum_{u=s+1}^{t} C(t,u) p^u (1-p)^{t-u}$$

# Simulation of airline overbooking

- An airline has a flight with **7** seats. They always sell 12 tickets for this flight. If ticket holders show up independently with probability **p**, estimate the following values
  - Expected value of the number of ticket holders who show up
  - Probability that the flight being overbooked
  - Expected value of the number of ticket holders who can't fly due to the flight is overbooked.

# Conditional expectation

- Expected value of X conditioned on event A:

$$E[X|A] = \sum_{x \in D(X)} xP(X = x|A)$$

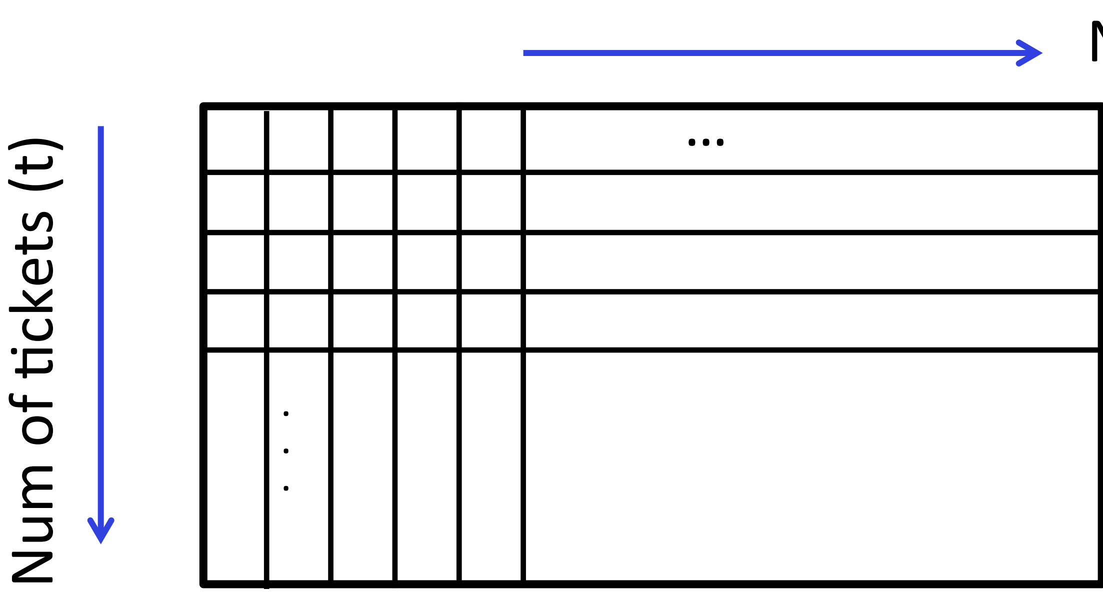
- Expected value of the number of ticketholders not flying

$$E[NF|overbooked] = \sum_{u=s+1}^{t} (u-s) \frac{\binom{t}{u} p^u (1-p)^{t-u}}{\sum_{v=s+1}^{t} \binom{t}{v} p^v (1-p)^{t-v}}$$

# Simulate the arrival

- Expected value of the number of ticket holders who show up

***nt=100000, t= 12, s=7, p=0.1, 0.2, … 1.0***

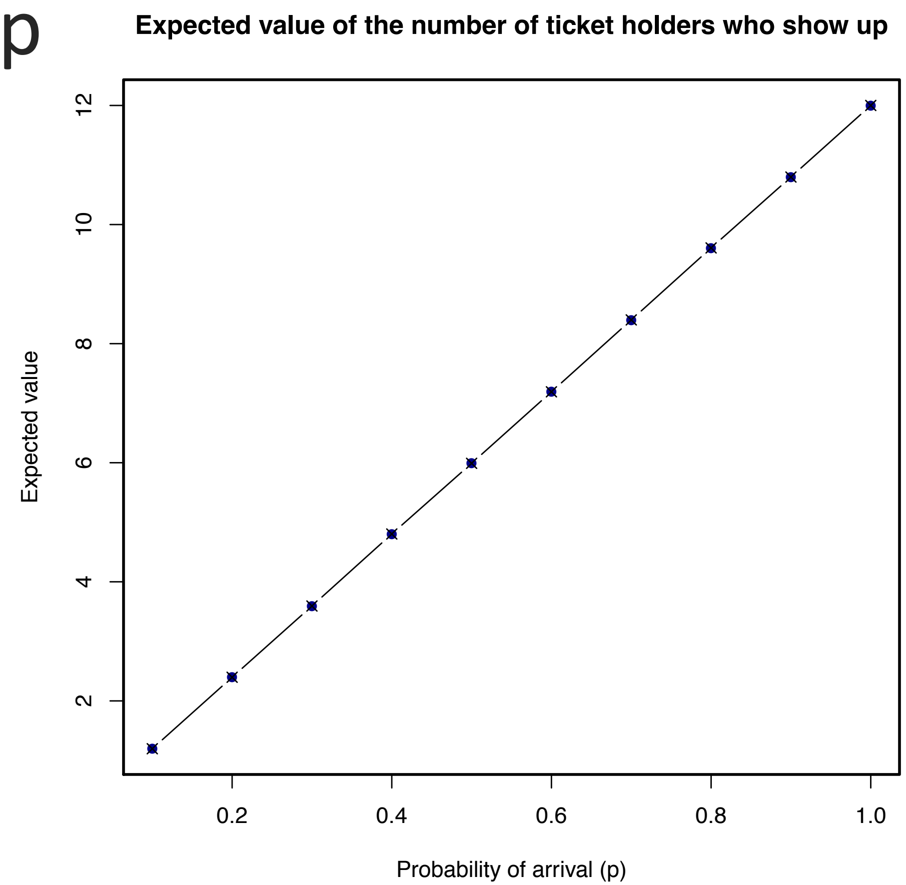



We generate a matrix of random numbers from uniform distribution in [0,1],
**Any number < p is considered an arrival**

# Simulate the arrival

- Expected value of the number of ticket holders who show up

***nt=100000, t= 12, s=7, p=0.1, 0.2, … 1.0***



# Simulate the expected probability of overbooking

- Expected probability of the flight being overbooked

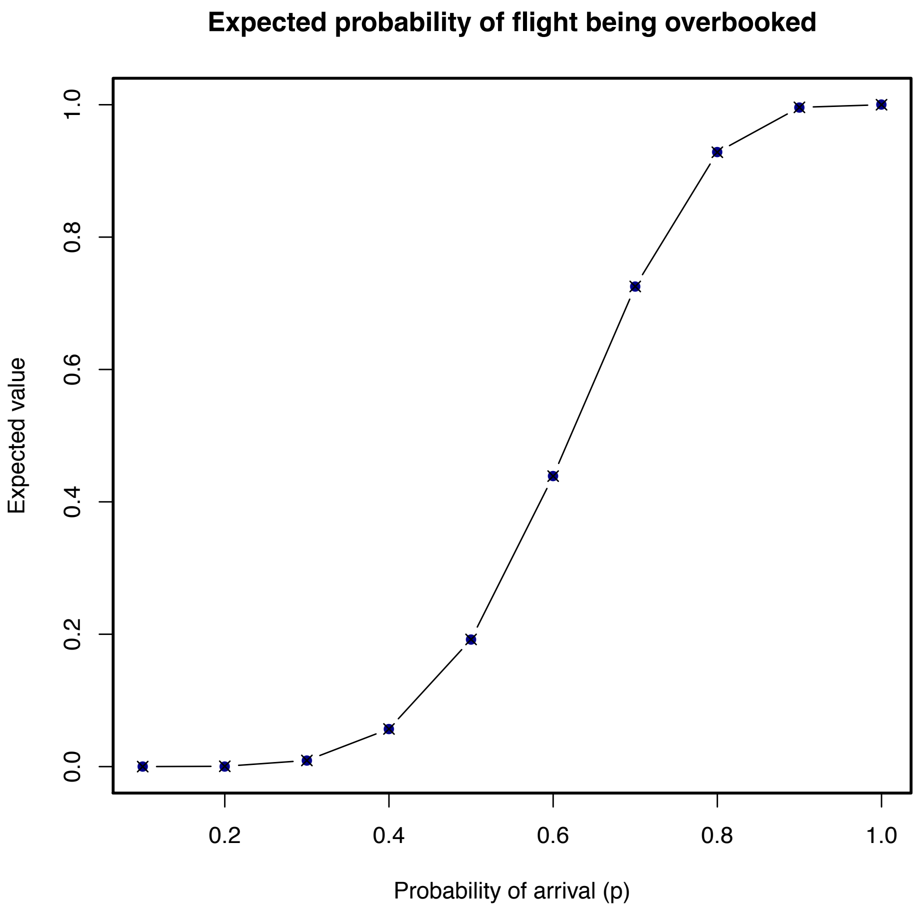
  ***t= 12, s=7, p=0.1, 0.2, … 1.0***

- **Expected probability** is equal to the **expected value of indicator function**. Whenever we have Num of arrival > Num of seats, we mark it with an indicator function. Then estimate with the sample mean of indicator functions.

# Simulate the expected probability of overbooking

- Expected probability of the flight being overbooked

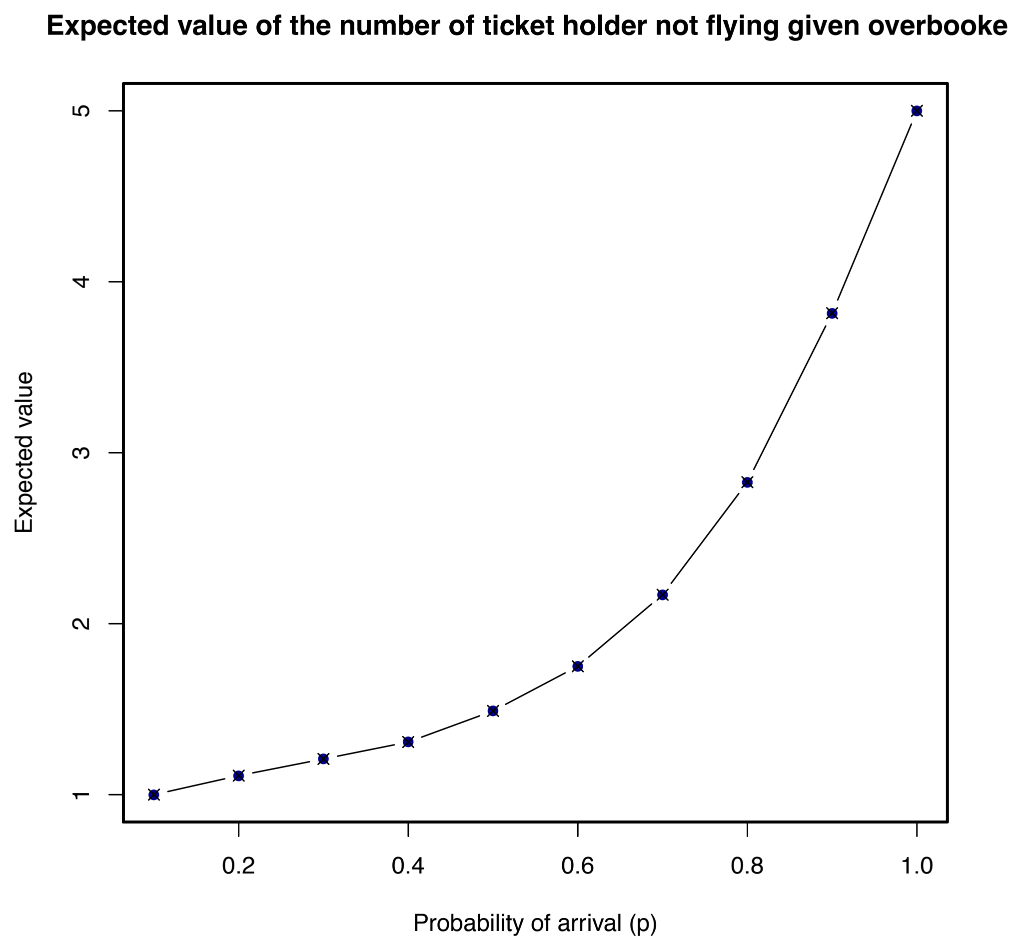
***nt=100000,***
***t= 12, s=7,***
***p=0.1, 0.2, … 1.0***



# Simulate the expected value of the number of grounded ticket holders given overbooked

- Expected value of the number of ticket holders who can’t fly due to the flight being overbooked

***Nt=200000,***
***t= 12, s=7,***
***p=0.1, 0.2, … 1.0***



# Assignments

- Finish Chapter 4 of the textbook
- Next time: Continuous random variable, classic known probability distributions

# Additional References

- Charles M. Grinstead and J. Laurie Snell "Introduction to Probability"
- Morris H. Degroot and Mark J. Schervish "Probability and Statistics"

# See you next time

*See You!*