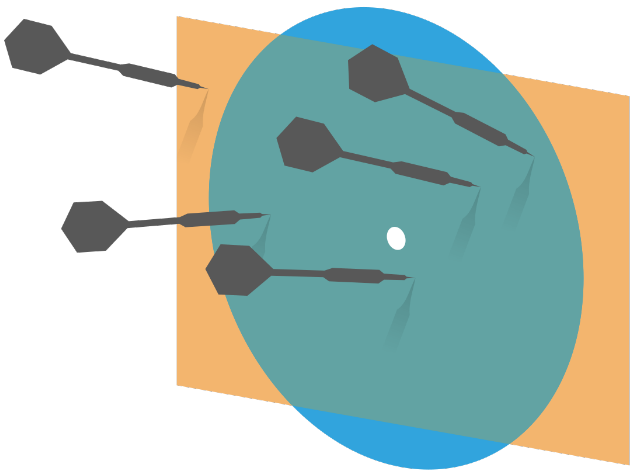
Probability and Statistics for Computer Science



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

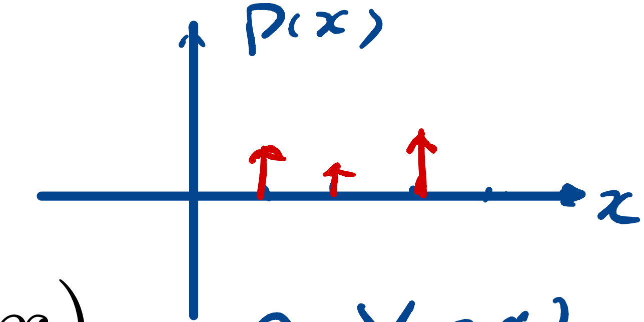
- Random Variable
  - *Expected value*
  - *Variance*

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

$0 \leq P(x) \leq 1$

$p(X=x)$

$p(x)$

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

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

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

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

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

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

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

- 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)

Y \ X	0	1
1	0	1
0	0	0

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

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.

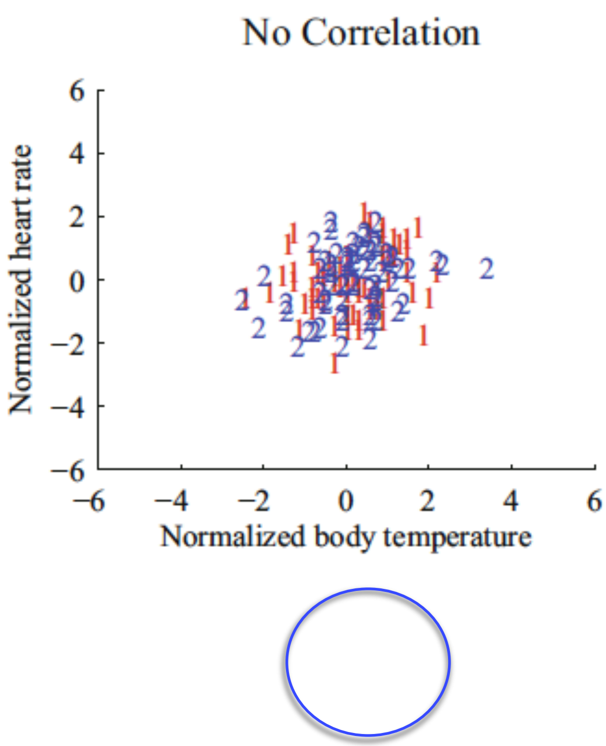
Correlation coefficient is normalized covariance

- The correlation coefficient can also be written as:

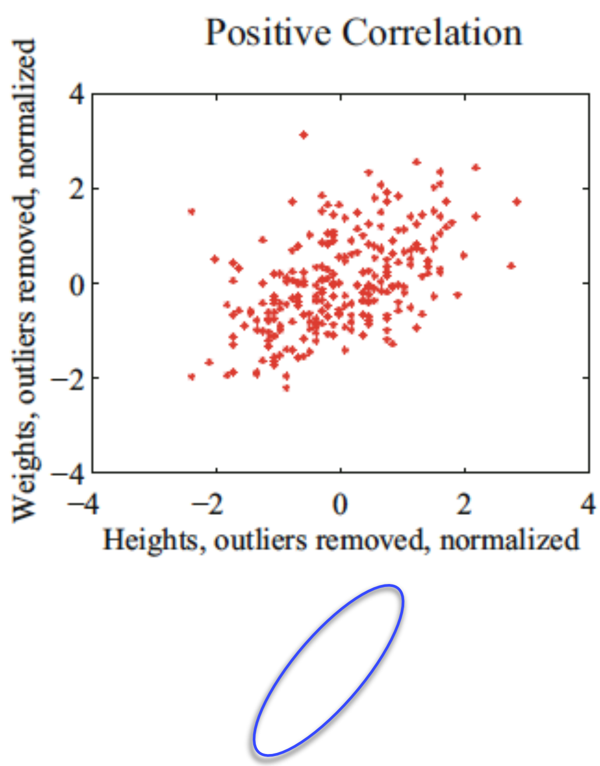
$$corr(X,Y) = \frac{E[XY] - E[X]E[Y]}{\sigma_X \sigma_Y}$$

Correlation seen from scatter plots

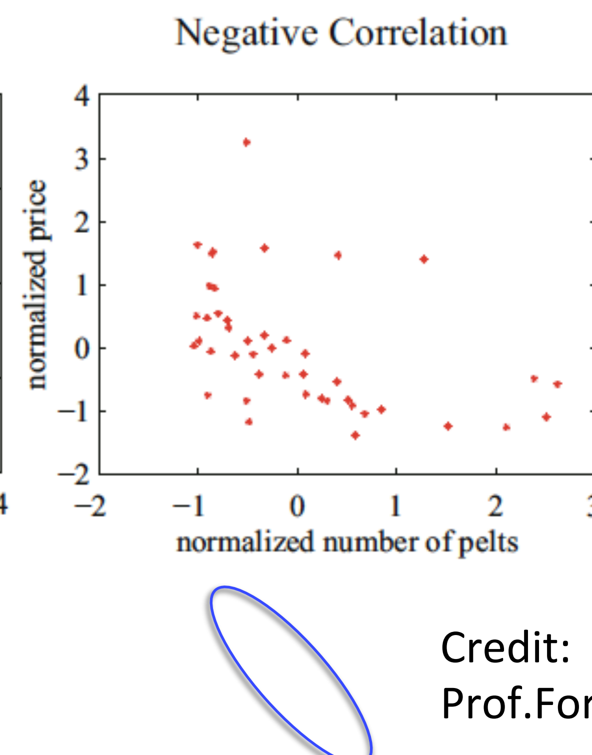
Zero Correlation



Positive correlation



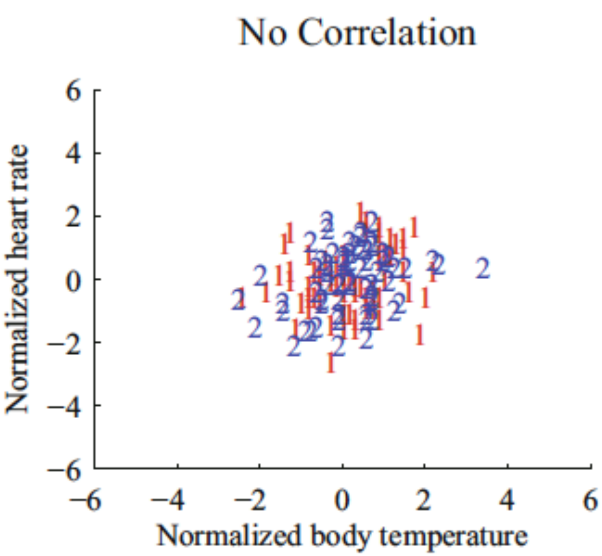
Negative correlation

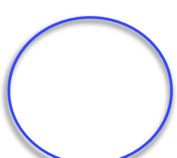


Credit: Prof.Forsyth

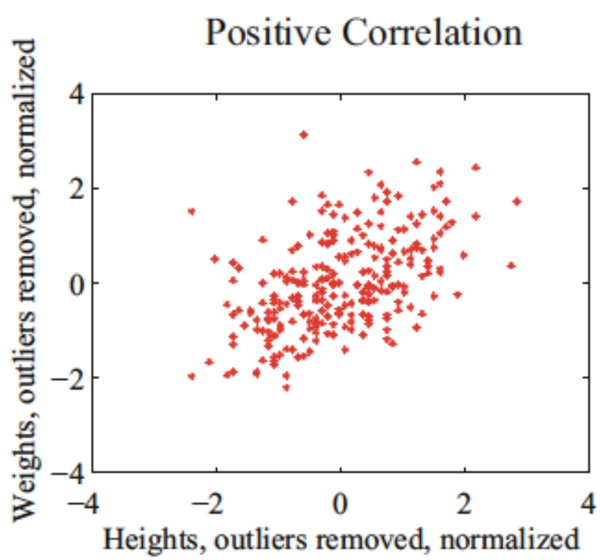
Covariance seen from scatter plots

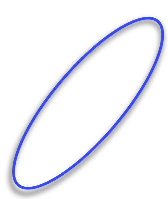
Zero
Covariance



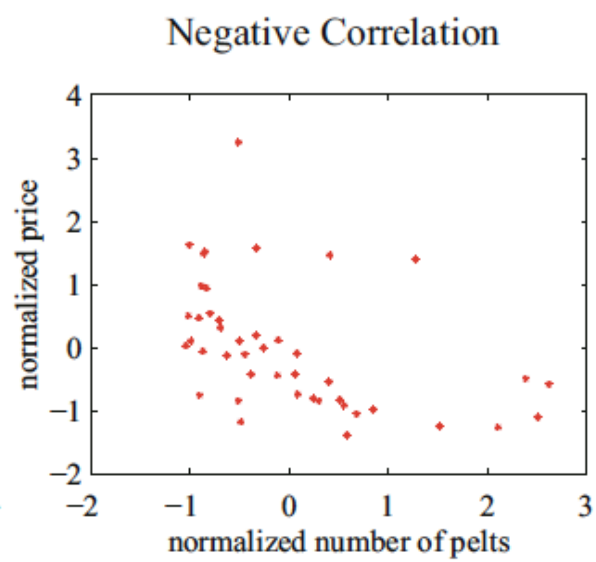


Positive
Covariance





Negative
Covariance



Credit:
Prof.Forsyth

When correlation coefficient or covariance is zero

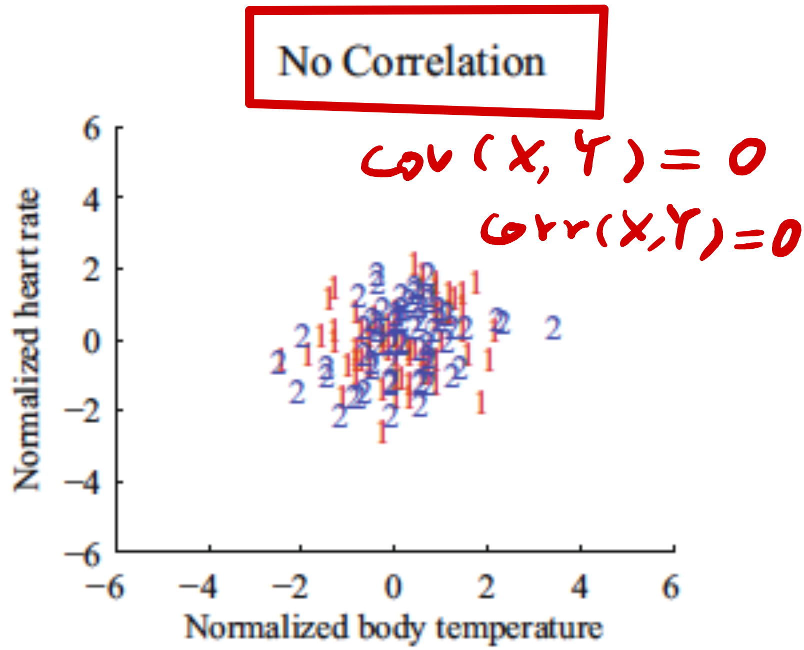
- The covariance is 0!

- 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

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

HW extra
points

If ~~events~~ 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.

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

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

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

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

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

These are equivalent! Uncorrelatedness

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

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

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

  $= 0$

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

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

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

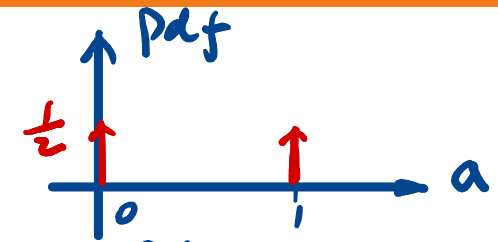
Q: What is this expectation?

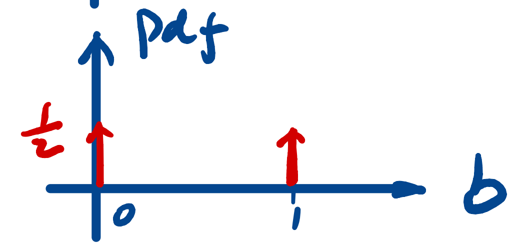
- We toss two Fair identical coins A & B independently for three times and 4 times respectively, for each head we earn \$1, we define X is the earning from A and Y is the earning from B. What is E[XY]?

A. 2 B. 3 C. 4

$$X = A_1 + A_2 + A_3$$

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





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

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

$\because$ 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

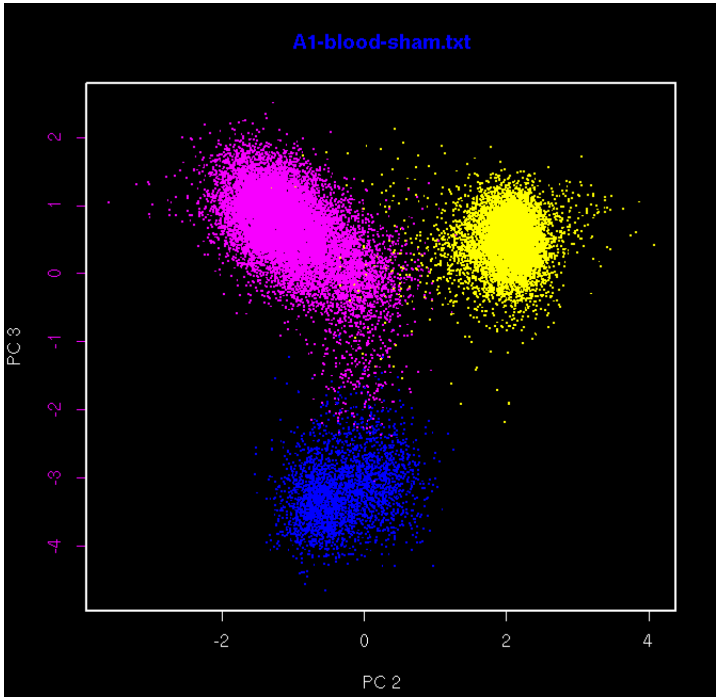
$E[Y] = ?$ $\frac{2}{3}$

$E[XY] = ?$ 0

X, Y are depedent!! but uncorrelated

Covariance example

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

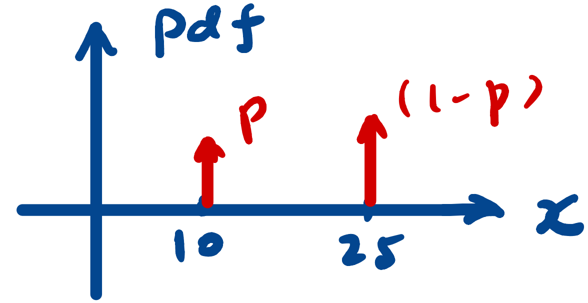


Random Variable Example for WLLN

Money box

10¢ 25¢

dime quarter

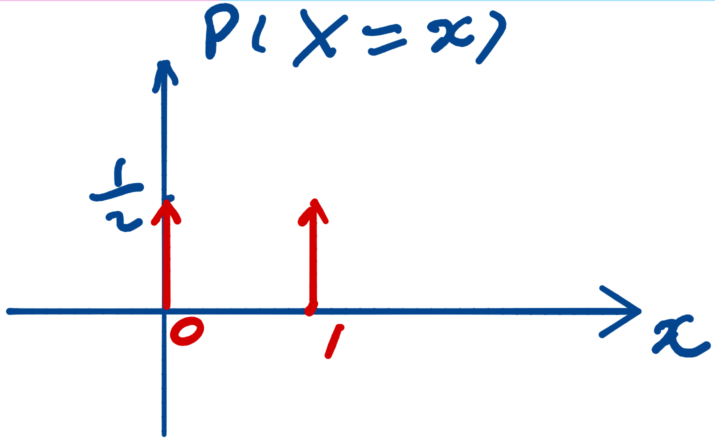


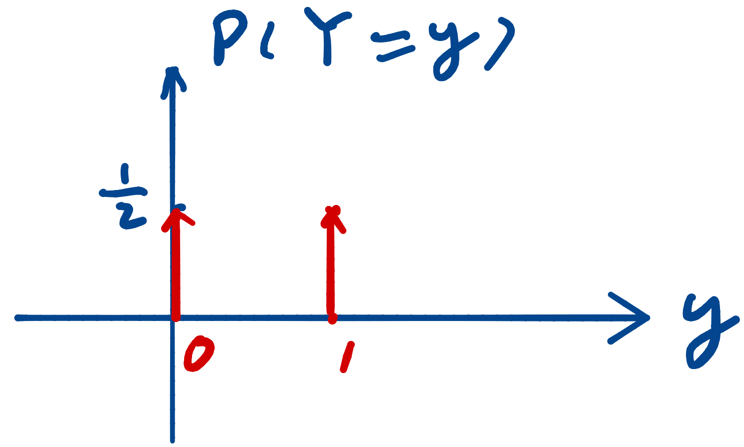
* shake and take one and put back

$E[X] = ?$

$10p + (1-p)25$

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

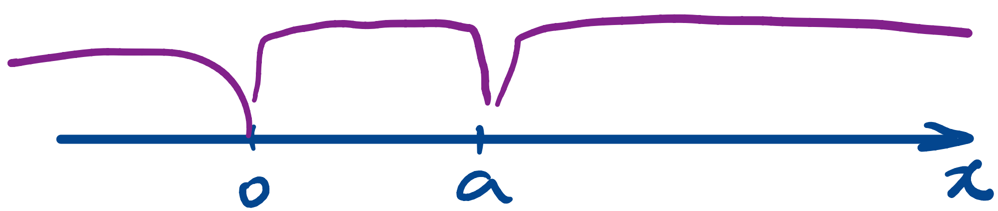
- For any random variable X that *only* takes $x \geq 0$ and constant *a* > 0

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

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

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

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

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

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

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

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

Chebyshev's inequality

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

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

- 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

$\omega > 0$

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

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

Proof of Chebyshev's inequality

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

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

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

$= var[X]$

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

$LHS = P(|X-E[X]| \geq a)$ $w = a^2$

$RHS = \frac{var[X]}{a^2}$

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 $a > 0$

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

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 $P(x)$

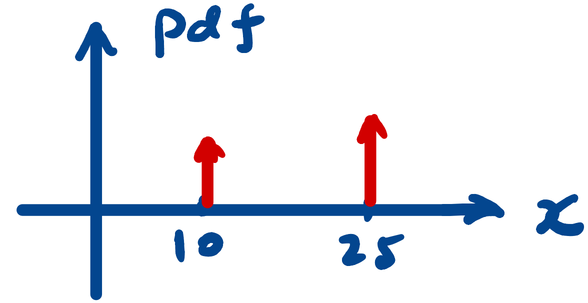
  then the numbers randomly generated from them are called **IID** samples
- The **sample mean** is a random variable

Random Variable Example

Money box

10¢ 25¢

dime quarter



* shake and take one and put back.

X_1 takes $x_1 = 10$

X_2 takes $x_2 = 10$

X_3 takes $x_3 = 25$

$\vdots$

X_N takes $x_N = 10$

$E[X] = ?$

$$\bar{X} = \frac{\sum X_i}{N}$$

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**

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

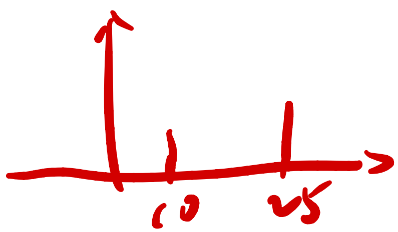
$E[\bar{x}]$

$var[\bar{x}]$

Expected value of sample mean of IID random variables

- By linearity of expected value

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



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

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

$$= E[X]$$

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 $P(x)$ → the same E[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

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

$var[kx] = k^2 var[x]$

X_i are indpt to X_j

$\Rightarrow cov(X_i, X_j) = 0 \quad i \neq j$

$$var[X_1 + X_2] = var[X_1] + var[X_2] + 2cov(X_1, X_2) \xrightarrow{} 0$$

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

Generalized

$$\Rightarrow var[\sum_{i=1}^{N} X_i] = \sum_{i=1}^{N} var[X_i] = N \cdot var[X]$$

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

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

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

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

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***.

- 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

$\varepsilon > 0$

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

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

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

$$P(|\bar{x} - E[x]| \geq \varepsilon) \leq \frac{var[x]}{N \cdot \varepsilon^2}$$

$N \to \infty$

$$\lim_{N \to \infty} LHS = 0$$

$\varepsilon \cdot E[x]$

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

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

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

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

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.

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

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!