Mathematical Concepts (G6012)

Lecture 19

Thomas Nowotny
Chichester I, Room CI-105
Office hours: Tuesdays 15:00-16:45
T.Nowotny@sussex.ac.uk

Reminder: Probability space

(Ω, P) is a probability space, if the following conditions hold:

$$P : \mathcal{P}(\Omega) \to [0,1]$$

$$\omega \mapsto P(\omega)$$

$$P(\Omega) = 1$$

$$P(\emptyset) = 0$$

$P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$

(Additivity)

Intuition

- The probability of events is like a “volume” of the sets that describe them.
- For the discrete probability spaces we are working with this translates into “number of (equally likely) outcomes that constitute each event.

Reminder: Independence

Definition: Two events A and B are **independent** if

$$P(A \cap B) = P(A) \cdot P(B)$$

Reminder: Conditional Probability

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

$$\Leftrightarrow \; P(A \cap B) = P(A|B)\, P(B)$$

Reminder: Bayes rule

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

Proof:

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

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

In other words Bayes theorem = definition of conditional probability.

Random variable

A **random variable** is a function

$$X : \Omega \to \mathbb{R}$$
$$\omega \mapsto X(\omega)$$

Examples:

$$\Omega = \{(i,j) : i = 1, \ldots 6, j = 1 \ldots 6\}$$ (two dice)

$$X(i,j) = i + j \qquad Z(i,j) = i^2 + j^2$$

$$Y(i,j) = i \cdot j$$ … and so on

Events for random variables

We can now write, e.g. $P(X = k)$:

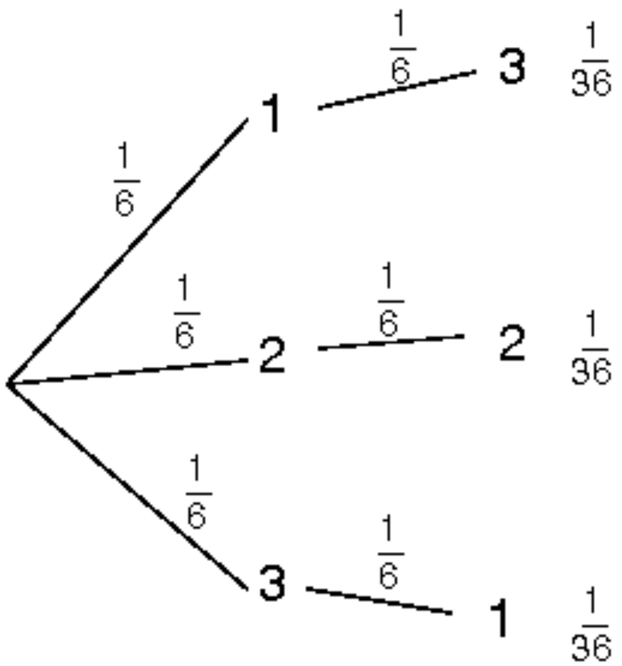
$$P(X = k) = P(\{\omega \in \Omega \ : \ X(\omega) = k\})$$

Example two dice, $X(i,j) = i + j$, i.e. X is
The sum of the two thrown dice.
What is the probability $P(X = 4)$?

Tree graph: **BB**

$$P(X = 4)$$

BB



So, $P(X = 4)$ is equal to

$$P(X = 4) = 3 \cdot \frac{1}{36} = \frac{1}{12}$$

Expectation value

- The **expectation value** is defined as

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega) P(\omega)$$

- This is equivalent to

$$\mathbb{E}X = \sum_{k \in \mathcal{X}} k P(X = k)$$

where $\mathcal{X}$ is the set of all possible values of X

Expectation value

- The expectation value is the average value we **expect** X to take, if we make many trials (this can be made more precise; it is called the law of large numbers, see below)

Example: Throwing a coin

$$\Omega = \{head, tail\} \qquad P(\{head\}) = P(\{tail\}) = \frac{1}{2}$$

$$X : \Omega \to \mathbb{R}$$

$$X(head) = 1 \quad X(tail) = 0$$

Example continued

$$\mathbb{E}X = X(head)P(\{head\}) + X(tail)P(\{tail\})$$

$$= 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}$$

Note: The expectation value does not need to be one of the possible values of X

Properties of the expectation value

- The expectation value is additive for any random variables:

$$\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$$

$$\mathbb{E}(k \cdot X) = k \cdot \mathbb{E}X$$

Proof: BB

- For the expectation value of independent random variables:

$$\mathbb{E}(X \cdot Y) = \mathbb{E}X \cdot \mathbb{E}Y$$

BB Proof

$$
\begin{aligned}
\mathbb{E}(X+Y) &= \sum_{\omega\in\Omega} \big(X(\omega)+Y(\omega)\big)P(\omega) \\
&= \sum_{\omega\in\Omega} \big(X(\omega)P(\omega)+Y(\omega)P(\omega)\big) \\
&= \sum_{\omega\in\Omega} X(\omega)P(\omega) + \sum_{\omega\in\Omega} Y(\omega)P(\omega) = \mathbb{E}X + \mathbb{E}Y
\end{aligned}
$$

Variance, Standard Deviation

- Variance is how much a RV **varies** around its expectation value

$$
\begin{aligned}
Var(X) &= \mathbb{E}((X - \mathbb{E}X)^2) \\
&= \mathbb{E}X^2 - (\mathbb{E}X)^2
\end{aligned}
$$

Proof: BB

- Standard deviation is the square root of variance

$$std(X) = \sigma_X = \sqrt{Var(X)}$$

BB Proof

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\big((x - \mathbb{E}X)^2\big) \\
&= \mathbb{E}\big(X^2 - 2X\mathbb{E}X + (\mathbb{E}x)^2\big) \\
&= \mathbb{E}X^2 - 2\mathbb{E}X \cdot \mathbb{E}X + (\mathbb{E}X)^2 \\
&= \mathbb{E}X^2 - (\mathbb{E}X)^2
\end{aligned}$$

Covariance

- So far we have characterized a single random variable
- Covariance characterizes the relationship between two random variables, how much they “co-vary”.

$$X : \Omega \to \mathbb{R} \qquad\qquad Y : \Omega \to \mathbb{R}$$

$$\begin{aligned} \mathrm{Cov}(X, Y) &= \mathbb{E}\big((X - \mathbb{E}X)(Y - \mathbb{E}Y)\big) \\ &= \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y \end{aligned}$$

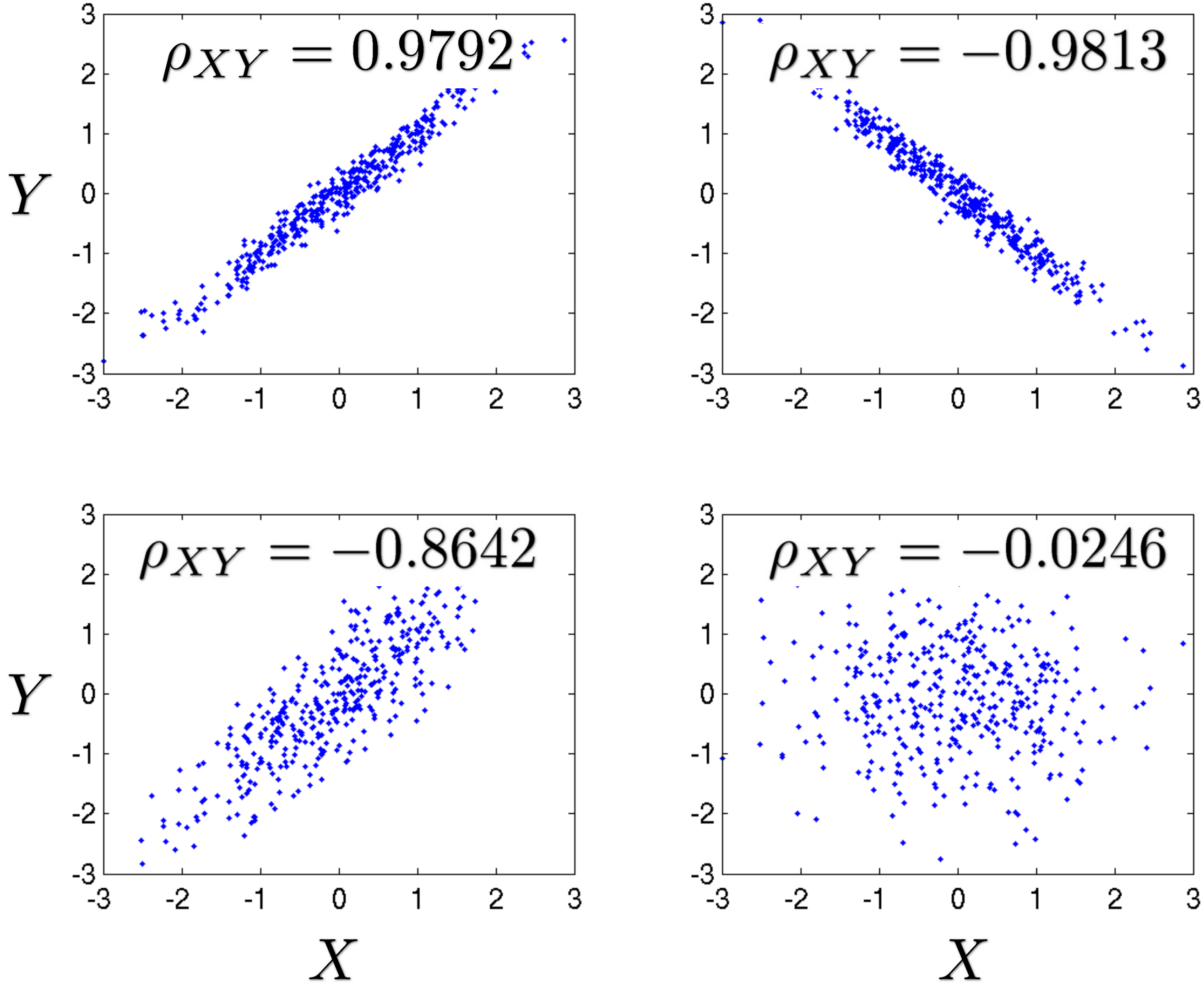
Correlation

- Correlation is a normalized form of covariance:

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \, \sigma_Y}$$

σ_X and σ_Y are the standard deviation of X and Y

Examples for correlation



Binomial distribution

- Is the probability distribution for the number of successes (1s) in a so-called Bernoulli process

- Bernoulli process:

$$\Omega = \{0,1\}^n = \big\{(\omega_1, \ldots, \omega_n) : \omega_i \in \{0,1\}\big\}$$

$$P(\{\omega_i = 1\}) = p \qquad\qquad P(\{\omega_i = 0\}) = 1 - p$$

And all such events are independent, such that

$$P(\{(\omega_1, \ldots, \omega_n)\}) = p^k (1-p)^{n-k}$$

k = number of 1s, n-k= number of 0s.

Binomial distribution

$$X : \Omega \to \mathbb{R}$$

$$X\big((\omega_1, \ldots, \omega_n)\big) = \sum_{i=1}^{n} \omega_i$$

X is the number of 1s (number of successes)

What is $P(\{X = k\})$?

$$P(\{(\omega_1, \ldots, \omega_n) \in \Omega : \sum_{i=1}^{n} \omega_i = k\})$$

In other words the probability of all elementary events with k 1s and n-k 0s.

Binomial distribution

$$= \sum_{\substack{(\omega_1, \ldots, \omega_n) \\ k\,1s,\; n-k\,0s}} P(\{(\omega_1, \ldots, \omega_n)\})$$

$$= \boxed{N(k, n-k)} \cdot (p)^k \, (1-p)^{n-k}$$

We only need to calculate the number of possible arrangements of k 1s and (n-k) 0s.

BB Number of combinations (1)

1	0	0	1	0	0	1	0

n possibilities to choose position of first “1”

Each, (n-1) possibilities to choose the next “1”: n(n-1)

…

Finally (n-k+1) possibilities for last “1”: n(n-1)…(n-k+1)

But we counted too many – we counted twice if a “1” is first placed onto one of the eventual “1 postions” or first onto another of the ”1-positions”: Need to correct by k(k-1)…*2*1

BB Number of combinations (2)

This leaves us with # configurations

$$\frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 2 \cdot 1}$$

We can replace the numerator with

$$n!/n-k)!$$

Which leads to # configurations

$$\frac{n!}{k!(n-k)!}$$ (so-called "binomial coefficient")

Binomial distribution

$$P(\{X = k\}) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

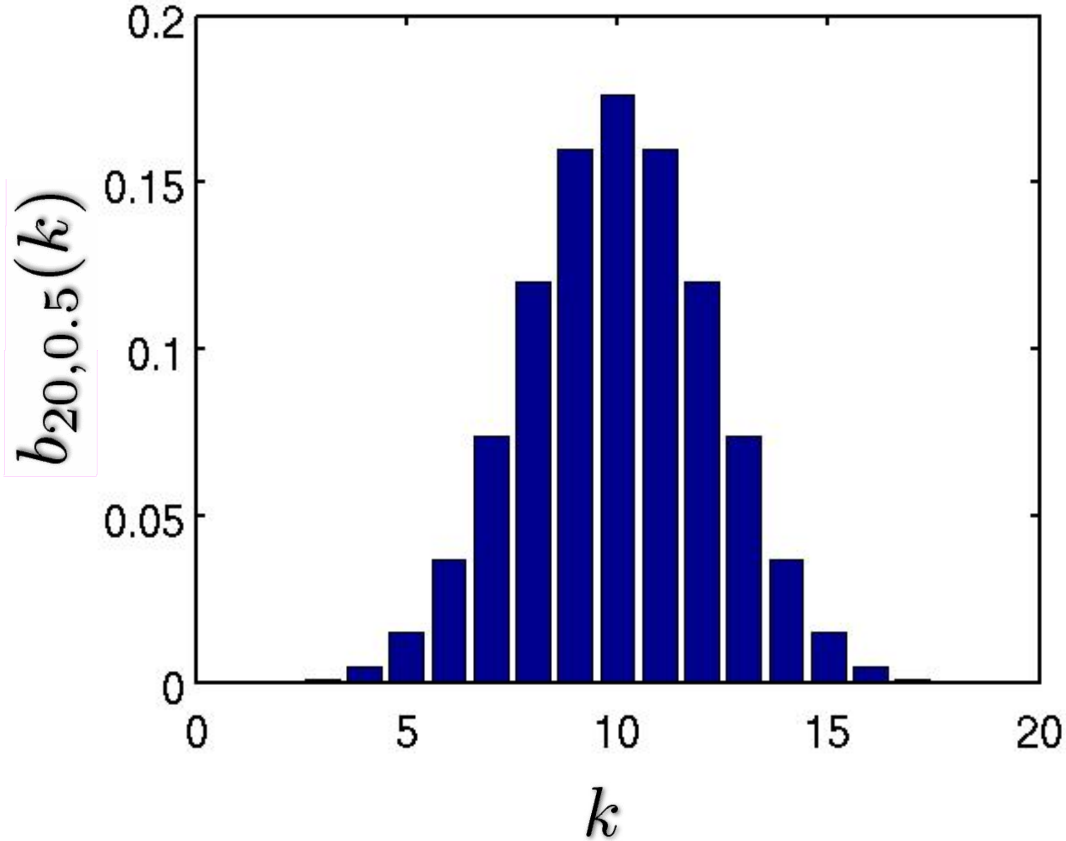
$$P(\{X = k\}) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}$$

Binomial coefficient

The binomial distribution is sometimes written as

$b_{n,p}(k)$

Binomial distribution properties



Expectation value

$$\mathbb{E}\sum_{i=1}^{n}\omega_i = n\cdot p$$

Standard deviation

$$\sigma\left(\sum_{i=1}^{n}\omega_i\right) = \sqrt{np(1-p)}$$

Note: Expectation value and maximum are *not* the same

Law of large numbers

- There are several different laws of large numbers.
- Here I would like to show you one example to give a feel for what these laws are about.

Law of large numbers for Bernoulli processes

$$\Omega_n = \{0,1\}^n$$

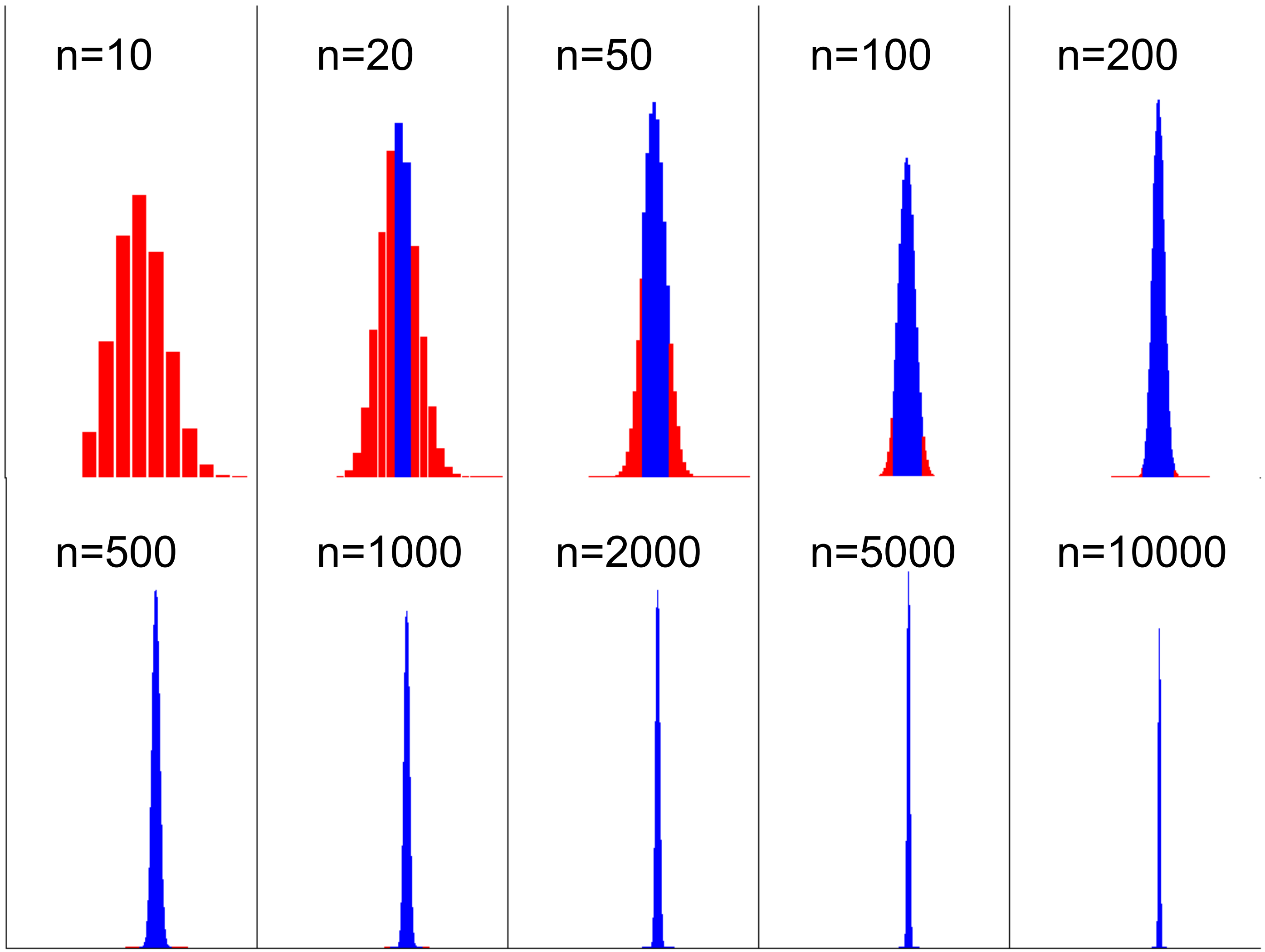
Random variable $X_n = \sum_{i=1}^{n} \omega_i$

has distribution $b_{n,p}(k)$

Consider $x_n = X_n / n$ and $\epsilon > 0$

$P(|x_n - p| \geq \epsilon)$ is the area under the tails:

$$P(|x_n - p| \geq \epsilon) \quad \text{for} \quad \epsilon = 0.1 \quad \text{and} \quad p = 0.4$$



Law of large numbers

For any $\epsilon > 0$

$$\lim_{n\to\infty} P(|x_n - \mathbb{E}x_n| \geq \epsilon) = 0$$

The probability for x_n to be more than ϵ away from its expectation value $\mathbb{E}x_n = p$ converges to 0 for $n \to \infty$.