INF5820/INF9820

LANGUAGE TECHNOLOGICAL APPLICATIONS

Jan Tore Lønning, 2016 fall

Probabilities 2

Today

- Some maths:
  - Notation for sums and products
  - Exponential function and logarithms
- More on probabilities
  - Conditional probabilities and Bayes' formula
  - Random variables
  - Probability distributions – binomial distribtuion

Notation

- $\sum_{i=1}^{n} a_i = a_1 + a_2 + a_3 + \dots + a_n$
- $\sum_{i=1}^{7} i = 1 + 2 + 3 + 4 + 5 + 6 + 7 = 28$
- $\sum_{i=2}^{5} i^2 = 4 + 9 + 16 + 25 = 54$
- $\prod_{i=1}^{n} a_i = a_1 * a_2 * a_3 * \dots * a_n$
- $\prod_{i=1}^{7} i = 1 * 2 * 3 * 4 * 5 * 6 * 7 = 7! = 5040$

Exponentiation

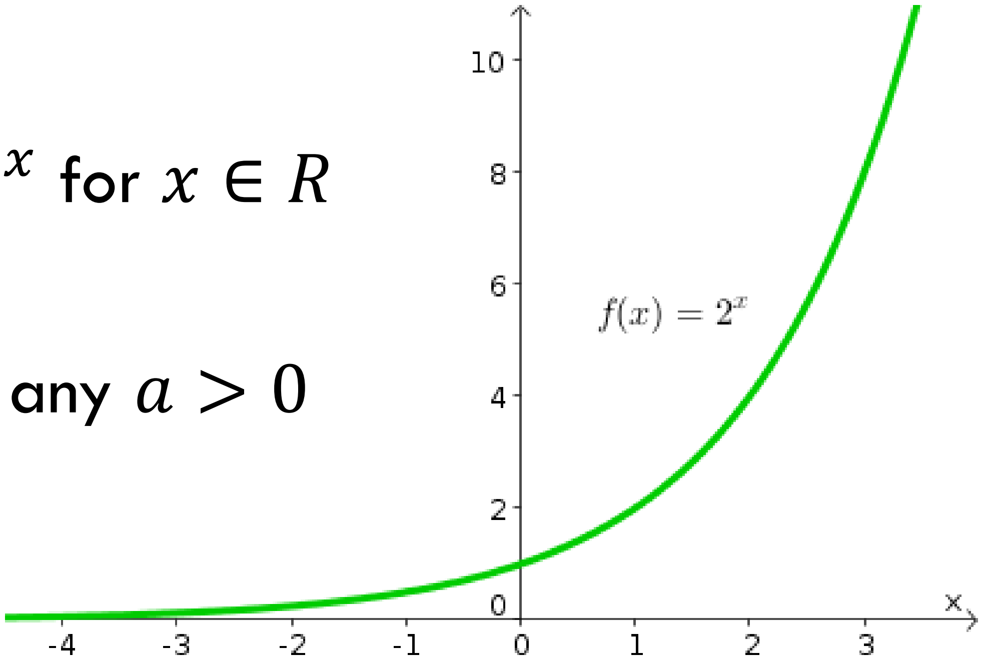
- Assume $a, b > 0$
- $a^m \cdot a^n = a^{(m+n)}$
- $a^{-m} = \frac{1}{a^m}$
- $(a^m)^n = a^{(m \cdot n)}$
- $a^{\left(\frac{1}{n}\right)} = \sqrt[n]{a}$ because
- $(a^{\left(\frac{1}{n}\right)})^n = a^{\left(\frac{1}{n} \cdot n\right)} = a$

$$\prod_{i=1}^{n} a^{b_i} = a^{\left(\sum_{i=1}^{n} b_i\right)}$$

Exponential function

0	1	2	3	4	5	6	7	8	9	10
1	2	4	8	16	32	64	128	256	512	1024

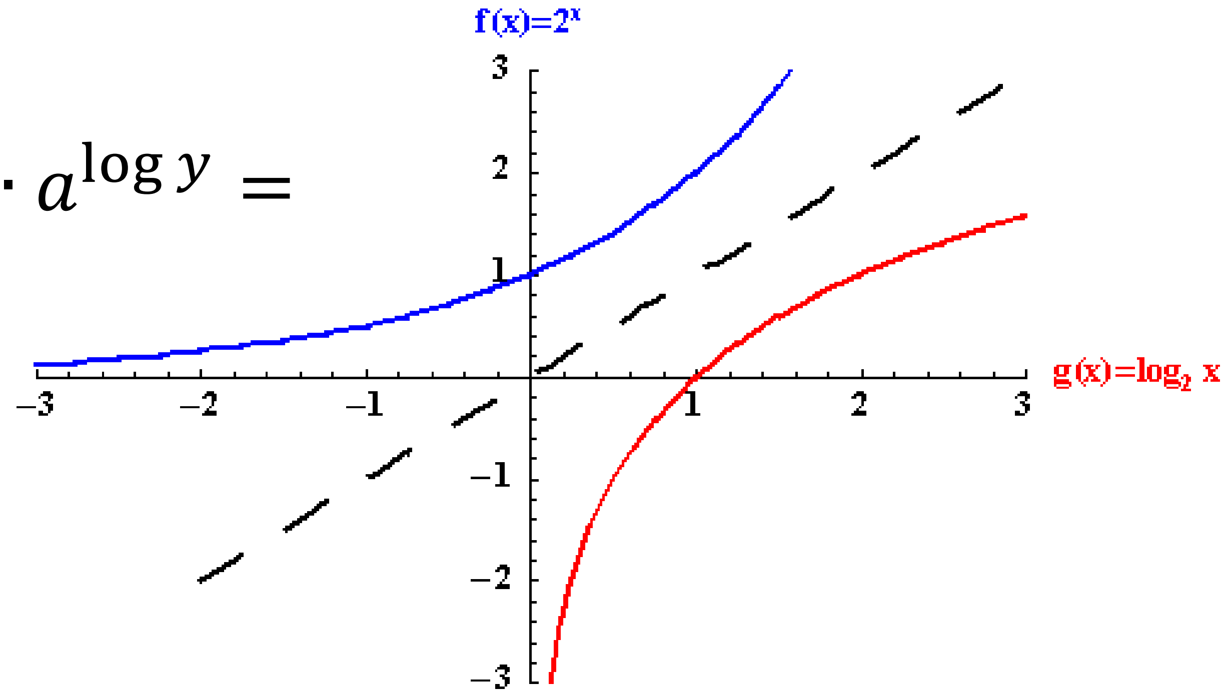
- 2^n for n = 0, 1, 2, …
- Extends to $f(x) = 2^x$ for $x \in R$
- Also $f_a(x) = a^x$ for any $a > 0$



Logarithms

- When $y = a^x$
- $x = \log_a y$
- Inverse function of exp
- $a^{\log_a x} = x$
- $x \cdot y = a^{\log x} \cdot a^{\log y} = a^{(\log x + \log y)}$

$$\prod_{i=1}^{n} x_i = \prod_{i=1}^{n} a^{\log x_i} = a^{\left(\sum_{i=1}^{n} \log x_i\right)}$$



Today

- Some maths:
  - Notation for sums and products
  - Exponential function and logarithms
- More on probabilities
  - Conditional probabilities and Bayes' formula
  - Random variables
  - Probability distributions – binomial distribtuion

Basic concepts

- Random experiment (or trial) (no: forsøk)
  - Observing a chance event
- Outcomes (utfallene)
  - The possible results of the experiment
- Sample space (utfallsrommet)
  - The set of all possible outcomes

Event

- An event (begivenhet) is a set of elementary outcomes

	Experiment	Event	Formally
2	Rolling a dice	Getting 5 or 6	{5,6}
3	Flipping a coin three times	Getting at least two heads	{HHH, HHT, HTH, THH}
5	The first word you hear tomorrow	You hear a noun	{u | u is a noun}
6	Throwing a dice until you get 6	An odd number of throws	{1,3,5, …}
7	The maximum temperature at Blindern	Between 20 and 22	$\{t \mid 20 \leq t \leq 22\}$

Probability measure, sannsynlighetsmål

- A probability measure P is a function from events to the interval [0,1] such that:

1. $P(\Omega) = 1$
2. $P(A) \geq 0$
3. If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$
   - And if A1, A2, A3, … are disjoint, then $P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$

Conditional probability

- $P(A \cap B)$
  - Both A and B happens
- Conditional probability (betinget sannsynlighet)

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

  - The probability of A happens if B happens
- Multiplication rule $P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$
- A and B are independent iff $P(A \cap B) = P(A)P(B)$

Bayes theorem

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

- Jargon:
  - P(A) – prior probability
  - P(A|B) – posterior probability
- Extended form

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

Example

- Disease: 1 out of 1000 are infected
- Test:
  - Detects 99% of the infected
  - 2% of the non-infected get a positive test
- Given a positive test: what is the chance you are ill?

What are probabilities?

- Example throwing a dice:

1. Classical view:
   - The six outcomes are equally likely
2. Frequenist:
   - If you throw the dice many, many, many times, the number of 6s approach 16.6666...%
3. Bayesian: subjective beliefs

Today

- Some maths:
  - Notation for sums and products
  - Exponential function and logarithms
- More on probabilities
  - Conditional probabilities and Bayes' formula
  - Random variables
  - Probability distributions – binomial distribtuion

Random variable

- A variable X in statistics is a property (feature) of an outcome of an experiment.
  - Formally it is a function from a sample space (utfallsrom) Ω to a value space Ω_X.
- When the value space Ω_X is numeric (roughly a subset of R^n), it is called a random variable
- There are two kinds:
  - Discrete random variables
  - Continuous random variables
- A third type of variable: categorical variable, when Ω_X is nonnumeric

- Motivation for talking about random variables:
  - It is convenient to consider mathematical concepts, e.g. mean and variance
  - We can study their distribution abstracting away from what the actual outcomes

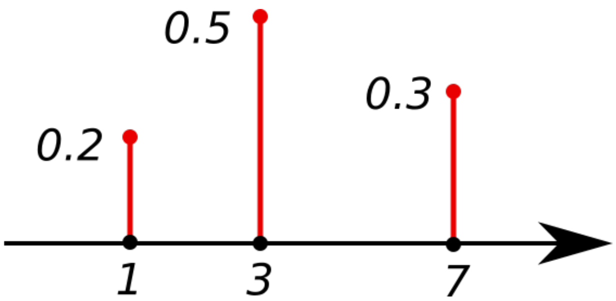
Examples

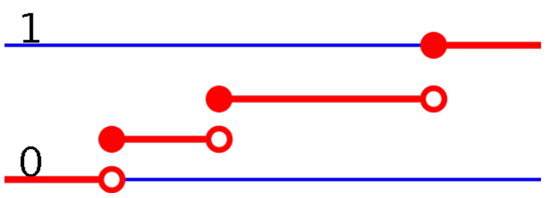
1. Throwing two dice,
   - Ω={(1,1), (1,2),…(1,6),(2,1),…(6,6)}
   1. The number of 6s is a random variable X, Ω_X={0, 1, 2}
   2. The number of 5 or 6s is a random variable Y, $\Omega_Y = \Omega_X$
   3. The sum of the two dice, Z, Ω_Z={2, 3, …, 12}
2. A random person:
   1. X, the height of the person Ω_X=[0, 3] (meters)
   2. Y, the gender Ω_Y={0, 1} (1 for female)

- Ex 2.1 is continuous, the other are discrete

Discrete random variable

- The value space is a finite or a countable infinite set of numbers {x1, x2,…, xn, …}
- The **probability mass function**, pmf, p, which to each value yields
  - $p(xi) = p(X=xi) = P(\{\omega \in \Omega \mid X(\omega)=x\})$
- The **cumulative distribution function**, cdf,
  - $F(x_i) = p(X \leq x_i) = P(\{\omega \in \Omega \mid X(\omega) \leq x_i\})$

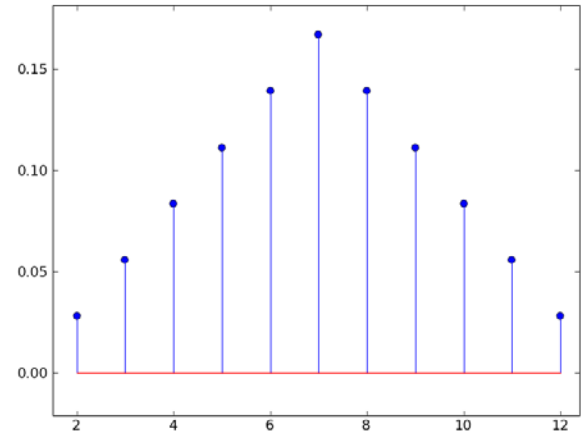


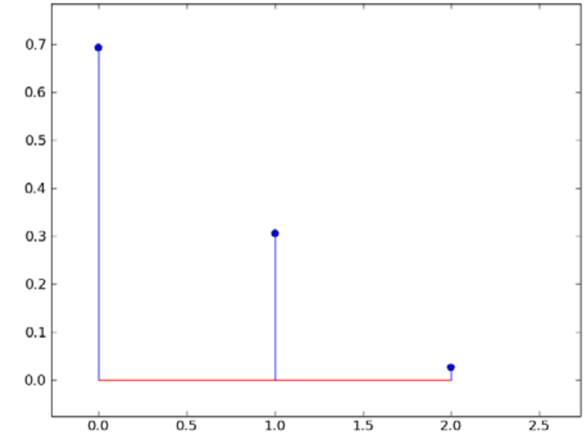


Diagrams: Wikipedia

Examples

- Throwing two dice,
  - $\Omega=\{(1,1), (1,2),\ldots(1,6),(2,1),\ldots(6,6)\}$
  - (1.3) The sum of the two dice, Z, $\Omega_Z=\{2, 3, \ldots, 12\}$
    - $p_Z(2) = P(\{(1,1)\} = 1/36$
    - $p_Z(7) = 6/36$
    - $F_Z(7) = 1+2+\ldots+6=21/36$
  - (1.1) The number of 6s X, $\Omega_X=\{0, 1, 2\}$
    - $p_X(2) = P(\{(6,6)\} = 1/36$
    - $p_X(1) = P(\{(6,x)|x\neq 6\}+ P(\{(x,6)|x\neq 6\}=10/36$
    - $px(0) = 25/36$





Mean – example

- Throwing two dice, what is the mean value of their sum?
- (2+3+…+7+
3+4+…+8+
4+…   +9+
5+…   +10+
6+…   +11+
7+…   +12)/36=
- (2+2*3+3*4+4*5+5*6+6*7+5*8+…2*11+12)/36=
- (1/36)2+(2/36)*3+(3/36)*4+… +(1/36)*12 =
- p(2)*2 + p(3)*3 + p(4)*4 + … p(12)*12 =
- Σp(x)*x

Mean of a discrete random variable

- The mean (or expectation) (forventningsverdi) of a discrete random variable X:

$$\mu_X = E(X) = \sum_x p(x)x$$

- Useful to remember

$$\mu_{(X+Y)} = \mu_X + \mu_Y$$

$$\mu_{(a+bX)} = a + b\mu_x$$

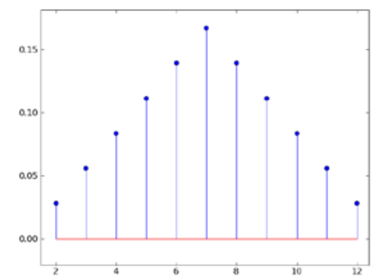
Examples:
One dice: 3.5
Two dices: 7
Ten dices: 35

Today

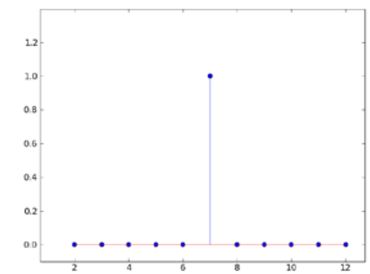
- Some maths:
  - Notation for sums and products
  - Exponential function and logarithms
- More on probabilities
  - Conditional probabilities and Bayes' formula
  - Random variables
  - Probability distributions – binomial distribution

Examples of distributions

- (1.3) The sum of the two dice, Z, i.e.
  - $p_Z(2) = 1/36, \ldots, p_Z(7) = 6/36$ etc

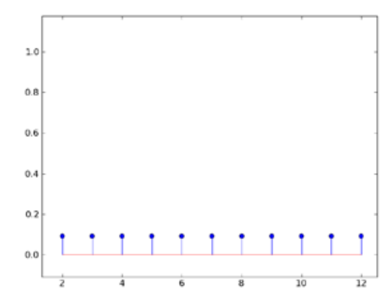


- (3.2) p_2 given by:
  - $p_2(7)=1$
  - $p_2(x)= 0$ for $x \neq 7$



- (3.3) p_3 given by:
  - $p_3(x)= 1/11$ for $x = 2,3,\ldots,12$
-

  The three have the same mean but are very different



Think about

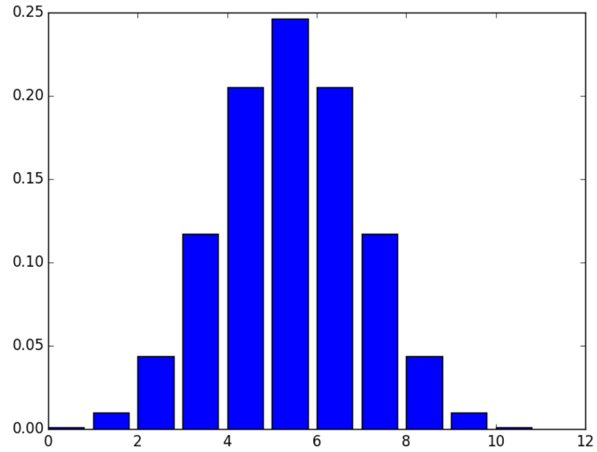
- Flip a coin 10 times, count the number of heads
- You expect 5 heads, but not exactly 5
  - 6 is OK
- When do you start to worry whether the coin is unfair?
  - 8 heads?
  - 9 heads?

- This is the task for inferential statistics

Tossing a fair(?) coin

- The cumulative distribution function: ``How likely is it to get N or fewer tails?´´

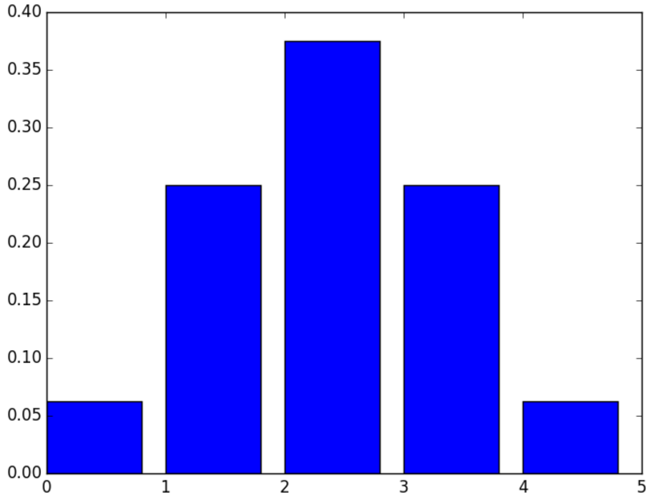
10:



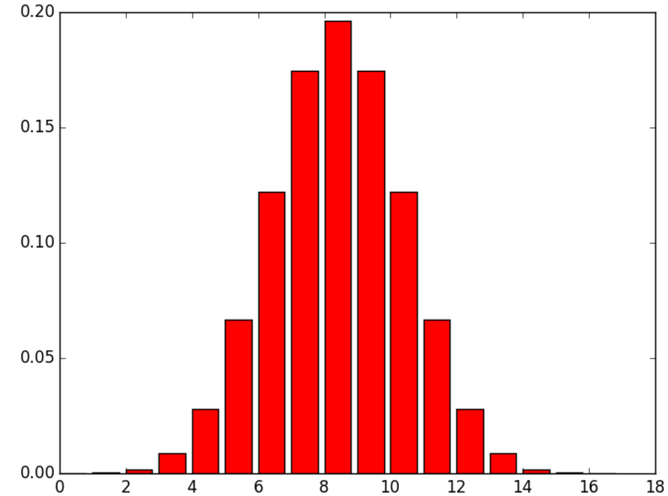
N	pmf(N)	cdf(N)
0	0.001	0.001
1	0.010	0.011
2	0.044	0.055
3	0.117	0.172
4	0.205	0.377
5	0.246	0.623
6	0.205	0.828
7	0.117	0.945
8	0.044	0.989
9	0.010	0.999
10	0.001	1.000

p=0.5

N=4:

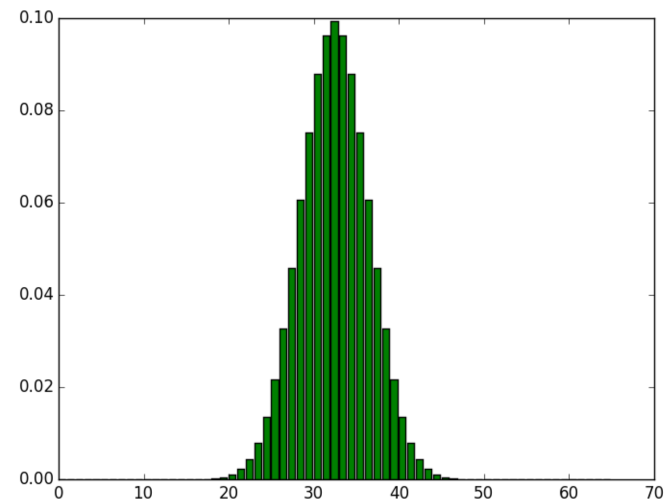


N=16:



N	1	4	16	64	256
σ^2	0.25	1	4	16	64
σ	0.5	1	2	4	8

N=64:



- The relative variation gets smaller with growing N
- The pmf graph approaches a bell shape

Bernoulli trial

- One experiment, two outcomes
- $\Omega_X=\{0, 1\}$
- Write p for p(1)
- Then p(0) = 1-p

Examples:

- Flipping a fair coin, p=1/2
- Rolling a dice, getting a 6, p=1/6

- The mean/expectation: 0*p(0)+1*p(1)=0+p=p

Sampling

Ordered sequences:

- Choose *k* items from a population of *n* items with replacement: n^k
- Without replacement (permutation):
  - : n(n-1)(n-2)...(n-k+1)=$\frac{n!}{(n-k)!}$

Unordered sequences

- Without replac.: $\frac{1}{k!}\left(\frac{n!}{(n-k)!}\right) = \left(\frac{n!}{k!(n-k)!}\right) = \binom{n}{k}$
  - = the number of ordered sequences/
    The number of ordered sequences containing the same *k* elements

Binomial distribution

- Binomial distribution (binomisk fordeling)
- Conducting *n* Bernoulli trials with the same probability and counting the number of successes

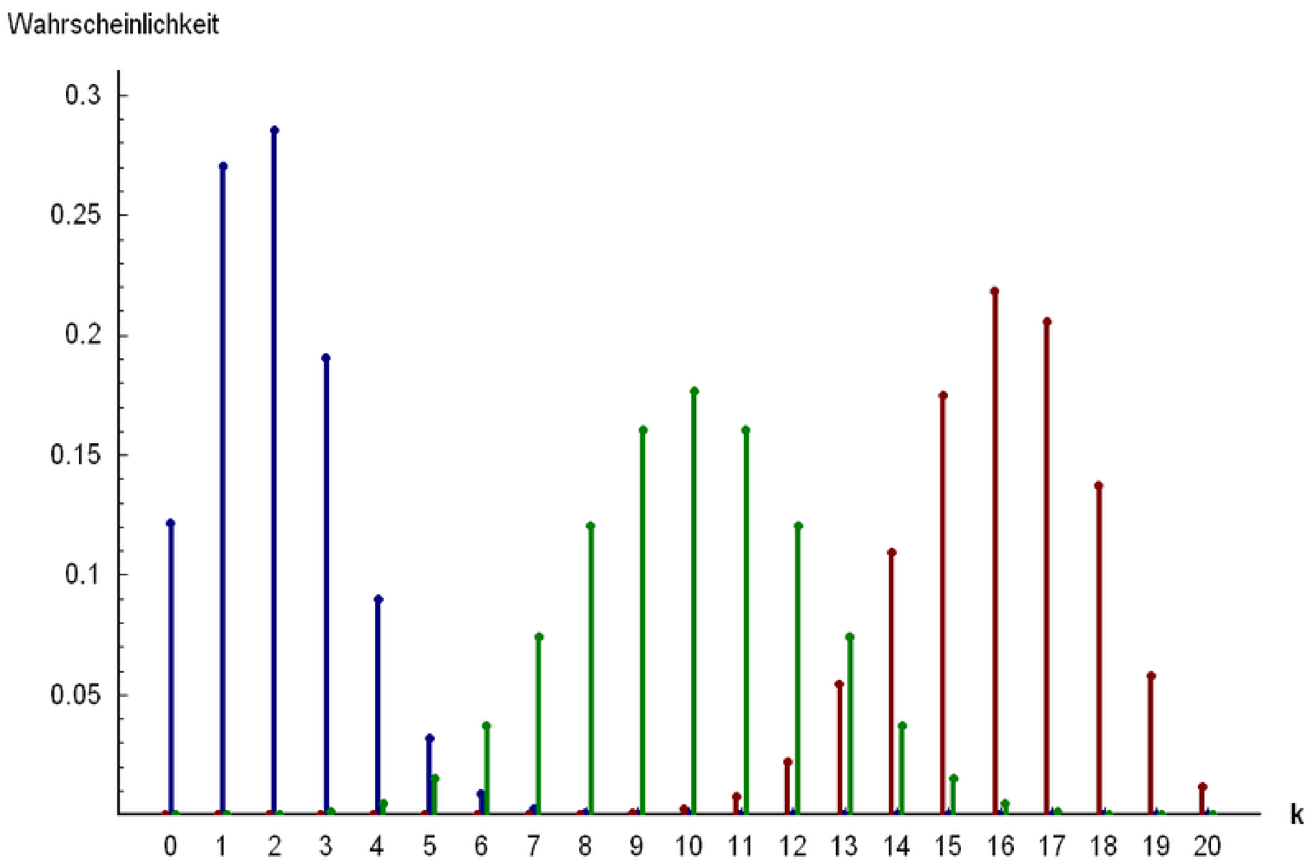
- Example flipping a fair coin *n* times, p(k):
  - n=2: p(0)=1/4, p(1)=1/2, p(2) =1/4
  - n=3: p(0)=1/8, p(1)=3/8, p(2)=3/8, p(3)=1/8
  - n=4: (1,4,6,4,1)/16
  - n=5: (1,5,10,5,1)/32
- n: $p(k) = \binom{n}{k}\left(\frac{1}{2}\right)^n$ where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
-

Binomial distribution

- Binomial distribution (binomisk fordeling)
- General form:
  - $0<p<1$
  - *n* a natural number
- B(n,p) is given by $$b(k;n,p)=\binom{n}{k}p^k(1-p)^{(n-k)}$$

for k = 0, 1, …n, where $$\binom{n}{k}=\frac{n!}{k!(n-k)!}$$

Binomial distribution



- n = 20
- p = 0.1 (blue), p = 0.5 (green) and p = 0.8 (red)