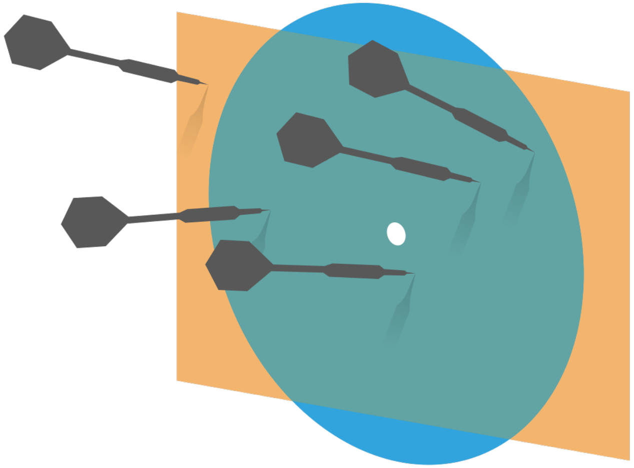
Probability and Statistics for Computer Science



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, 02.23.2021

Last time

- Random Variable
  - *Expected value*
  - *Variance & covariance*

Objectives

- Random Variable
  - Review
  - Covariance
  - *The weak law of large numbers*
  - *Simulation & example of airline overbooking*

Expected value

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

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

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

Linearity of Expectation

Expected value of a function of X

Q:

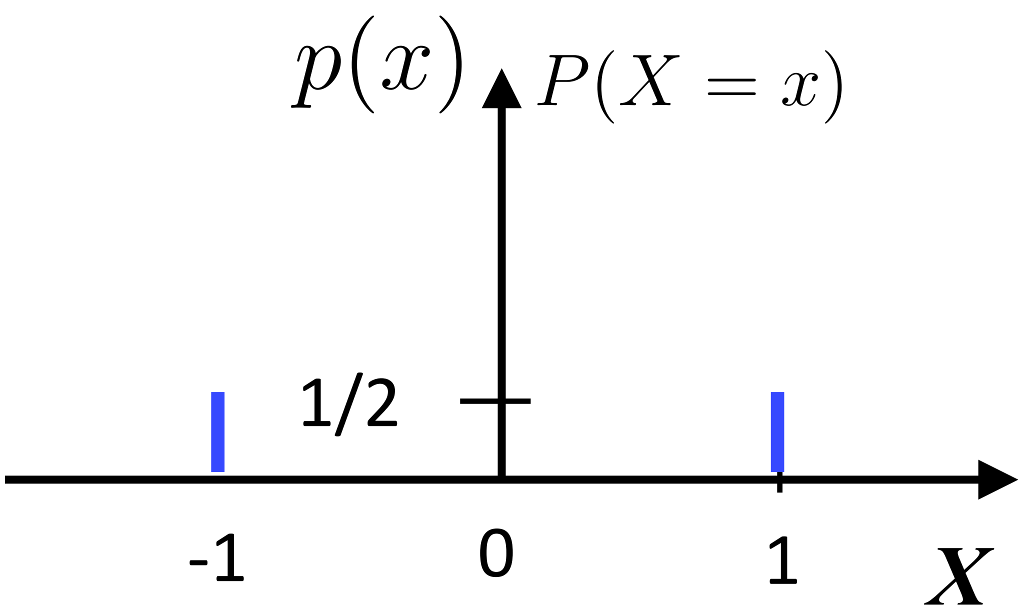
What is E[E[X]]?

A. E[X]
B. 0
C. Can’t be sure

Probability distribution

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

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

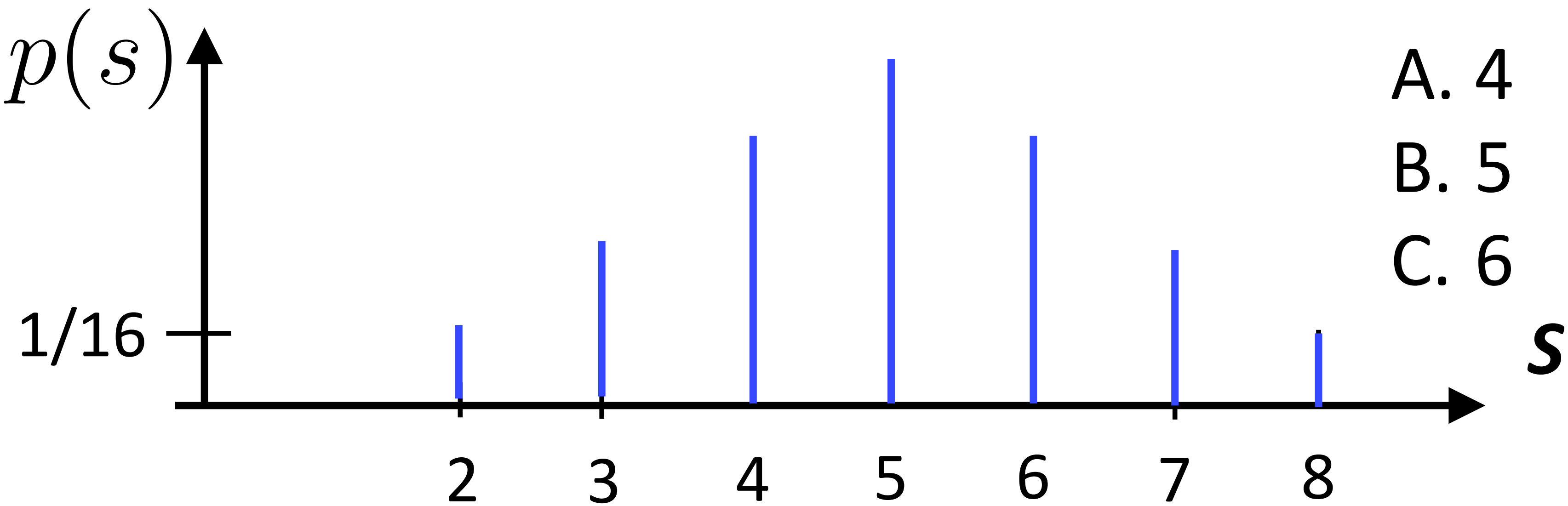


A. 0
B. 1
C. 2
D. 3
E. 5

Probability distribution

- Given 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 **E[S]** ?



A. 4

B. 5

C. 6

A neater expression for variance

- Variance of Random Variable X is defined as:

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

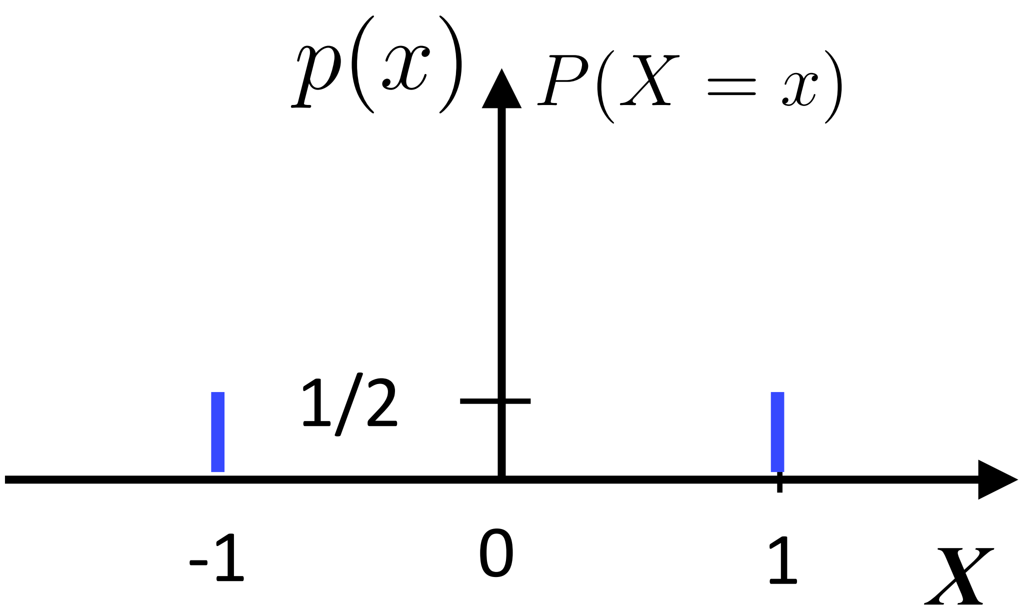
- It's the same as:

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

Probability distribution and cumulative distribution

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

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



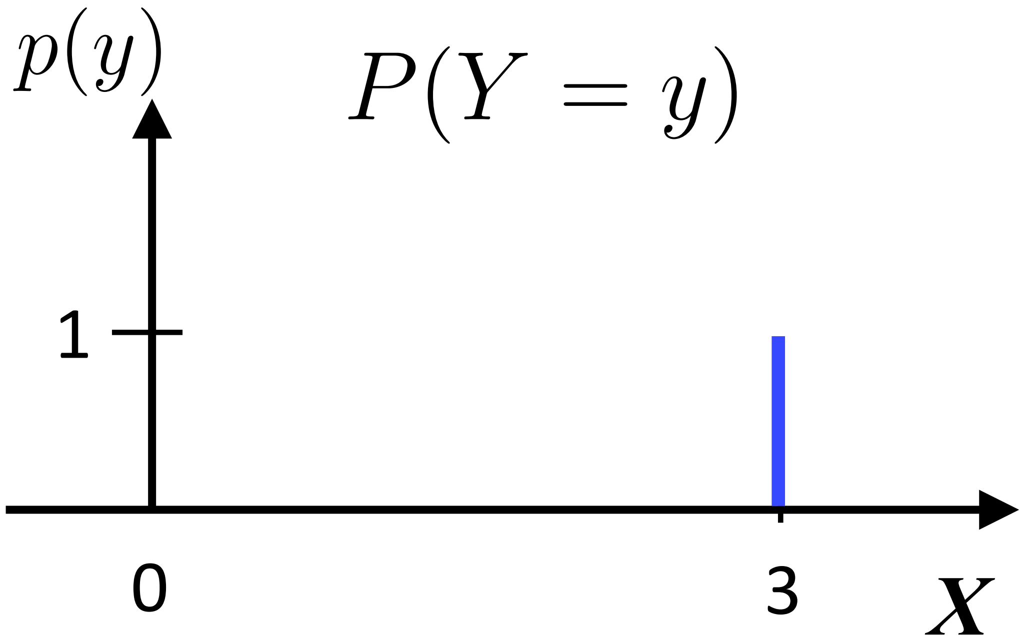
A. 0
B. 1
C. 2
D. 3
E. -1

Probability distribution

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

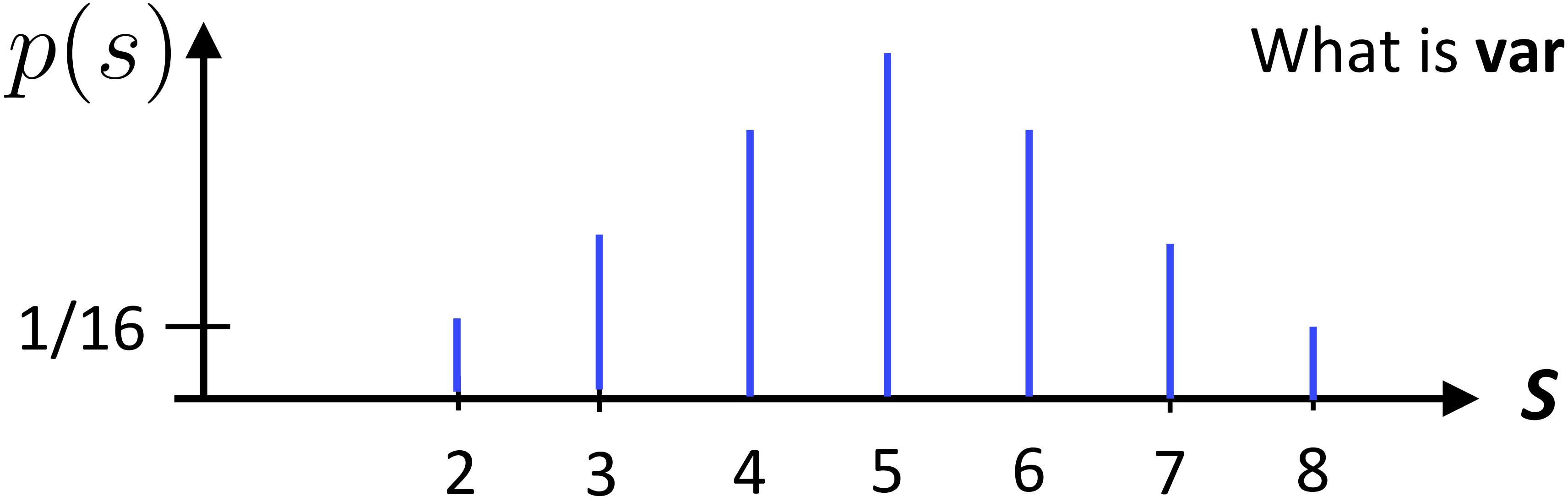
var[2|***X***| +1]? Let ***Y*** = 2|***X***|+1

$$P(Y = y)$$



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]** ?

Motivation for covariance

- Study the relationship between random variables
- Note that it's the un-normalized correlation
- Applications include the fire control of radar, communicating in the presence of noise.

Covariance

- The **covariance** of random variables X and Y is

$$cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$

- Note that

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

A neater form for covariance

- A neater expression for **covariance** (similar derivation as for variance)

$$cov(X, Y) = E[XY] - E[X]E[Y]$$

Correlation coefficient is normalized covariance

- The correlation coefficient is

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

- When X, Y takes on values with equal probability to generate data sets $\{(x,y)\}$, the correlation coefficient will be as seen in Chapter 2.

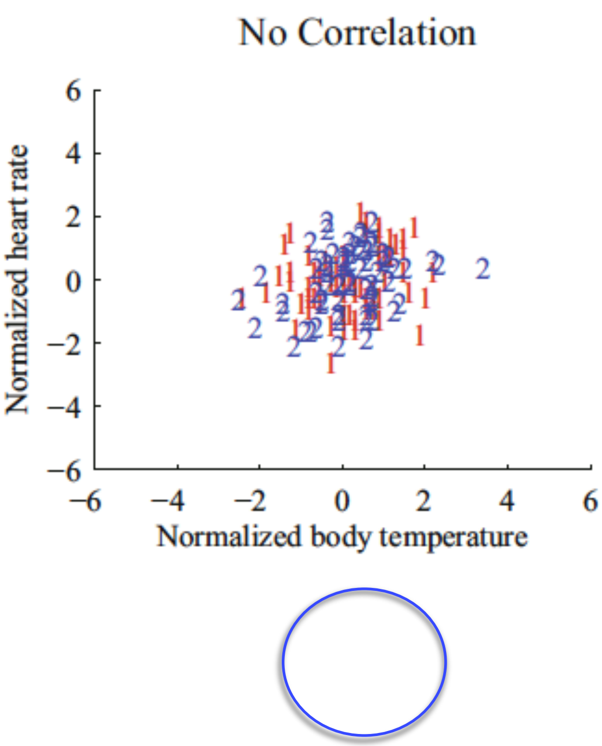
Correlation coefficient is normalized covariance

- The correlation coefficient can also be written as:

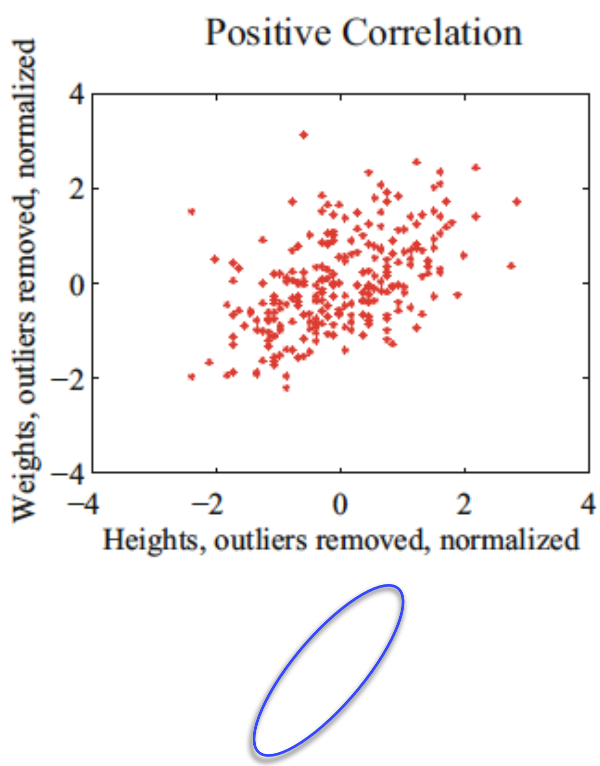
$$corr(X,Y) = \frac{E[XY] - E[X]E[Y]}{\sigma_X \sigma_Y}$$

Correlation seen from scatter plots

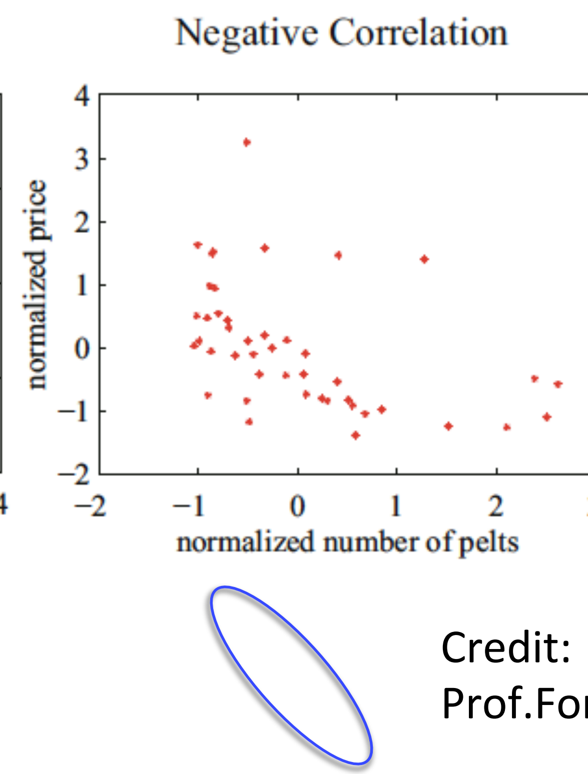
Zero Correlation



Positive correlation



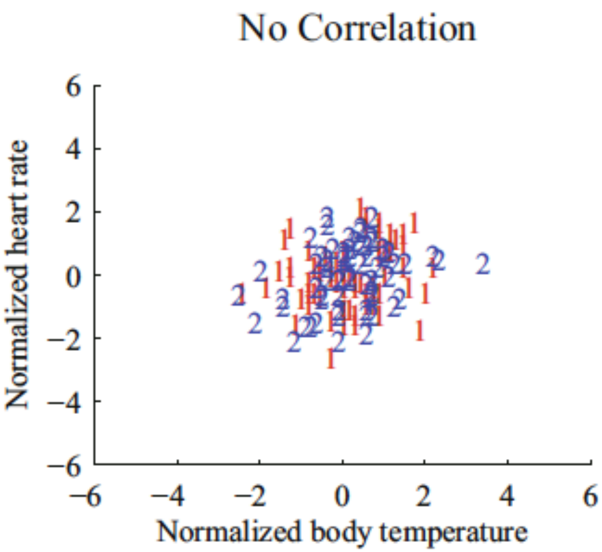
Negative correlation

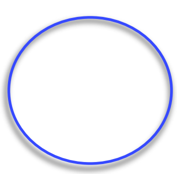


Credit: Prof.Forsyth

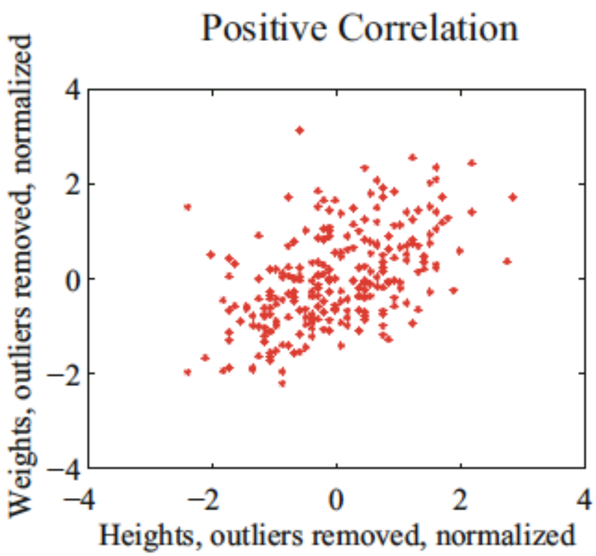
Covariance seen from scatter plots

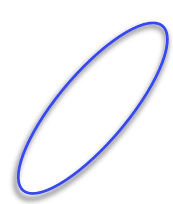
Zero **Covariance**



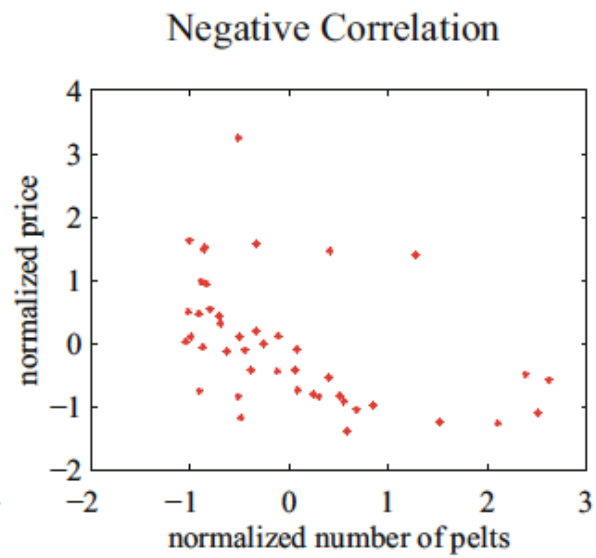


Positive **Covariance**





Negative **Covariance**



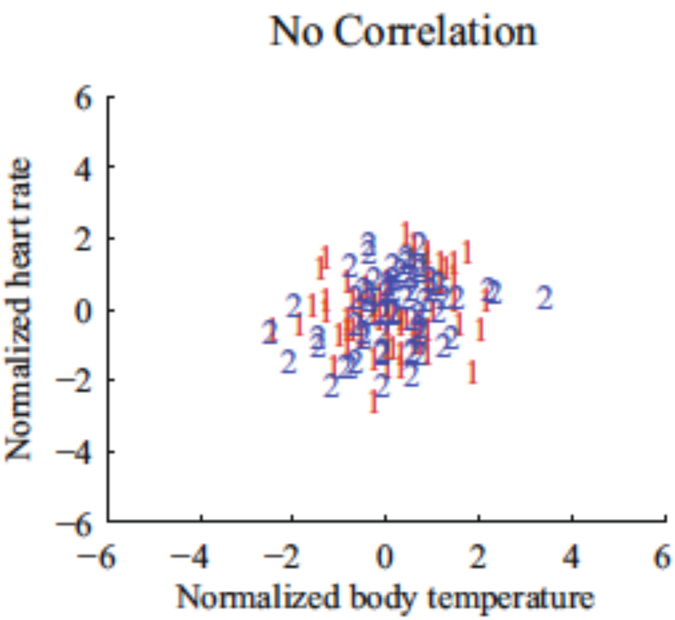
Credit: Prof.Forsyth

When correlation coefficient or covariance is zero

- The covariance is 0!
- That is:

$$E[XY] - E[X]E[Y] = 0$$

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



- This is a necessary property of independence of random variables * (not equal to independence)

Variance of the sum of two random variables

$$var[X+Y] = var[X] + var[Y] + 2cov(X,Y)$$

If events X &Y are independent, then

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

Proof:

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

These are equivalent! Uncorrelatedness

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

  $cov(X, Y) = 0$

  $var[X + Y] = var[X] + var[Y]$

Q: What is this expectation?

- We toss two fair 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]?

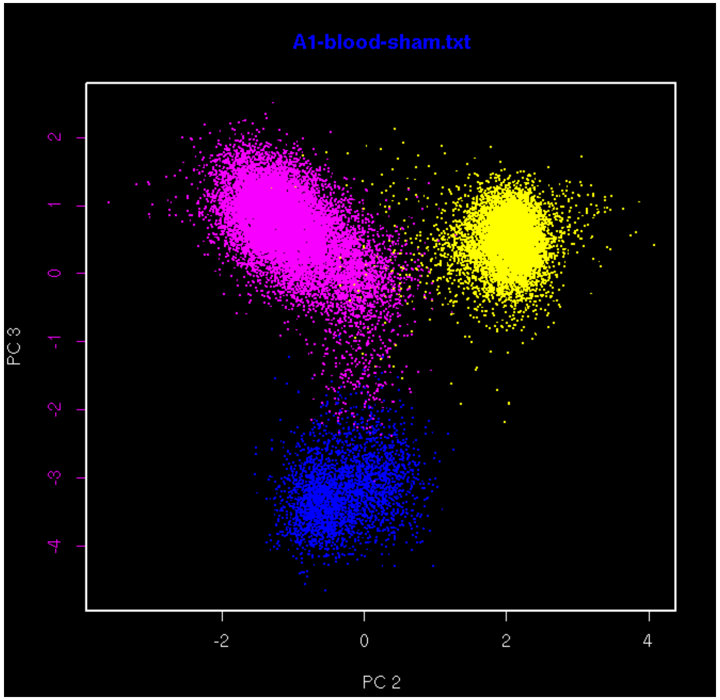
A. 2 B. 3 C. 4

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 $Y=X^2$.

Covariance example

It's an underlying concept in principal component analysis in Chapter 10



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 $E[X]=20p-10$

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

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

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

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

- We can rewrite it as

$\omega > 0$

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

Proof of Chebyshev's inequality

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

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

Proof of Chebyshev's inequality

- Apply Markov inequality to $U = (X - E[X])^2$

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

- Substitute $U = (X - E[X])^2$ and $w = a^2$

$$P((X - E[X])^2 \geq a^2) \leq \frac{var[X]}{a^2}$$ Assume $a > 0$

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

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 $X_1, \ ..., \ X_N$.
- If $X_1, \ ..., \ X_N$ 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**

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

Expected value of sample mean of IID random variables

- By linearity of expected value

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

Expected value of sample mean of IID random variables

- By linearity of expected value

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

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

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

Variance of sample mean of IID random variables

- By the scaling property of variance

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

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

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

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

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

- Given each X_i has identical $P(x)$, $var[X_i] = var[X]$

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

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

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

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

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

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

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

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

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

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

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

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:

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

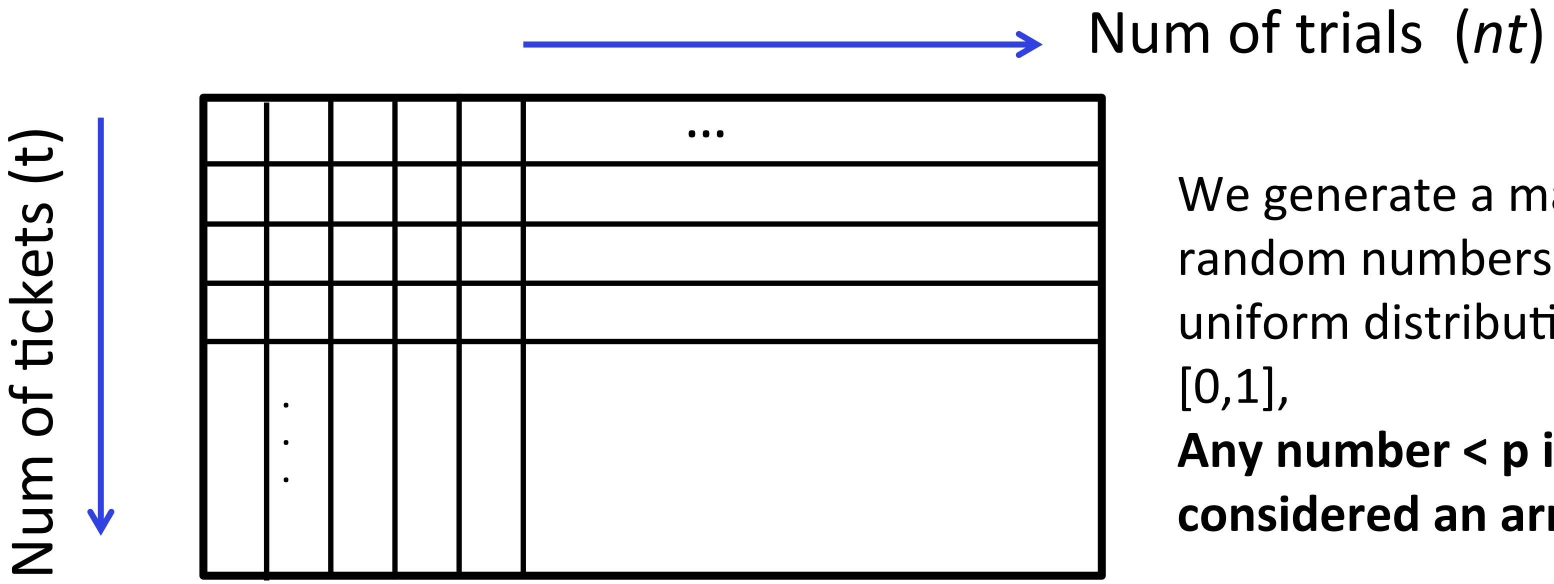
- Expected value of the number of ticketholders not flying

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

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

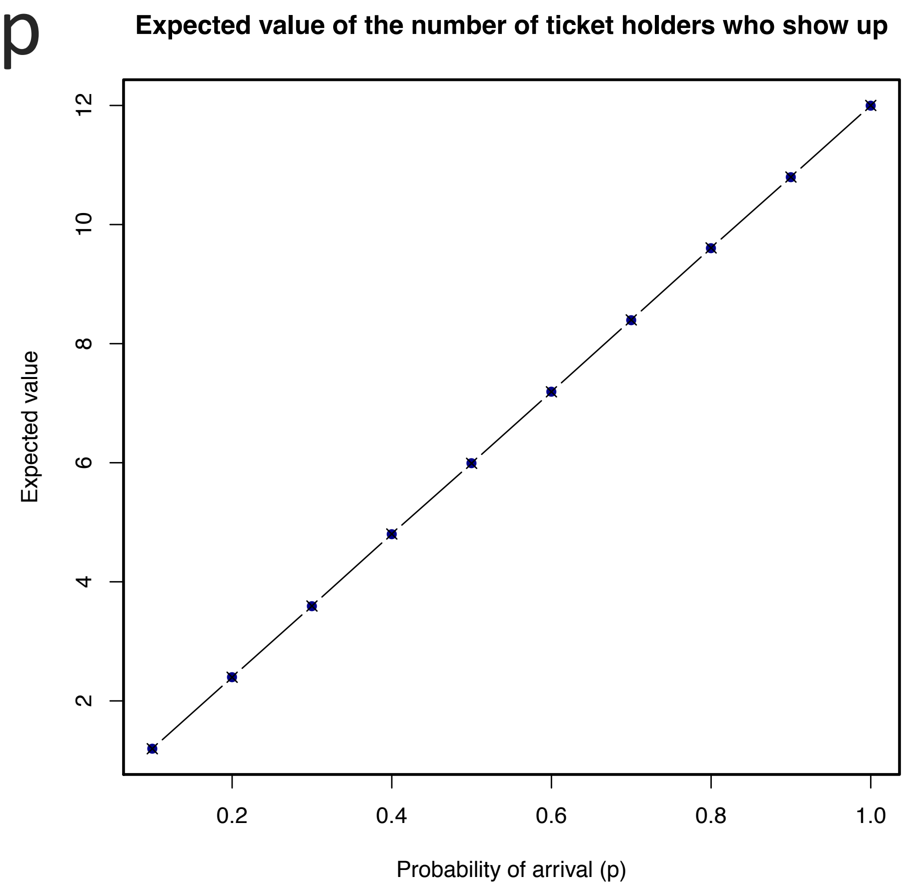


We generate a matrix of random numbers from uniform distribution in [0,1],
Any number < p is considered an arrival

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

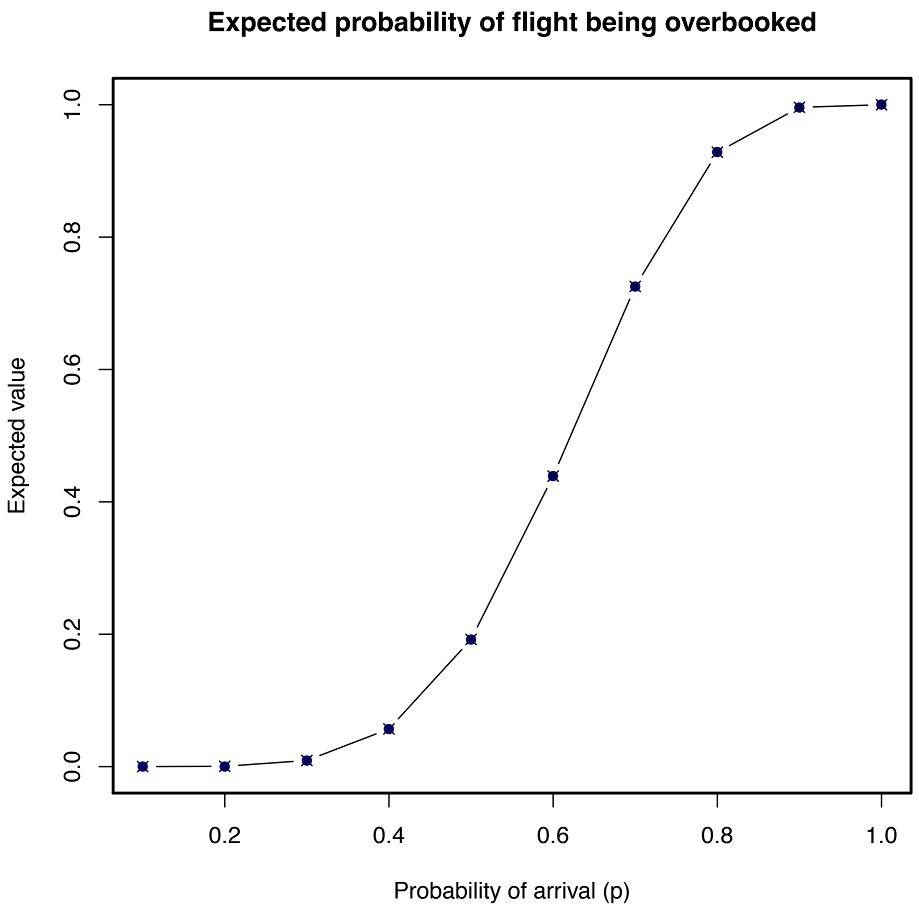
t= 12, s=7, p=0.1, 0.2, … 1.0

- **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

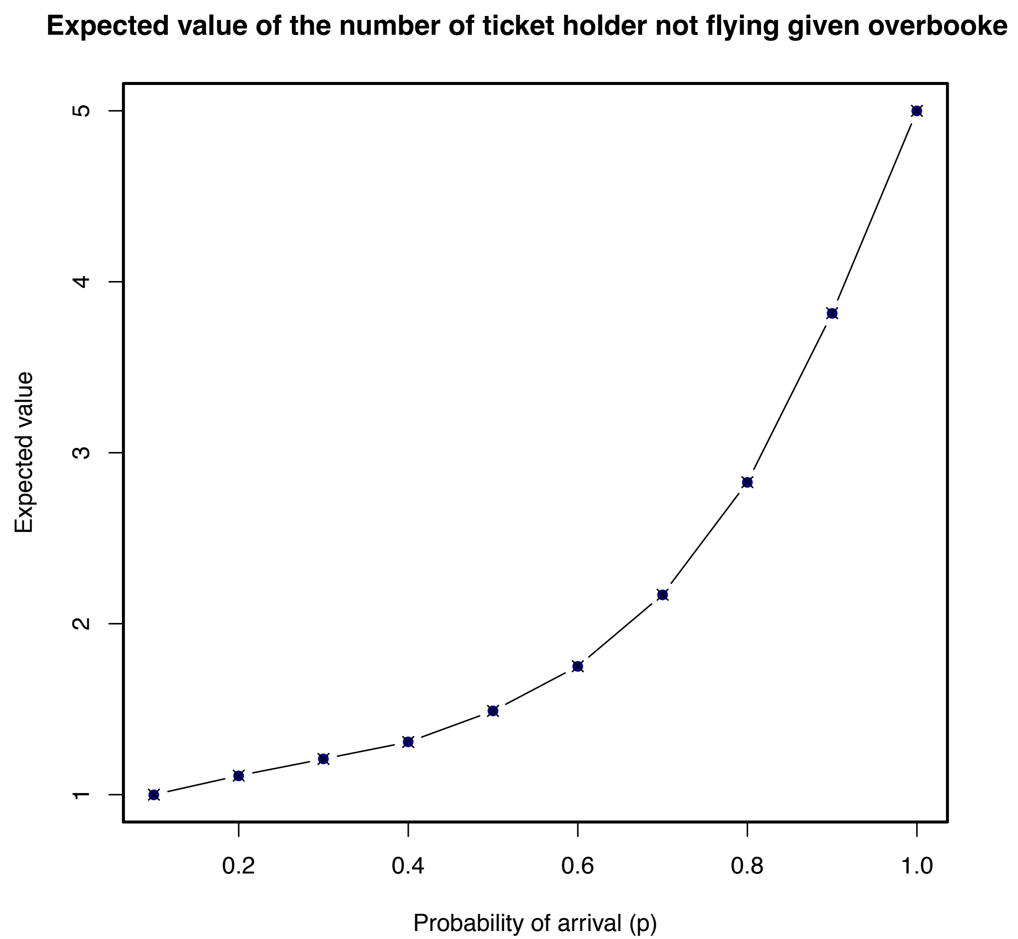
nt=100000,
t= 12, s=7,
p=0.1, 0.2, … 1.0



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

- Continue to work on HW4
- Read Module Week 5
- 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!