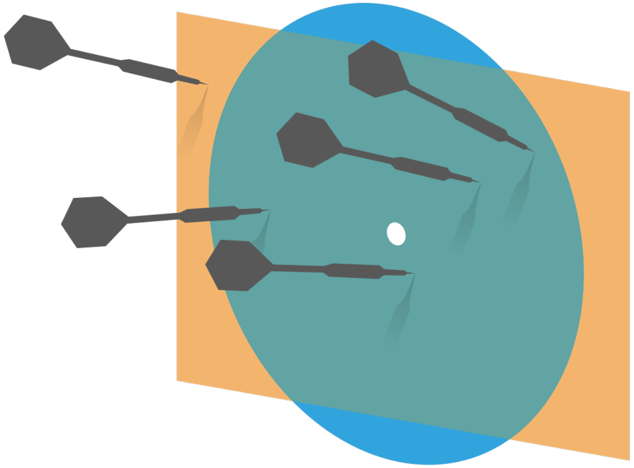
Probability and Statistics for Computer Science



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, 2.18.2021

Objectives

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

Conditional probability distribution of random variables

- The conditional probability distribution of X *given* Y is

$$P(x|y) = \frac{P(x,y)}{P(y)} \qquad P(y) \neq 0$$

Conditional probability distribution of random variables

- The conditional probability distribution of X *given* Y is

$$P(x|y) = \frac{P(x,y)}{P(y)} \quad P(y) \neq 0$$

- The joint probability distribution of two random variables ***X*** and ***Y*** is

$$P(\{X = x\} \cap \{Y = y\})$$

$$\sum_x P(x|y) = 1$$

Get the marginal from joint distri.

- We can recover the individual probability distributions from the joint probability distribution

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

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

Joint probabilities sum to 1

- The sum of the joint probability distribution

$$\sum_y \sum_x P(x, y) = 1$$

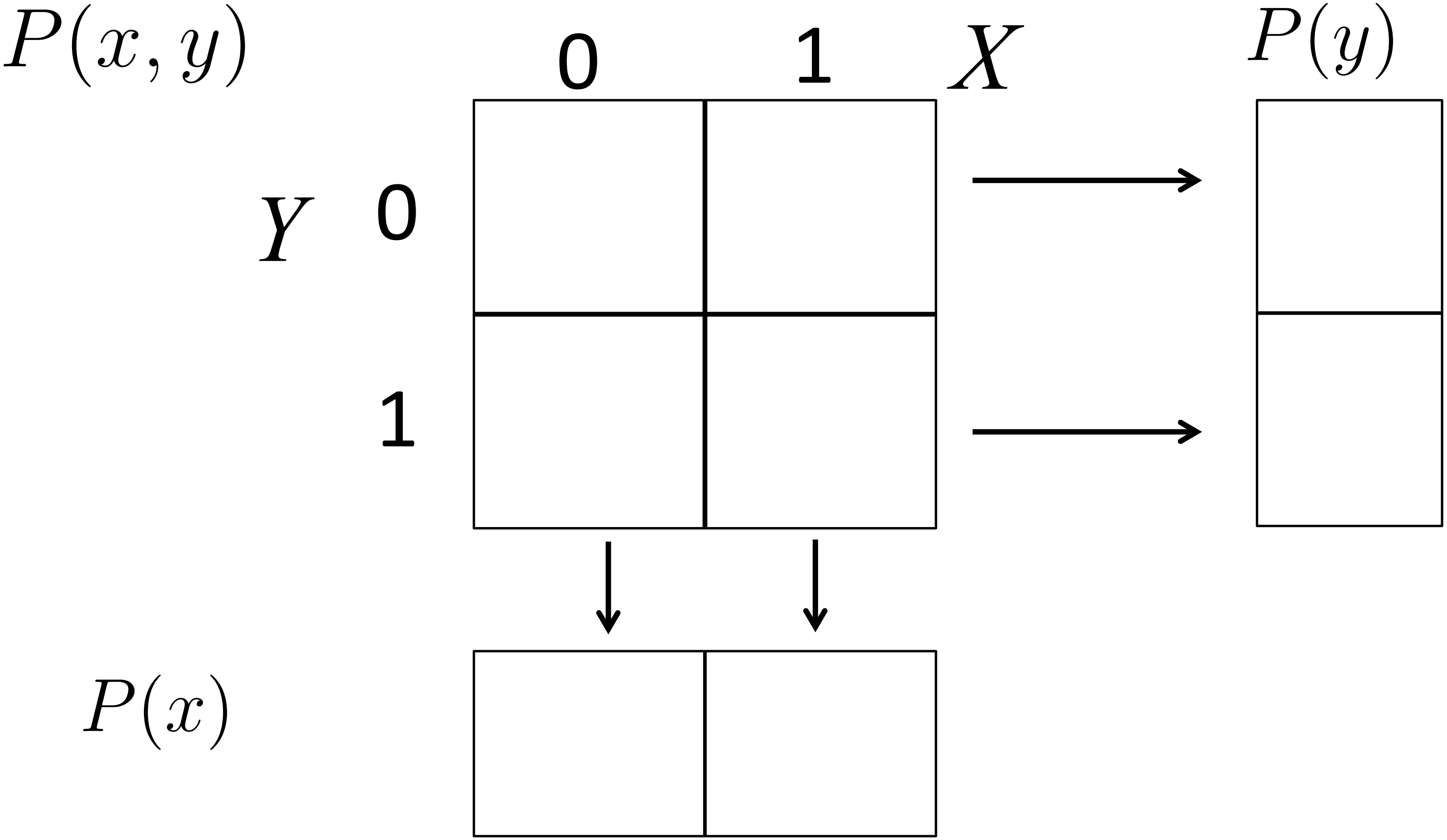
Joint Probability Example

- Tossing a coin twice, we define random variable $X\ and\ Y$ for each toss.

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

$$Y(\omega) = \begin{cases} 1 & outcome\ of\ \omega\ is\ head \\ 0 & outcome\ of\ \omega\ is\ tail \end{cases}$$

Joint probability distribution example



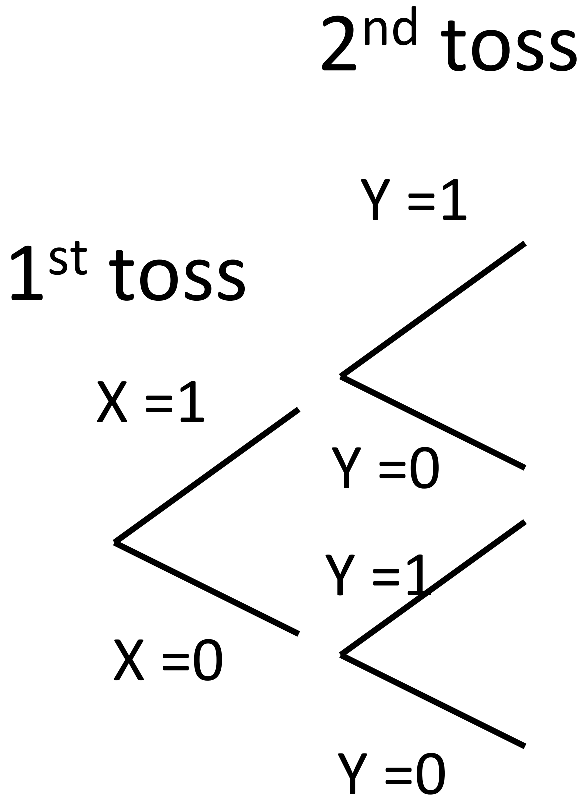
Joint Probability Example

Now we define Sum ***S*** = $X + Y$, Difference ***D*** = $X - Y$. ***S*** takes on values {0,1,2} and ***D*** takes on values {-1, 0, 1}

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

$$Y(\omega) = \begin{cases} 1 & outcome\ of\ \omega\ is\ head \\ 0 & outcome\ of\ \omega\ is\ tail \end{cases}$$

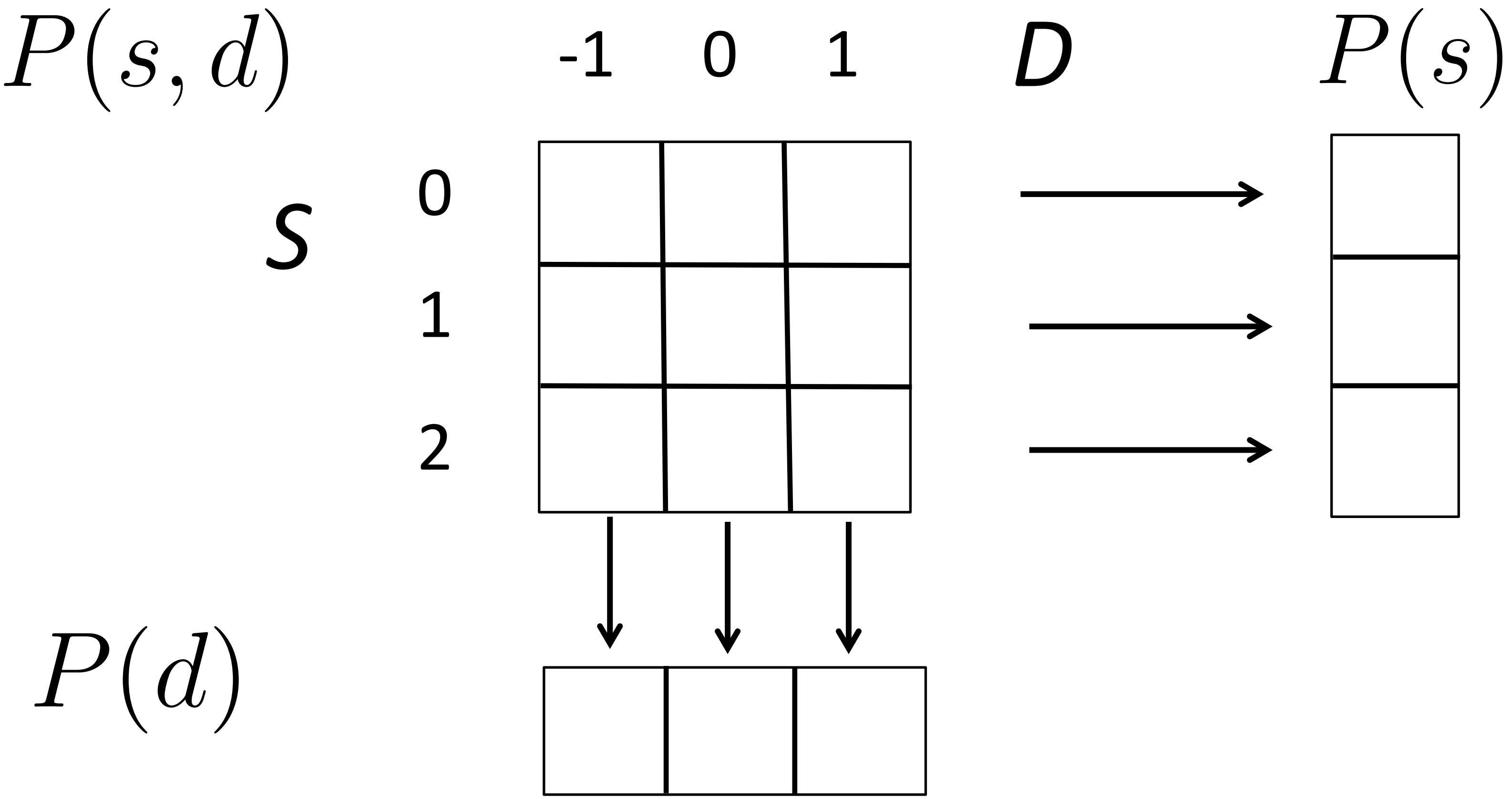
Joint Probability Example



S	D	P(s, d)
2	0	
1	1	
1	-1	
0	0	

Suppose coin is fair, and the tosses are independent

Joint probability distribution example



Independence of random variables

- Random variable $X\ and\ Y$ are independent if

$$P(x,y) = P(x)P(y)\ for\ all\ x\ and\ y$$

- In the previous coin toss example
  - Are X and Y independent?
  - Are ***S*** and ***D*** independent?

Joint probability distribution example

$P(x, y)$

Y \ X	0	1	$P(y)$
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

	0	1
$P(x)$	$\frac{1}{2}$	$\frac{1}{2}$

Joint probability distribution example

$P(s,d)$

S \ D	-1	0	1
0	0	$\frac{1}{4}$	0
1	$\frac{1}{4}$	0	$\frac{1}{4}$
2	0	$\frac{1}{4}$	0

$P(s)$

0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

$P(d)$

-1	0	1
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Joint probability distribution example

$P(s,d)$

S \ D	-1	0	1
0	0	$\frac{1}{4}$	0
1	$\frac{1}{4}$	0	$\frac{1}{4}$
2	0	$\frac{1}{4}$	0

$P(s)$

0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

$P(d)$

-1	0	1
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$$P(S=1,D=0) \neq P(S=1)P(D=0)$$

Conditional probability distribution example

$$P(s|d) = \frac{P(s,d)}{P(d)}$$

S \ *D*	-1	0	1
0	0	$\frac{1}{2}$	0
1	1	0	1
2	0	$\frac{1}{2}$	0

Bayes rule for random variable

- Bayes rule for events generalizes to random variables

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$\longrightarrow \quad P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

$$= \frac{P(y|x)P(x)}{\sum_x P(y|x)P(x)}$$

Total Probability

Conditional probability distribution example

$$P(s|d) = \frac{P(s,d)}{P(d)}$$

S \ *D*	-1	0	1
0	0	$\frac{1}{2}$	0
1	1	0	1
2	0	$\frac{1}{2}$	0

$$P(D=-1|S=1) = \frac{P(S=1|D=-1)P(D=-1)}{P(S=1)} = \frac{1 \times \frac{1}{4}}{\frac{1}{2}}$$

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**

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

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

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 $\boldsymbol{X}$ 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

Variance and standard deviation

- The variance of a random variable X is

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

- 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:

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

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

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

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

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

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.

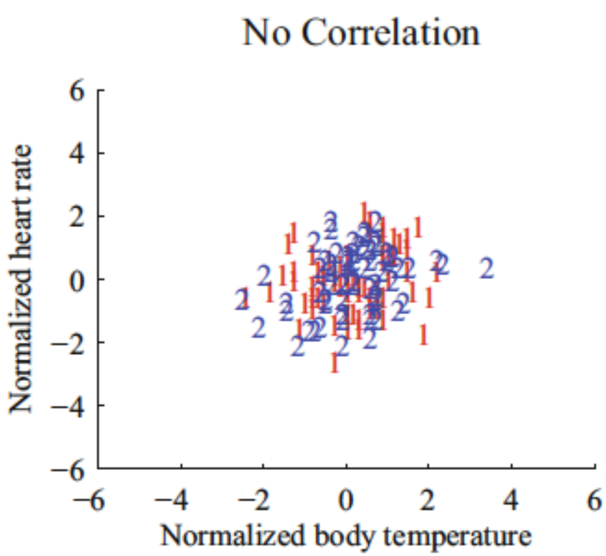
Correlation coefficient is normalized covariance

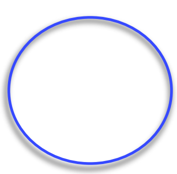
- The correlation coefficient can also be written as:

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

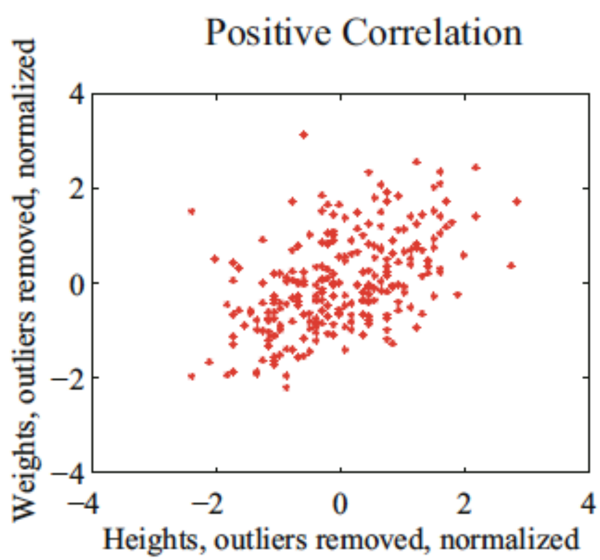
Correlation seen from scatter plots

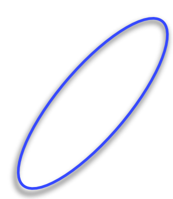
Zero Correlation



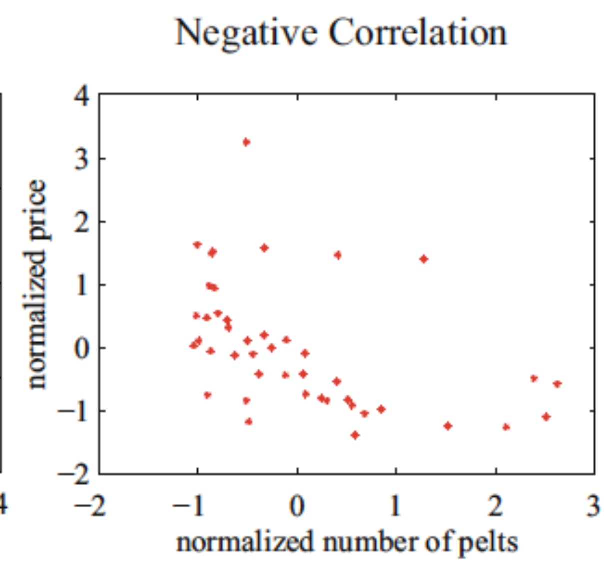


Positive correlation





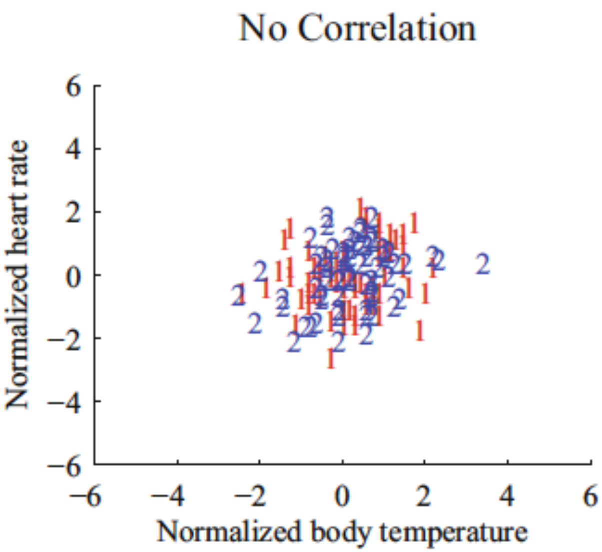
Negative correlation

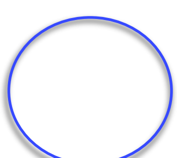


Credit: Prof.Forsyth

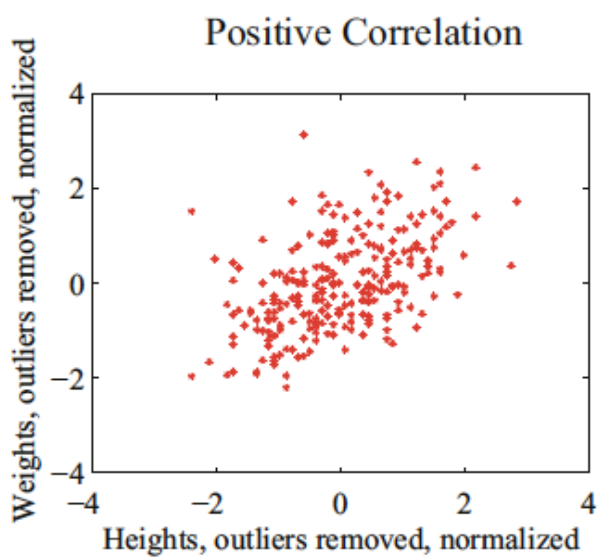
Covariance seen from scatter plots

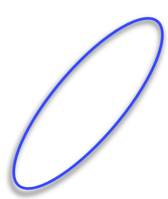
Zero Covariance ↓



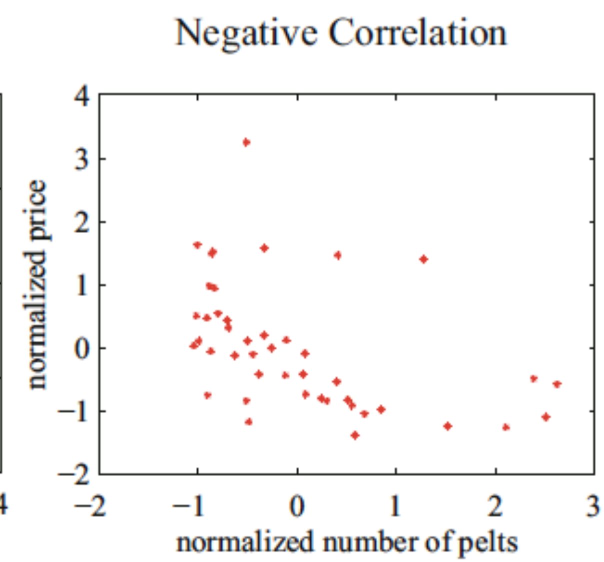


Positive Covariance ↓





Negative Covariance ↓



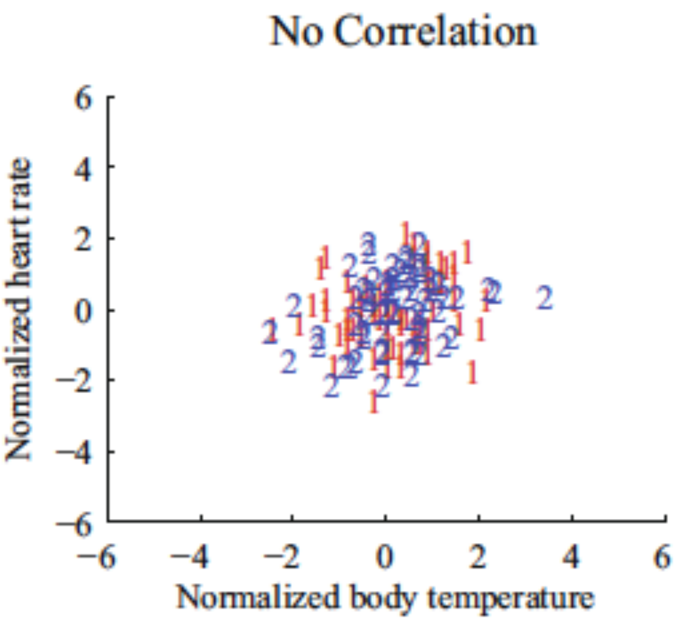
Credit: Prof.Forsyth

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

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

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 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)?

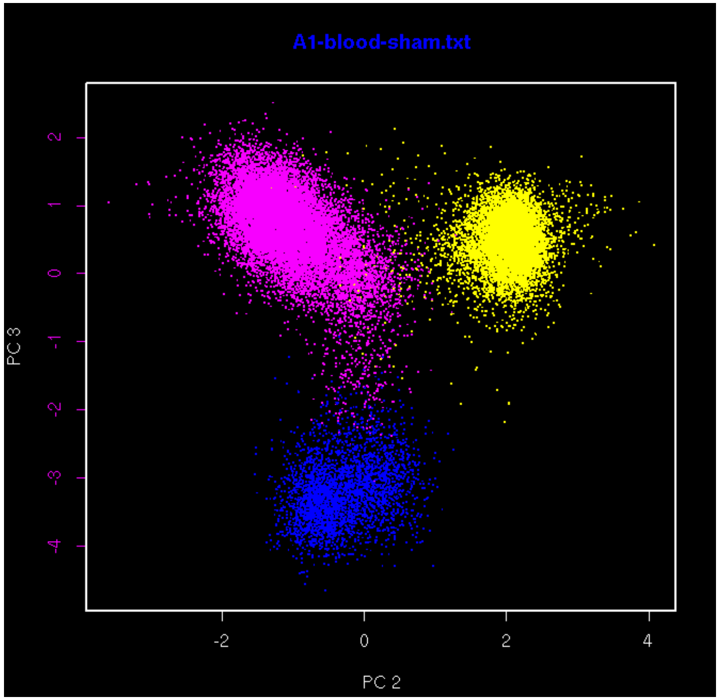
A. $2 B. $3 C. $4

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



Assignments

- Finish week4 module
- Next time: Markov and 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!