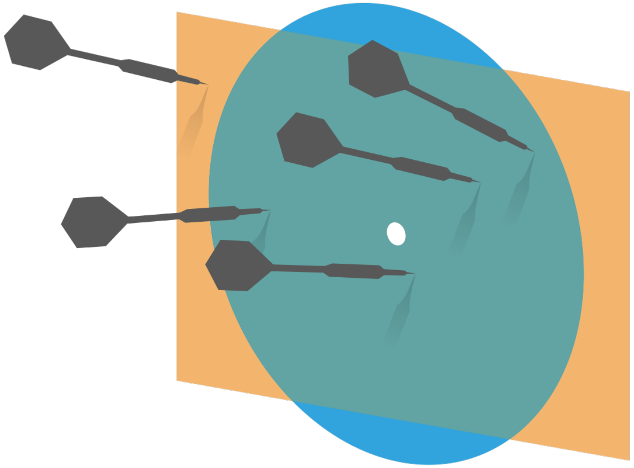
Probability and Statistics for Computer Science



Credit: wikipedia

"I have now used each of the terms mean, variance, covariance and standard deviation in two slightly different ways." ---Prof. Forsythe

Hongye Liu, Teaching Assistant Prof, CS361, UIUC, 2.13.2020

Last time

- Random Variable
  - The Definition of Random Variable
  - *Probability distribution* of random variable
  - *Joint & Conditional probability distribution* of random variable

Content

- Random Variable
  - *Expected value*
  - *Variance & covariance*
  - *Markov’s inequality*

Three important facts of Random variables

- Random variables have **probability functions**
- Random variables can be **conditioned** on events or other random variables
- Random variables have **averages**

Content

- Random Variable
  - ***Expected value***
  - *Variance & covariance*
  - *Markov’s inequality*

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 the values X can take

Expected value

- The **expected value** of a random variable X is

<= 1

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

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

Expected value: profit

- A company has a project that has ***p*** probability of earning 10 million and ***1-p*** probability of losing 10 million.
- Let X be the return of the project.

Expected value as mean

- Suppose we have a data set $\{x_i\}$ of N data points. Let's define a random variable X taking on each of the data points with equal probability $1/N$.

$$E[X] = \sum_i x_i P(x_i) = \frac{1}{N} \sum_i x_i = mean(\{x_i\})$$

- The expected value is also called the mean.

Linearity of Expectation

- For random variables X and Y and constants k,c
  - Scaling property
    $$E[kX] = kE[X]$$
  - Additivity
    $$E[X + Y] = E[X] + E[Y]$$
  - And $E[kX + c] = kE[X] + c$

Linearity of Expectation

- Proof of the additive property

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

Q. What's the value?

- What is $E[E[X] + 1]$?

A. E[X]+1 B. 1 C. 0

Expected value of a function of X

- If $\boldsymbol{f}$ is a function of a random variable X, then $Y = \boldsymbol{f}(X)$ is a random variable too
- The expected value of $Y = \boldsymbol{f}(X)$ is

Expected value of a function of X

- If $\boldsymbol{f}$ is a function of a random variable X, then $Y = \boldsymbol{f}(X)$ is a random variable too

- The expected value of $Y = \boldsymbol{f}(X)$ is

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

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?

Q: Is this statement true?

If there exists a constant such that $\boldsymbol{P}(X \geq a) = 1$, then $E[X] \geq a$. It is:

A. True
B. False

Content

- Random Variable
  - *Expected value*
  - ***Variance & covariance***
  - *Towards the week law of large numbers*

Variance and standard deviation

- The variance of a random variable X is

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

- The standard deviation of a random variable X is

$$std[X] = \sqrt{var[X]}$$

Properties of variance

- For random variable X and constant k

$$var[X] \geq 0$$

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

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:

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

A neater expression for variance

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

A neater expression for variance

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

$$var[X] = E[(X - \mu)^2] \quad where\ \mu = E[X]$$

A neater expression for variance

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

$$var[X] = E[(X - \mu)^2] \quad where\ \mu = E[X]$$

$$= E[X^2 - 2X\mu + \mu^2]$$

A neater expression for variance

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

$$\begin{aligned} var[X] &= E[(X - \mu)^2] \quad where\ \mu = E[X] \\ &= E[X^2 - 2X\mu + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2E[X]E[X] + (E[X])^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Variance: the profit example

- For the profit example, what is the variance of the return? We know E[X]= 20p-10

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

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.

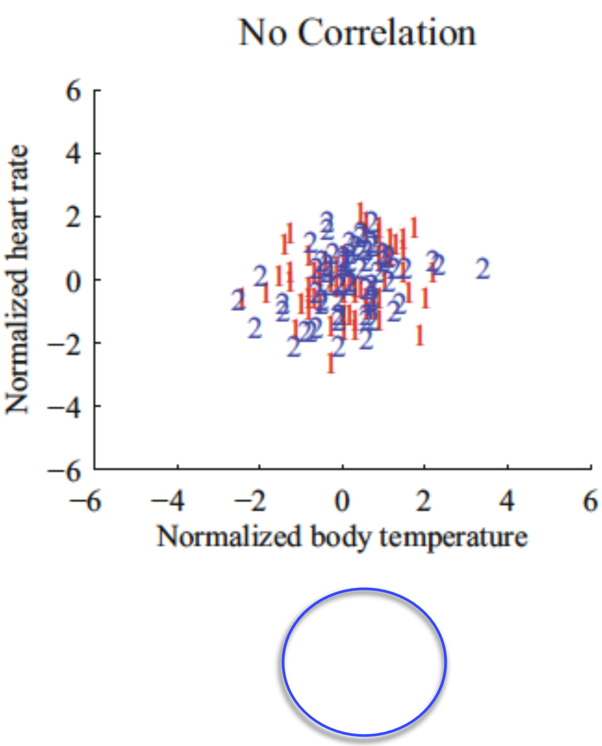
Correlation coefficient is normalized covariance

- The correlation coefficient can also be written as:

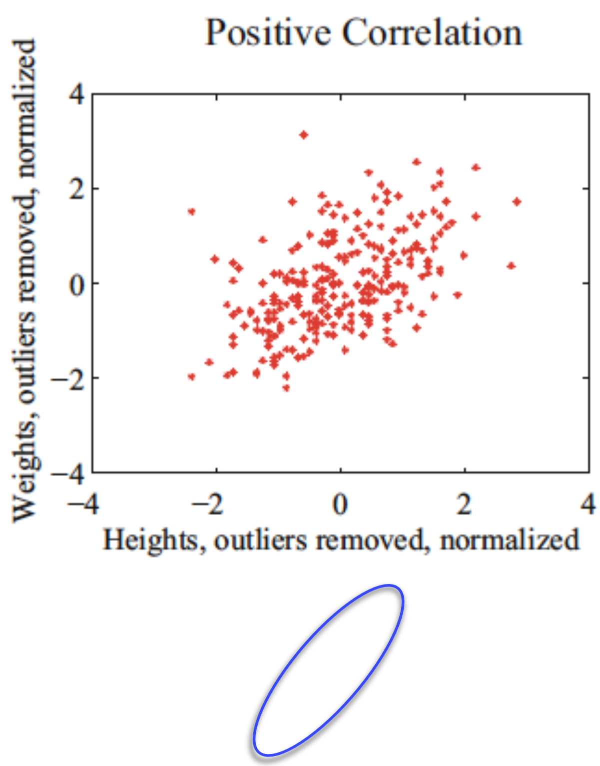
$$corr(X,Y) = \frac{E[XY] - E[X]E[Y]}{\sigma_X \sigma_Y}$$

Correlation seen from scatter plots

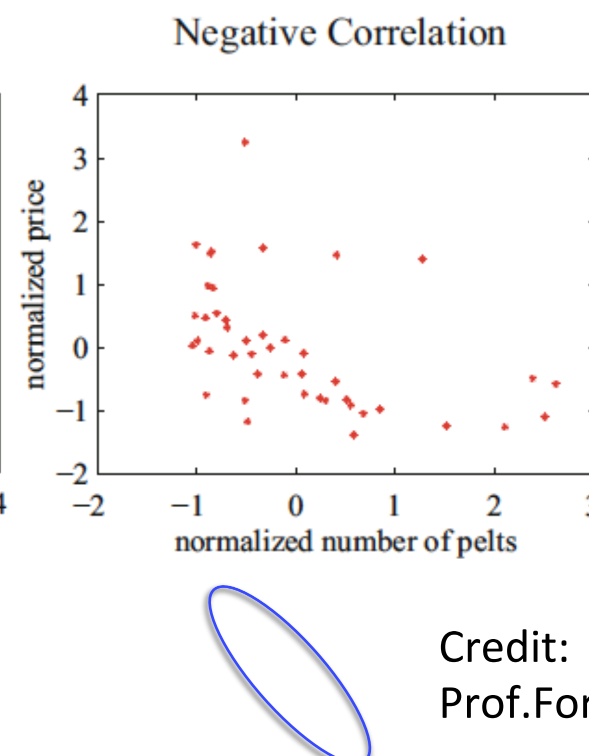
Zero Correlation



Positive correlation



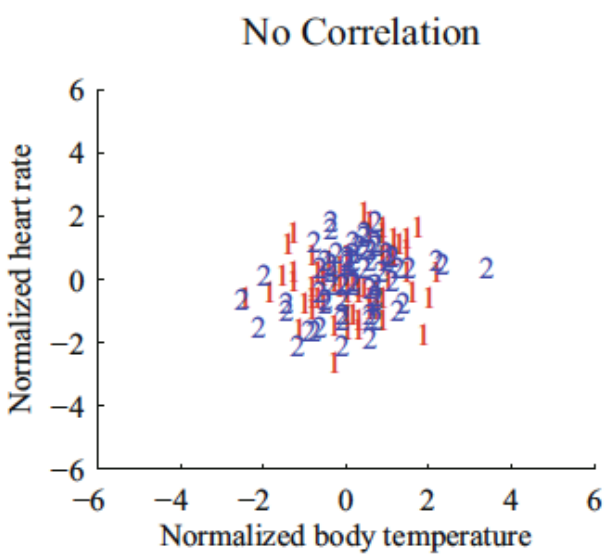
Negative correlation

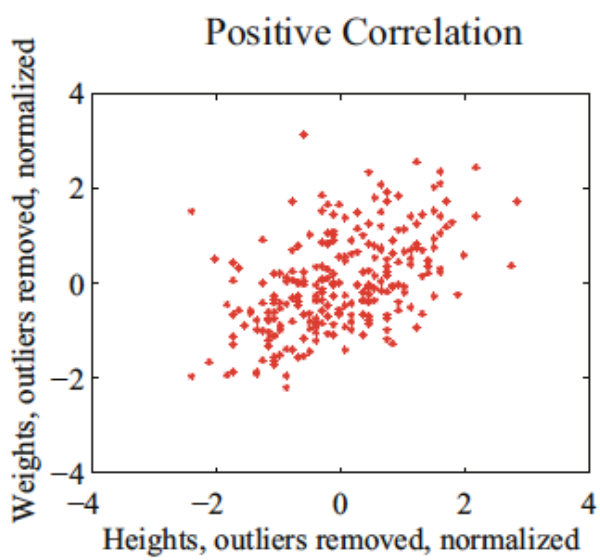


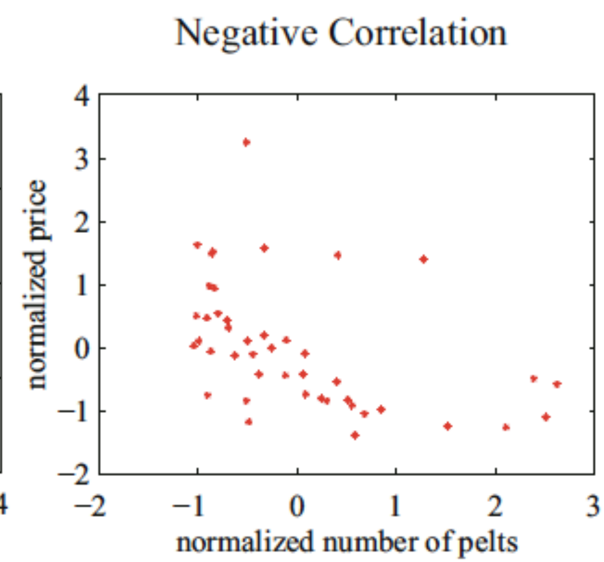
Credit: Prof.Forsyth

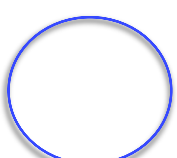
Covariance seen from scatter plots

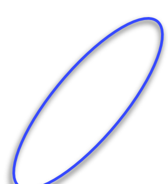
Zero **Covariance**	Positive **Covariance**	Negative **Covariance**











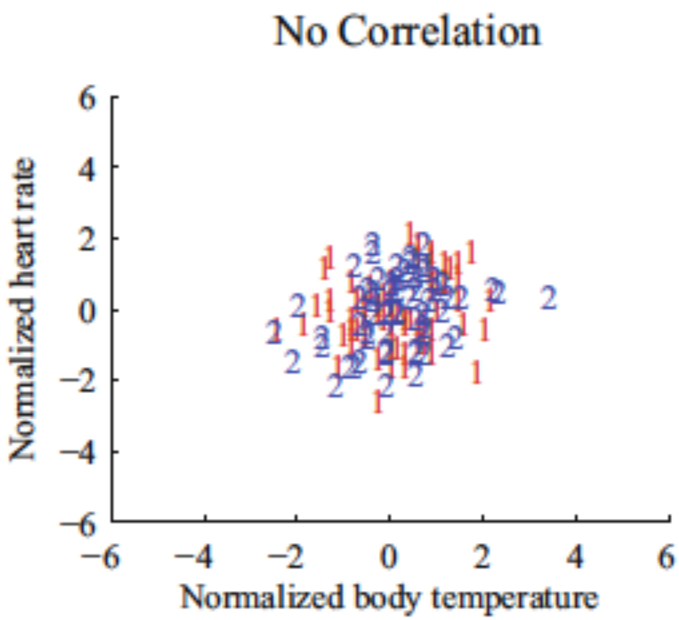
Credit: Prof.Forsyth

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

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

Properties of independence in terms of expectations

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

Proof of independence in terms of expectation (1)

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

$$E[XY] = \sum_{(x,y)\in D_x \times D_y} xyP(X = x \cap Y = y)$$

$$= \sum_{x\in D_x} \sum_{y\in D_y} (xyP(x, y))$$

$$= \sum_{x\in D_x} \sum_{y\in D_y} (xyP(x)P(y))$$

$$= \sum_{x\in D_x} xP(x) \sum_{y\in D_y} yP(y)$$

$$= (\sum_{x\in D_x} xP(x))(\sum_{y\in D_y} yP(y))$$

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

Properties of independence in terms of expectations

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

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

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

Q: What is this expectation?

- We toss two 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)?

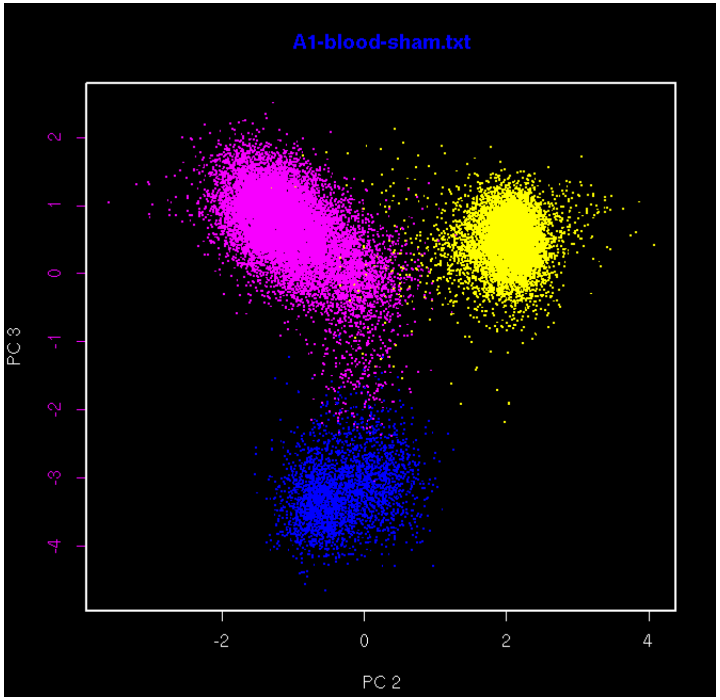
A. \$2 B. \$3 C. \$4

Group Discussion

- 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



Content

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

Markov inequality

- The inequality that was the foundation of many probabilistic theories
- Discovered by Andrei Markov who also invented Markov Chain model

Andrei Markov
1856 - 1922

Indicator function

- An indicator function for an event A is a function of x such that

- The expected value of the indicator function is the probability of event A

$$E[\mathbb{I}_{[A]}(x)] = ?$$

Indicator function

- An indicator function for an event $\boldsymbol{A}$ is a function of x such that

$$\mathbb{I}_{[\boldsymbol{A}]}(x) = \begin{cases} 1 & event\ occurs\ for\ the\ value\ x \\ 0 & otherwise \end{cases}$$

- The expected value of the indicator function is the probability of event $\boldsymbol{A}$

$$\mathrm{E}[\,\mathbb{I}_{[\boldsymbol{A}]}\,(x)] = ?$$

Indicator functions

- An indicator function for an event $\boldsymbol{A}$ is a function of x such that

$$\mathbb{I}_{[\boldsymbol{A}]}(x) = \begin{cases} 1 & event\ occurs\ for\ the\ value\ x \\ 0 & otherwise \end{cases}$$

- The expected value of the indicator function is the probability of event $\boldsymbol{A}$

Indicator functions

- An indicator function for an event $\boldsymbol{A}$ is a function of x such that

$$\mathbb{I}_{[\boldsymbol{A}]}(x) = \begin{cases} 1 & event\ occurs\ for\ the\ value\ x \\ 0 & otherwise \end{cases}$$

- The expected value of the indicator function is the probability of event $\boldsymbol{A}$

$$\boxed{\mathrm{E}[\,\mathbb{I}_{[A]}\,(x)] = 1\times \mathrm{P}(A)+0\times(1-\mathrm{P}(A)) = \mathrm{P}(A)}$$

Markov's inequality

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

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

- So, a random variable is unlikely to have the absolute value much larger than the mean of its absolute value

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

$$P(|X| \geq 10E[|X|]) \leq 0.1$$

Proof of Markov's inequality

$$\mathbb{I}_{[|X|\geq a]}(X) = \begin{cases} 1 & if\ |X| \geq a \\ 0 & otherwise \end{cases}$$

Proof of Markov's inequality

$$\mathbb{I}_{[|X| \geq a]}(X) = \begin{cases} 1 & if\ |X| \geq a \\ 0 & otherwise \end{cases}$$

$$\leq \frac{|X|}{a}$$

Proof of Markov's inequality

$$\mathbb{I}_{[|X| \geq a]}(X) = \begin{cases} 1 & if\ |X| \geq a \\ 0 & otherwise \end{cases}$$

$$\leq \frac{|X|}{a}$$

$$E[\mathbb{I}_{[|X| \geq a]}(X)] \leq \frac{E[|X|]}{a}$$

Proof of Markov's inequality

$$\mathbb{I}_{[|X|\geq a]}(X) = \begin{cases} 1 & if\ |X| \geq a \\ 0 & otherwise \end{cases}$$

$$\leq \frac{|X|}{a}$$

$$E[\mathbb{I}_{[|X|\geq a]}(X)] \leq \frac{E[|X|]}{a}$$

$\downarrow$

LHS =

Proof of Markov's inequality

$$\mathbb{I}_{[|X| \geq a]}(X) = \begin{cases} 1 & if\ |X| \geq a \\ 0 & otherwise \end{cases}$$

$$\leq \frac{|X|}{a}$$

$$E[\mathbb{I}_{[|X| \geq a]}(X)] \leq \frac{E[|X|]}{a}$$

↓

LHS = $P(|X| \geq a)$

Proof of Markov's inequality

$$\mathbb{I}_{[|X| \geq a]}(X) = \begin{cases} 1 & if\ |X| \geq a \\ 0 & otherwise \end{cases}$$

$$\leq \frac{|X|}{a}$$

$$E[\mathbb{I}_{[|X| \geq a]}(X)] \leq \frac{E[|X|]}{a}$$

$\downarrow$

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

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

Assignments

- Finish Chapter 4 of the textbook
- Next time: Proof of Chebyshev inequality & Weak law of large numbers, Continuous random variable

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!