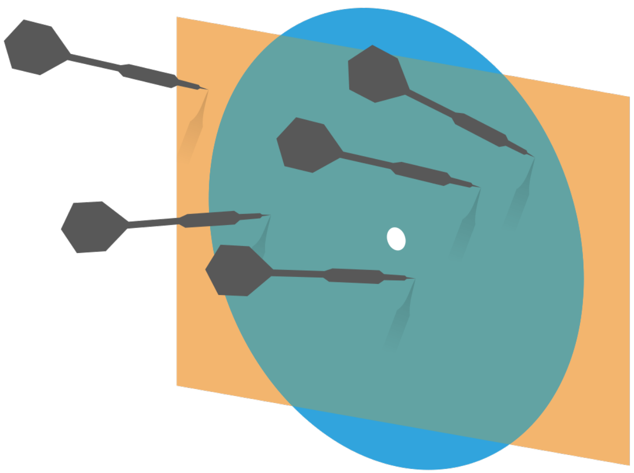
Probability and Statistics for Computer Science



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, 09.22.2021

Last time

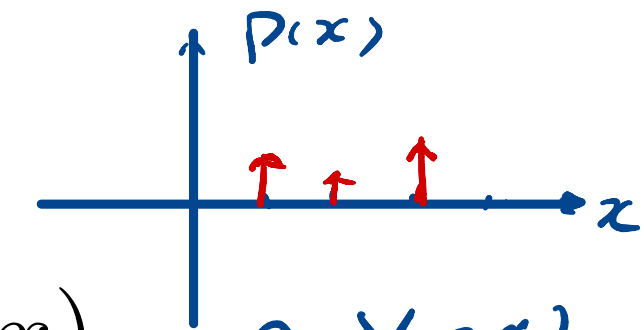
- Random Variable
  - *Expected value*
  - *Variance & covariance*

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



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

Theoretical mean

$p(X = x)$
$p(x)$

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

Linearity of Expectation

$$E[aX + bY] = aE[X] + bE[Y]$$

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

Expected value of a function of X

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

$$E[f(X,Y)] = \sum_x \sum_y f(x,y)\,\underbrace{p(x,y)}_{p(X=x \,\cap\, Y=y)}$$

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. * ML

Covariance

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

corr $= \frac{\sum \hat{x}\hat{y}}{N}$

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

$E[f(X,Y)] = \sum_x \sum_y f(x,y)\,p(x,y)$

- Note that

here $f(x,y) = (X - E[X])(Y - E[Y])$

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

A neater form for covariance

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

$$
\begin{array}{c|cc}
Y & & \\
1 & 0 & 1 \\
0 & 0 & 0 \\
\hline
 & 0 & 1 \quad X
\end{array}
$$

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

$=0$

what is $E[XY]$?

$\frac{1}{4}$

$\sum xy\, p(x,y)$

$1 \times \frac{1}{4}$

$\frac{1}{4} - \frac{1}{4}$

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.

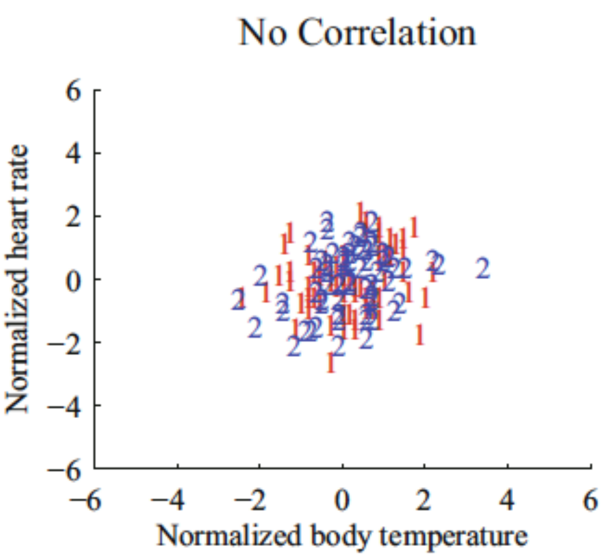
Correlation coefficient is normalized covariance

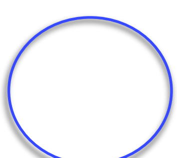
- The correlation coefficient can also be written as:

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

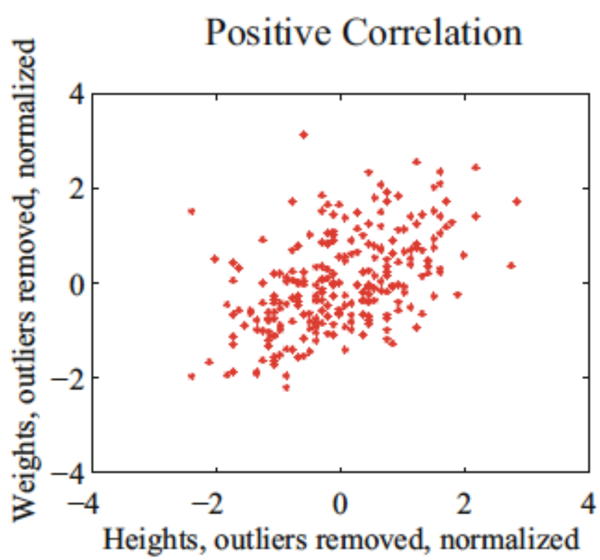
Correlation seen from scatter plots

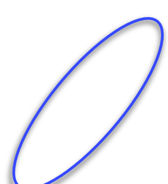
Zero Correlation



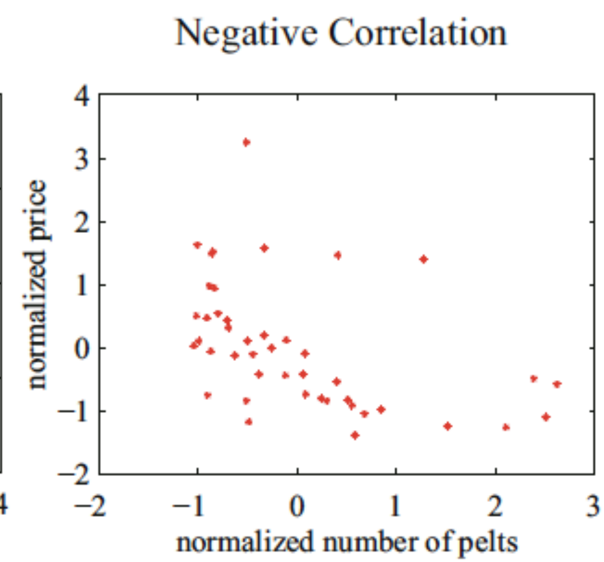


Positive correlation





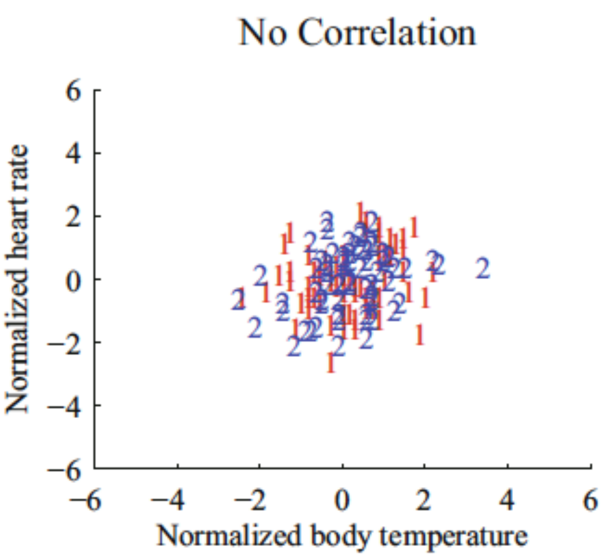
Negative correlation

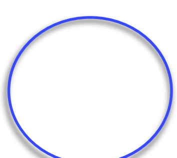


Credit: Prof.Forsyth

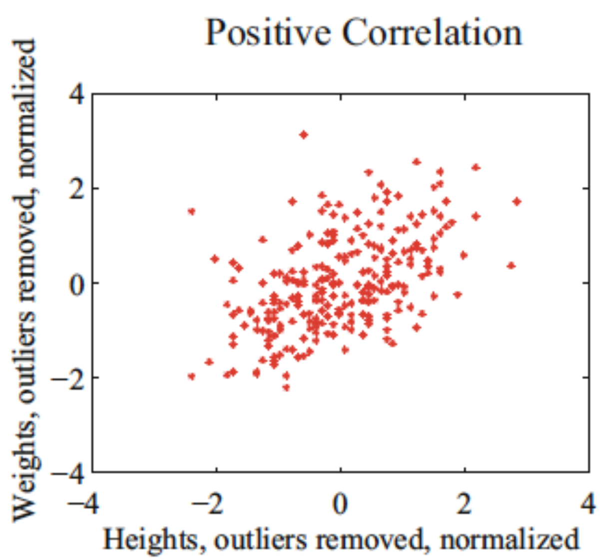
Covariance seen from scatter plots

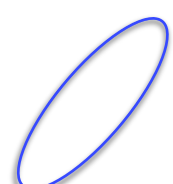
Zero Covariance ↓



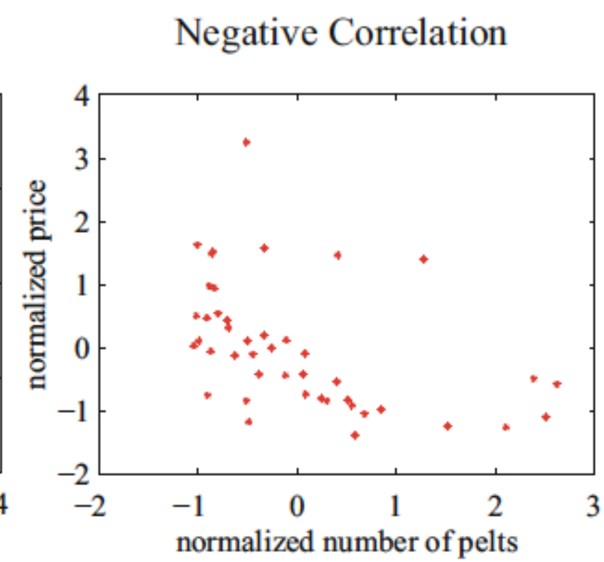


Positive Covariance ↓





Negative Covariance ↓



Credit: Prof.Forsyth

When correlation coefficient or covariance is zero

- The covariance is 0!
- That is:

Cov = 0

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

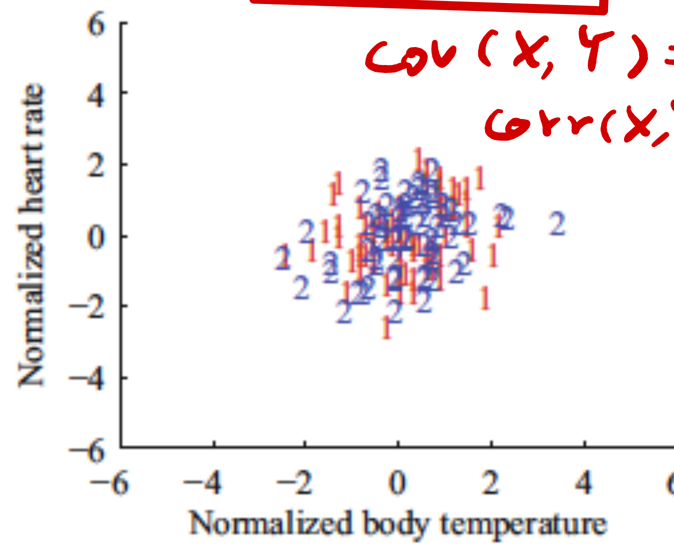
$$E[XY] = E[X]E[Y]$$

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

No Correlation

cov(X, Y) = 0

corr(X,Y) = 0



Variance of the sum of two random variables

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

HW extra points

If RVs X &Y are independent, then

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

$$E[XY] = \sum_y \sum_x xy\, P(x,y)$$

$P(x,y)$
$= P(X=x \cap Y=y)$

if X, Y are indpt.

$$P(x,y) = P(x)P(y) \quad \text{for all } x, y$$

$$RHS = \sum_y \sum_x x\,P(x)\, y\, P(y)$$

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

$$= E[Y]\,E[X]$$

$$= E[X]\,E[Y]$$

These are equivalent! Uncorrelatedness

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

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

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

$= 0$

$$cov(X, Y) = 0$$

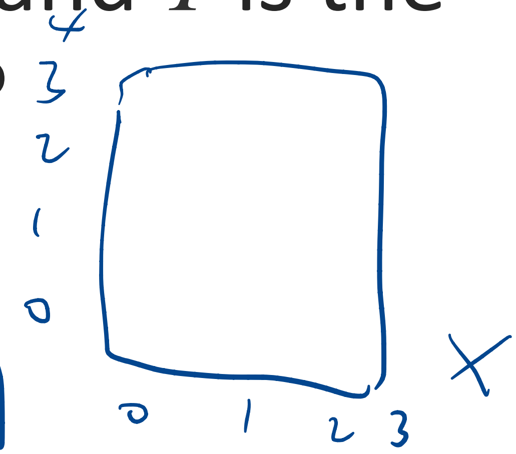
$\Leftrightarrow Corr(X, Y) = 0$

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

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 $E[XY]$?

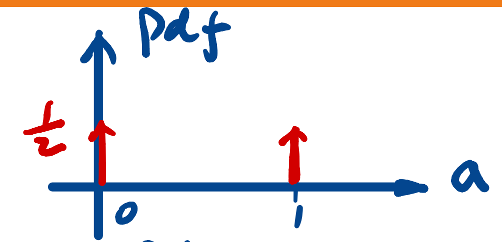
A. 2 B. 3 C. 4

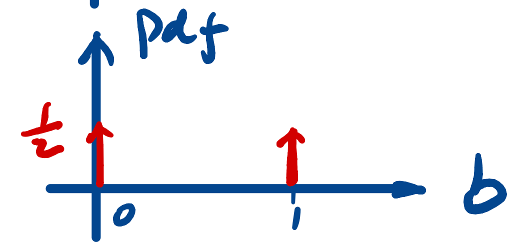


$E[XY] = E[X]E[Y]$

$$X = A_1 + A_2 + A_3$$

$$Y = B_1 + B_2 + B_3 + B_4$$





$$E[X] = E[A_1] + E[A_2] + E[A_3]$$
$$= 0.5 + 0.5 + 0.5 = 1.5$$

$$E[Y] = E\left[\sum B_i\right] = \sum_{i=1}^{4} E[B_i] = 4 \times 0.5 = 2$$

$\because$ X, Y are independent, $E[XY] = E[X]E[Y]$

$$E[XY] = E[X]E[Y] = 1.5 \times 2 = 3$$

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 $Y=X^2$.

$E[X]=?$ 0

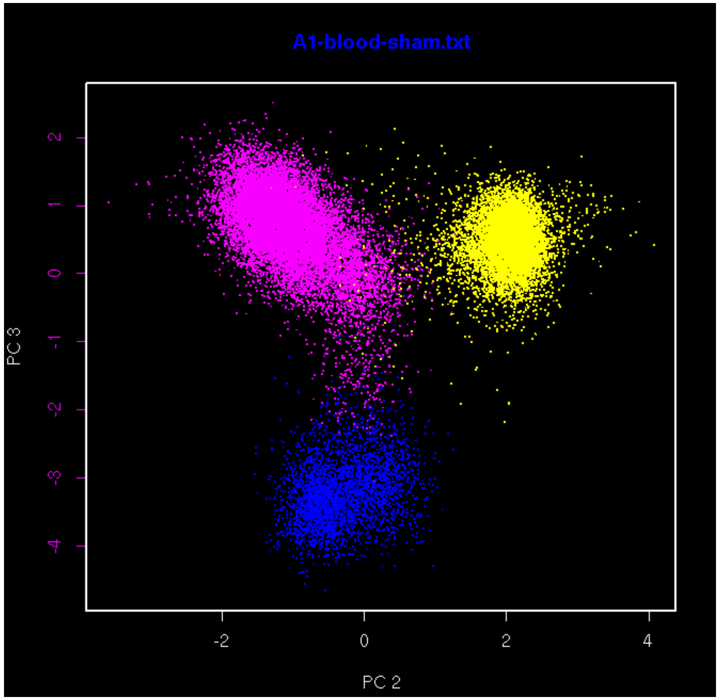
$E[Y]=?$ $\frac{2}{3}$

$E[XY]=?$ 0

X, Y are depedent!! but uncorrelated

Covariance example

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

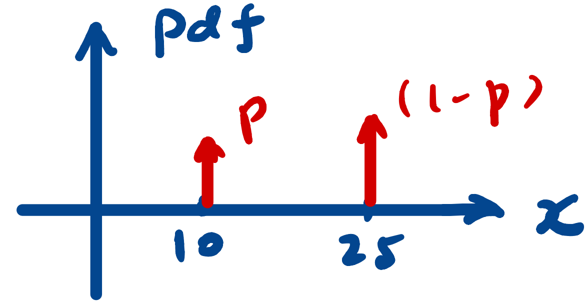


Random Variable Example for WLLN

Money box

10¢ 25¢

dime quarter



* shake and take one and put back

$X_1 = 10?\ 25?$

$X_2 = 10?\ 25?$

$\vdots$

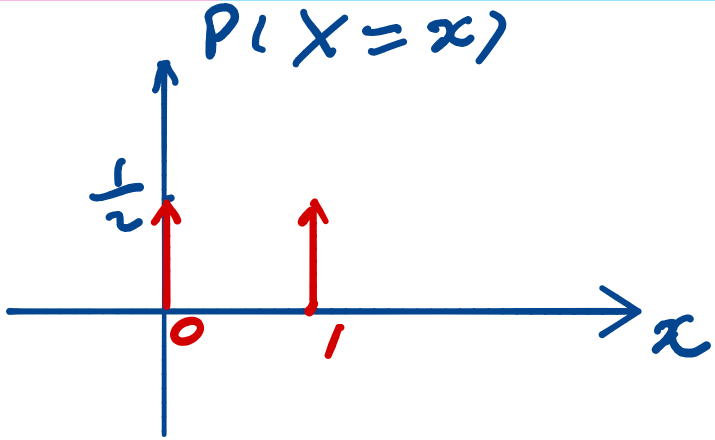
X_n

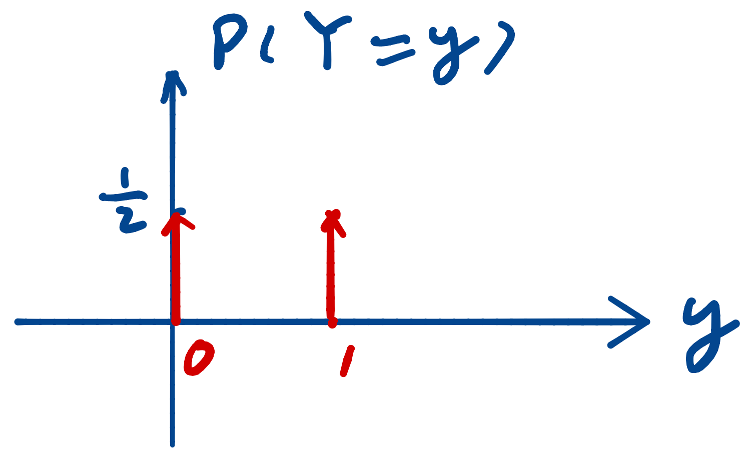
$E[X] = ?$

$10p + (1-p)25$

$\overline{X_n}$

2 RVs have the same distribution





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

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

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

Three experiments of 2 students

Report the sum of random number each finds after rolling a fair 4-die.

① each roll once, then add them.

② one of them rolls twice, then add them.

③ One of rolls once, then times with 2.

X Y

① $X+Y$

② X_1+X_2 ③ $2\cdot X_1$

Markov's inequality

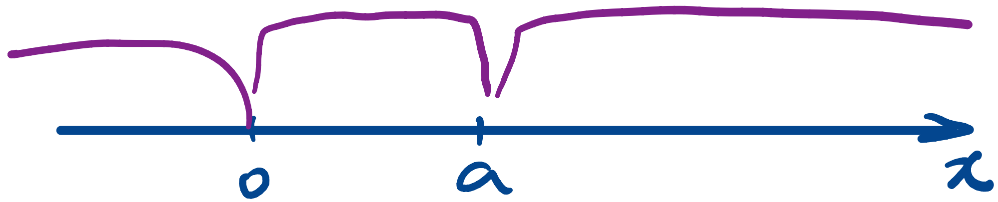
- For any random variable X that *only* takes $x \geq 0$ and constant $a > 0$

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

- 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



given: $x \geq 0 \quad a > 0$

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

$$\text{RHS} \geq 0 + \sum_{x\in[a,\infty)} x\,p(x) \geq \sum_{x\in[a,\infty)} a\,p(x)$$

$$= a \sum_{a\leq x<\infty} p(x)$$

$$= a\,p(X \geq a)$$

$$E[X] \geq a\,p(X \geq a)$$

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

Chebyshev's inequality

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

generic RV

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

- If we let a = kσ where σ = std[X] $= \sqrt{var[X]}$

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

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

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

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

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

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

Proof of Chebyshev's inequality

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

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

$P(X^2 < 9)$

$= P(|x| < 3)$

- 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 $a > 0$

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

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 $X_1, \ldots, X_N$.
- 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

x_1

x_2

$\vdots$

x_n

$\overline{X_n}$

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**

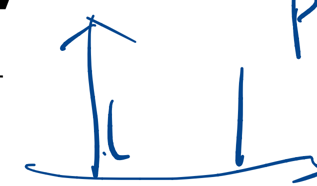
$$\overline{\mathbf{X}} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Expected value of sample mean of IID random variables

- By linearity of expected value

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

Pdf



$X_1 \quad X_2 \quad - - \quad X_N$

$E[X_1] = E[X_2] \cdots E[X_n]$

$= \frac{1}{N} \cdot N \cdot E[X_i]$

$= E[X]$

Expected value of sample mean of IID random variables

- By linearity of expected value

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

- 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

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

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

if X, Y are indep't

$var[kX] = k^2 var[X]$

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

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

$$= \frac{1}{N^2} \cdot N \cdot Var[X]$$

$$= \frac{1}{N} \cdot var[X]$$

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

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

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

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

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

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

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

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

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***.

$f_n \to c$

$N > n$

- For any positive number $\epsilon > 0$

ε

c

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

- 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

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

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

$$var[\bar{X}] = \frac{var[X]}{N}$$

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

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

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

if $N \to \infty$

RHS $\to 0$

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

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

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

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

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

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.

RV $X = \begin{cases} 1 & \text{if any E of interest happens} \\ 0 & \text{oW} \end{cases}$

$$E[X] = 1 \times P(\text{any Event of interest})$$

$$+ 0 \times P(\text{oW})$$

$$= P(\text{any Event of interest})$$

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:

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

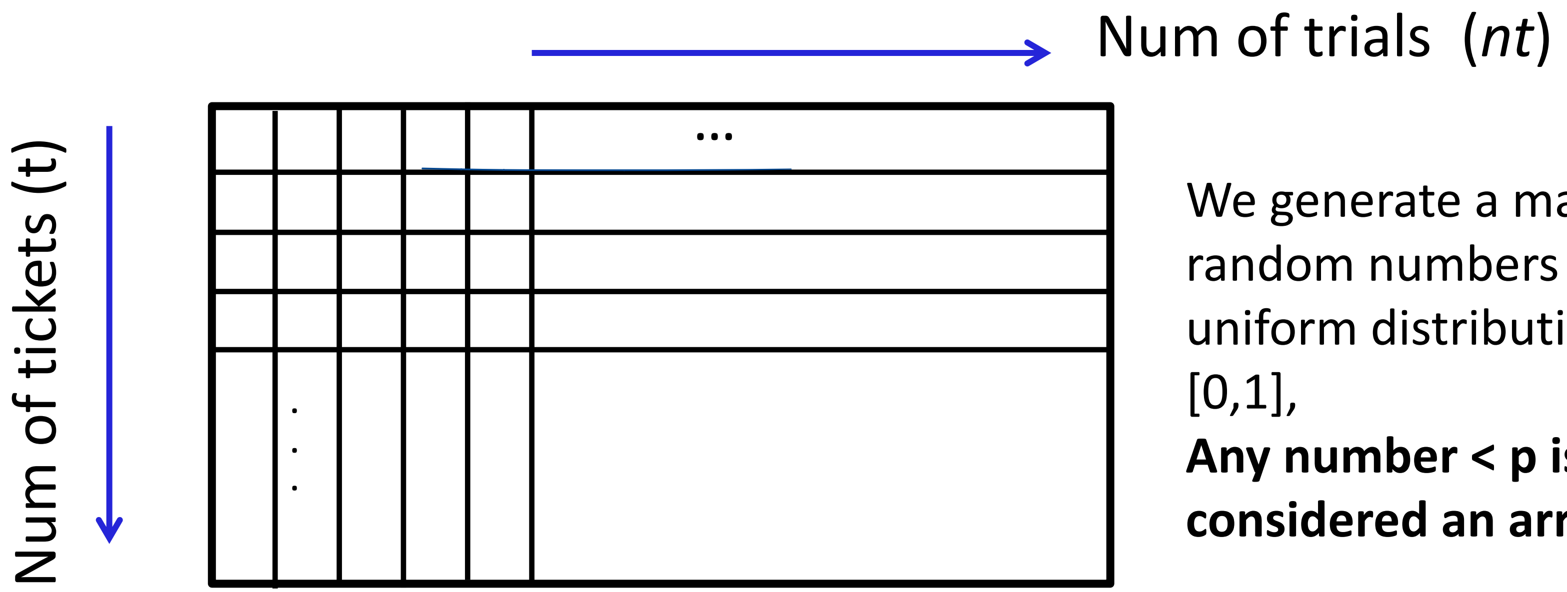
- Expected value of the number of ticketholders not flying

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

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

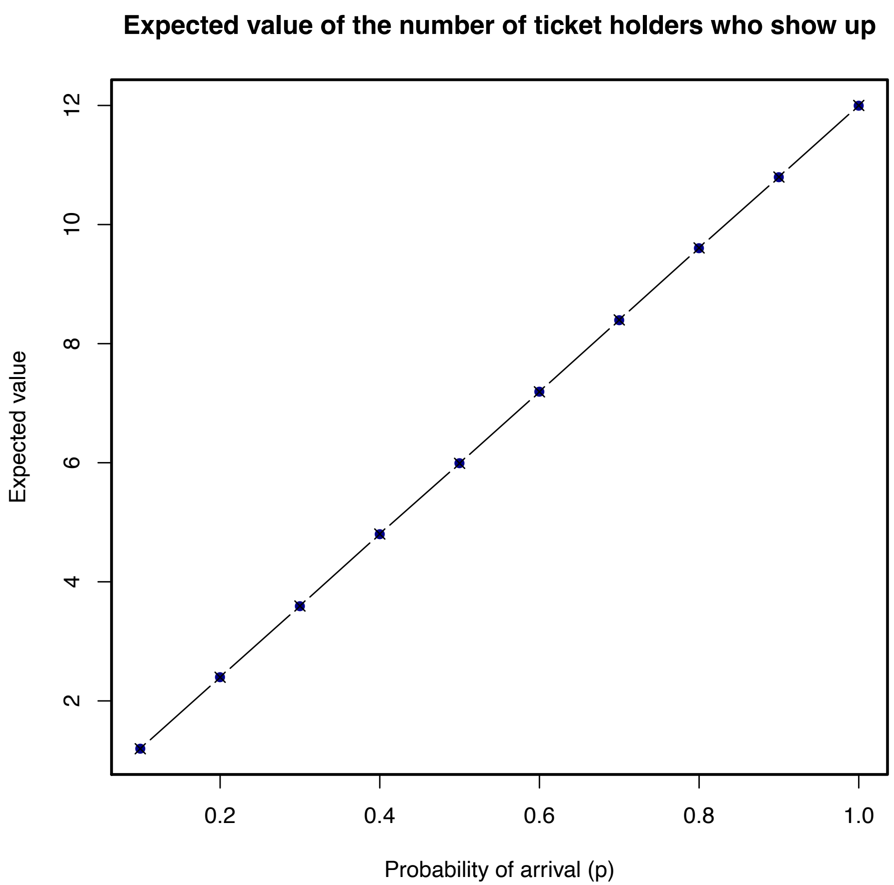


We generate a matrix of random numbers from uniform distribution in [0,1],
Any number < p is considered an arrival

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

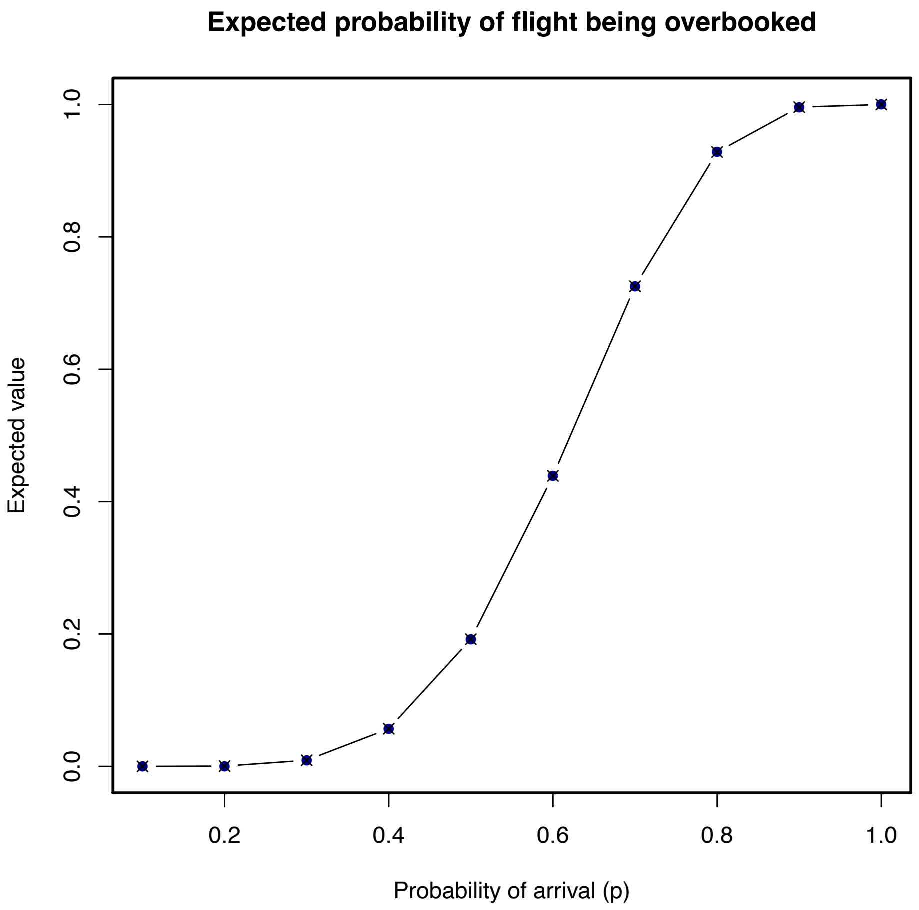
t= 12, s=7, p=0.1, 0.2, … 1.0

- **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

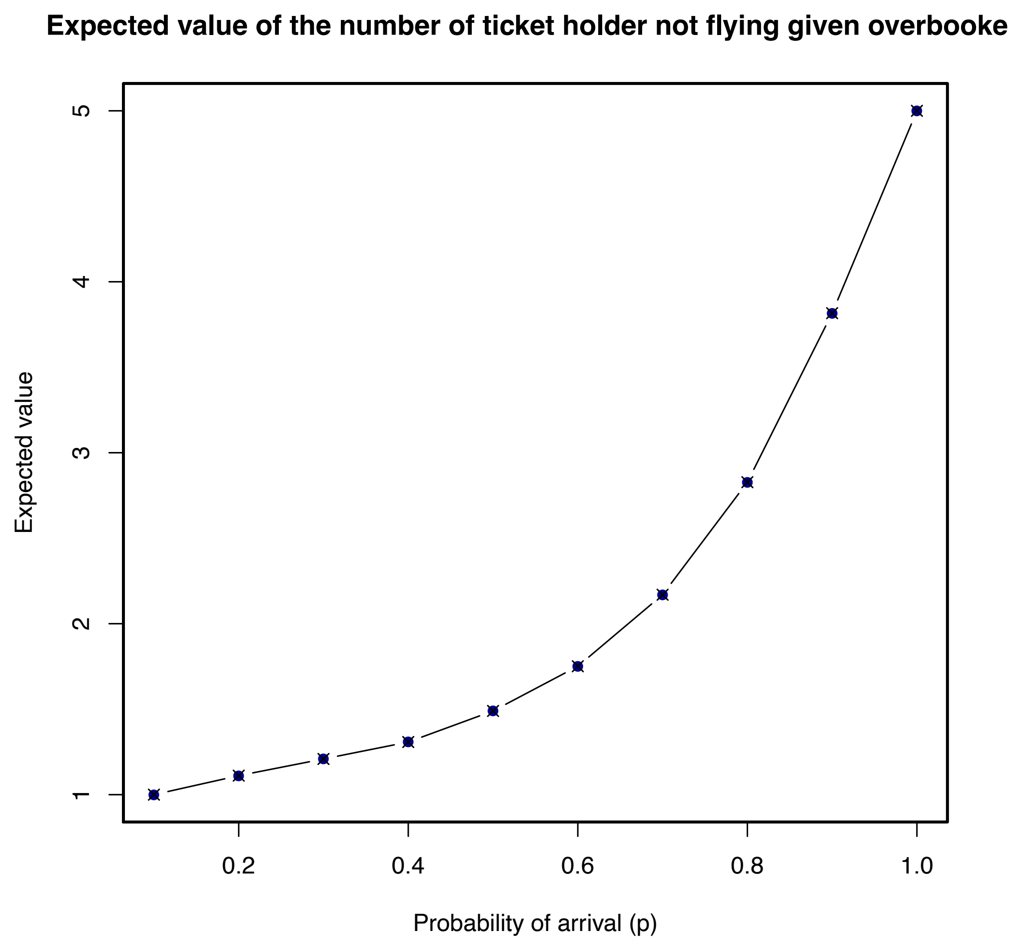
nt=100000,
t= 12, s=7,
p=0.1, 0.2, … 1.0



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
- 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!