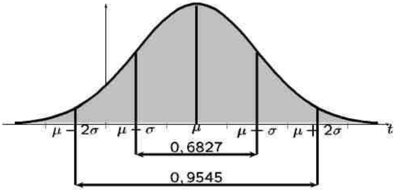
Statistics

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$



Bachelor Studiengang Informatik

Prof. Dr. Egbert Falkenberg

Fachbereich Informatik & Ingenieurwissenschaften

Wintersemester 21/22

Section 1

Calculus of Probability

Introduction I

compare: Pitman, Probability, chapter 1.2 Introduction

Probability theory: A branch of mathematics concerned with the analysis of random phenomena. The outcome of a random event cannot be determined before it occurs, but it may be any one of several possible outcomes. The actual outcome is considered to be determined by chance.
(source: https://www.britannica.com/topic/probability-theory)

Introduction II

- Probability is the measure of the likelihood that an event will occur. Probability is quantified as a number between 0 and 1, where, loosely speaking, 0 indicates impossibility and 1 indicates certainty. (source: https://en.wikipedia.org/wiki/Probability)
- Definition of probability?
- Problem: a mathematical theory precise enough for use in mathematics and comprehensive enough to be applicable to a wide range of phenomena
- Different approaches and discussion over centuries, but all were never fully accepted
- Kolmogoroff 1933: axiomatic definition of probability avoids former problems with inconsistent definitions and allows building up probability theory

Introduction IV

Question: Interpretation of probabilities in applications

Probabilities as approximations to long-run frequencies

- Probability of an event = expected or estimated relative frequency of A in a large number of trials. Theoretical probability P(A) = limit of relative frequencies $P_n(A)$ as $n \to \infty$ ("Law of Large Numbers").
- This justifies: theoretical probability = a useful approximation to relative frequency $P_n(A)$ for large values of n.
- Interpretation sensefull in a context of repeated trials

Introduction IV

Context of repeated trials not always appropriate

Example: Probability of a patient surviving an operation.

- Statement of a doctor: you survive the operation with a probability of 0.95.
- What does this means?
  - A survival in round about 95 of 100 operations???
  - Fatality of 5% in the past in similar operations: doctor has additional informations
    - your state of health is much better, or
    - you are much younger or
    - you are some different from the population with a fatality of 5%.

  But the doctor do not knows survival percentages for patients just like you.
- Thus: 95% chance of surviving = matter of opinion.

Introduction V

Another Interpretion of probability: probabilities as degrees of believe.

Bayesian view of probability: instead of frequency of some phenomenon, probability is interpreted as reasonable expectation representing a state of knowledge or as quantification of a personal belief.

- Which interpration is suitable depends on the context of the application.
- Here: frequentistic approach
- Inferential statistics from the viewpoint of a frequence interpretation of probabilities (**frequentistic inference**).
- Methods of **Bayesian Inference** not discussed, not possible in an introductory course.

Probability Spaces I

Compare Virtual Lab of Probability and Statistics $\rightarrow$ 1. Probability Spaces: 1, 2, 3, 5, 6 and Online Statistics: V: 2, 3, 4

Probability theory is based on the paradigm of a **random experiment.**

Definition: Random Experiment

- Outcome cannot be predicted with certainty, before the experiment is run
- Infinite number of repetitions under essentially the same conditions possible

Probability Spaces II

Definition: Sample space Ω of a random experiment is the set that includes all possible outcomes of the experiment

Example:

- Tossing one coin (1 for head or 0 for tail): $\Omega = \{0, 1\}$
- Tossing two distinct coins: $\Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$
- Capture a fly and measure its body weight (in milligrams): usually $\Omega = [0, \infty)$, even though most elements of this set are practically impossible

Probability Spaces III

Example: Sampling Experiments

- Population of objects: people or memory chips, for example
- One or more numerical measurements of interest: the height and weight of a person or the lifetime of a memory chip, for example
- Population of objects usually too large $\rightarrow$ random sample
- Basic types of sampling:
  - Sample with replacement
  - Sample without replacement

Questions

Which of the following statements are true or false?

t	f	
☐	☐	In the frequentistic approach probabilities are seen as approximations to long-run frequencies.
☐	☐	Any experiment can be interpreted as a random experiment.
☐	☐	The sample space of rolling a die is the interval [1,6].
☐	☐	If 10 balls are drawn from an urn with 100 from 1 to 100 by numbered balls, then in any case the numbers of the drawn balls are all different.

Probability Spaces IV

Definition: Events are subsets of the sample space of an experiment.

- Event = set of outcomes of the experiment
- Elementary event = event which consists of a single outcome in the sample space
- Event A occurs: outcome of the experiment is an element of A
- Event A does not occur: outcome of the experiment is not an element of A.

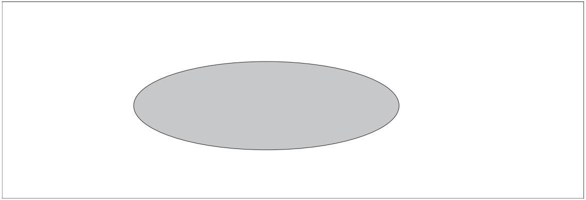
Probability Spaces V

Example:

- Tossing one coin (1 for head or 0 for tail):
  - $A = \{1\}$ occurs, if a head occurs after tossing the coin
  - Set of all possible events: $\Omega_1 = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$
- Tossing two distinct coins:
  - Set of all possible events: $\Omega_1 \times \Omega_1$
  - $A = \{(0,0)\}$ occurs, if two tails occure after tossing the two coins; A is an elementary event.
  - $B = \{(0,1),(1,1)\} =$ the second coin shows head; B is not an elementary event.

Probability Spaces VI

Visualizing Events as Subsets of the Sample Space:

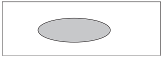
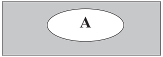
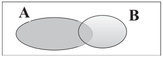
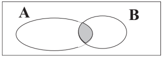
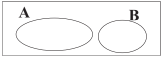
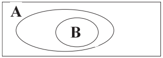


Interpretation:

- A point is picked at random from the square.
- Point in the square = outcome
- Region = event that the point is picked from that region.

Probability Spaces VII

Translations between events and sets:

Event language	Set language	Set notation	Venn diagram
outcome space	universal set	Ω	
event	subset of Ω	$A, B, C, \ldots$	
impossible event	empty set	$\emptyset$	
not A	complement of A	A^c	
either A or B or both	union of A and B	$A \cup B$	
both A and B	intersection of A and B	$A \cap B$	
A and B are mutually exclusive	A and B are disjoint	$A \cap B = \emptyset$	
if B then A	B is a subset of A	$B \subseteq A$	

Probability Spaces VIII

- **Problem:** Sometimes impossible to include all subsets of the sample space Ω as events.
- **Objektive:** assign probabilities to events in a random experiment
- The more events we include in the mathematical model of our random experiment, the harder it is to assign probabilities in a consistent way.
- Collection of events should be closed under certain set operations. $\rightarrow$ Collection of events $\mathcal{A}$ is required to be a σ-algebra.

Definition: A set $\mathcal{A}$ of subsets of $\Omega \neq \emptyset$ with:

- $\Omega \in \mathcal{A}$
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- $B_i \in \mathcal{A}, i \in \mathbb{N} \Rightarrow \bigcup_{i=1}^{\infty} B_i \in \mathcal{A}$

is called a $\sigma-$Algebra over Ω.

Probability Spaces IX

Definition of probability based on the rules of relative frequencies:

$P_n(A)$ = relative frequency of event A in n independent repetitions of a random experiment, where A_1 occurs n_1 times, A_2 occurs n_2 times, ...

$$\Rightarrow \quad 0 \leq P_n(A) \leq 1$$

$$P_n(\Omega) = 1$$

$$P_n(A_1 \cup A_2 \cup \ldots \cup A_k) = \frac{n_1+n_2+\ldots n_k}{n}$$

$$= \frac{n_1}{n} + \frac{n_2}{n} + \ldots \frac{n_k}{n} =$$

$$= P_n(A_1) + P_n(A_2) + \ldots + P_n(A_k)$$

if A_i pairwise disjoint

Probability Spaces X

- Random experiment with sample space Ω
- Intuitively, probability of an event = a measure of how likely the event is to occur when we run the experiment.

Definition: A probability measure P for a random experiment is a real-valued function defined on the collection $\mathcal{A}$ of events that satisfies the following axioms:

1. $P(A) \geq 0$ for every event $A \in \mathcal{A}$
2. $P(\Omega) = 1$
3. If $\{A_i \in \mathcal{A} \mid i \in I\}$ is a countable, pairwise disjoint collection of events then

$$P(\bigcup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$$

$(\Omega, \mathcal{A}, P)$ is called a probability space.

The Law of Large Numbers I

- Indefinitely repetitions of an experiment
- A event in the experiment
- $N_n(A)$ = number of times A occurred in the first n runs, i.e.

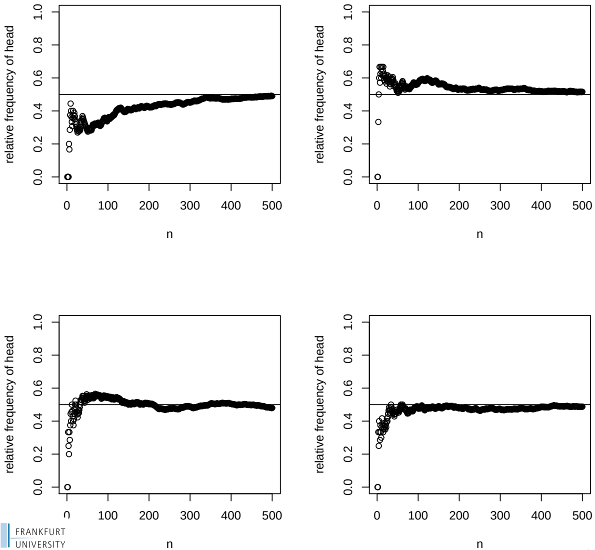
$$P_n(A) = \frac{N_n(A)}{n}$$

  relative frequency of A in the first n runs
- In case of a correct chosen probability measure we expect that in ome sense the relative frequency of each event should converge to the probability of the event:

$$P_n(A) \to P(A) \quad \text{as} \quad n \to \infty$$

The Law of Large Numbers II

Example: Tossing a coin repeatedly



The Law of Large Numbers III

Illustrating the Law of Large Numbers
Source: Wolfram Demonstrations Project

http://demonstrations.wolfram.com/IllustratingTheLawOfLargeNumbers/

Law of large numbers comparing relative versus absolute frequency
Source: Wolfram Demonstrations Project

http://demonstrations.wolfram.com/LawOfLargeNumbersComparingRelativeVersusAbsoluteFrequencyOfC/

- Precise statement: **law of large numbers**, one of the fundamental theorems in probability
- $\Rightarrow$ In the data from n runs of the experiment, the observed relative frequency $P_n(A)$ can be used as an approximation for $P(A)$.
- Approximation is called the empirical probability of A.

Questions

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Which of the following statements are true or false?

t	f	
☐	☐	Every subset of Ω is an event.
☐	☐	If $A \in \mathcal{A}$ and $\|A\| = 1$ then A is an elementary event.
☐	☐	If A is an event we know that $0 \leq P(A) \leq 1$.
☐	☐	For all $A, B \in \mathcal{A}$ we have $P(A \cup B) \leq P(A) + P(B)$.
☐	☐	For $A, B \in \mathcal{A}$ the event $A \cup B$ can be interpreted that either A or B has happened.
☐	☐	For $A \in \mathcal{A}$ the event A^c means that A has not happened.

Basic Rules I

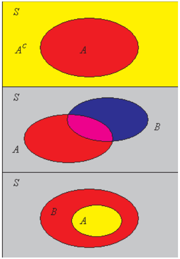
- Probability is defined as a function of events.
- The events are represented as sets.
- The probabbility function satisfies the basic rules of proportion. These are the rules for fractions or percentages in a population, and for relative areas of regions in a plane.

Random experiment with sample space S, probability measure P and $A, B \in \mathcal{A}$:

$$P(A^c) = 1 - P(A),\ P(\emptyset) = 0$$

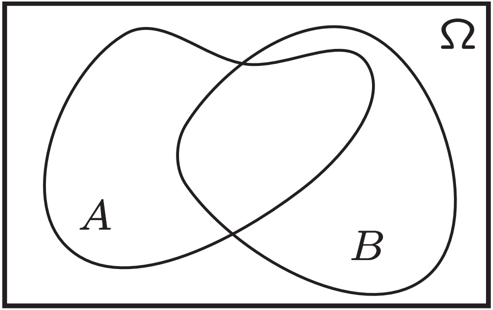
$$P(A \setminus B) = P(A) - P(A \cap B)$$

$$A \subset B \Rightarrow P(A) \leq P(B)$$



Basic Rules II

Example:

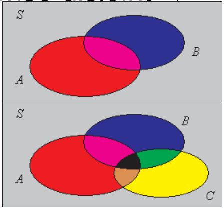


$P(A) = 0.6$, $P(B) = 0.55$, $P(A \cap B) = 0.3 \Rightarrow$
$P(A^c) = 0.4$, $P(B^c) = 0.45$, $P(A \cup B) = 0.85$,
$P(A^c \cap B) = P(B \setminus (A \cap B)) = 0.25$

Basic Rules III

The Inclusion-Exclusion Formula:

$A_1, A_2, \dots, A_k \in \mathcal{A}$ not necessarily pairwise disjoint $\Rightarrow$



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

$$P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i) - \sum_{1 \le i_1 < i_2 \le k} P(A_{i_1} \cap A_{i_2}) + \dots$$

$$+(-1)^{l+1} \sum_{1 \le i_1 < i_2 < \dots < i_l \le k} P(A_{i_1} \cap \dots \cap A_{i_l}) + \dots$$

$$+(-1)^{k+1} P(A_1 \cap \dots \cap A_k)$$

Basic Rules IV

Example Rencontre-Problem: 4 professors leave after dinner a restaurant. What is the probability that at least one professor gets his own coat?

- Let the professors and the coats are numbered from 1 to 4.
- The outcome of the experiment can be described by a bijective function from the set of professors to the set of coats.
- For example $\binom{1234}{3412}$:
  professor 1 gets the coat of professor 3, ... professor 4 gets the coat of professor 2.

Basic Rules V

Ω		set of permutations over 1,2,3,4
A_i		professor i gets his own coat
$P(A_i)$	=	$\frac{3!}{4!} = \frac{1}{4}$ for all i
$P(A_i \cap A_j)$	=	$\frac{1}{3 \cdot 4}$ for all $i \neq j$
$P(A_i \cap A_j \cap A_k)$	=	$\frac{1}{4!} = \frac{1}{2 \cdot 3 \cdot 4}$ for all $i \neq j \neq k$
$P(A_1 \cap A_2 \cap A_3 \cap A_4)$	=	$\frac{1}{4!} = \frac{1}{1 \cdot 2 \cdot 3 \cdot 4}$
A		at least one professor gets his own coat

Basic Rules VI

The rule of inclusion-exclusion leads to:

$$
\begin{aligned}
P(A) &= P(A_1 \cup A_2 \cup A_3 \cup A_4) \\
&= P(A_1) + \dots + P(A_4) - P(A_1 \cap A_2) - \dots - P(A_3 \cap A_4) \\
&\quad + P(A_1 \cap A_2 \cap A_3) + \dots + P(A_2 \cap A_3 \cap A_4) \\
&\quad - P(A_1 \cap \dots \cap A_4) \\
&= 4 \cdot \frac{1}{4} - \binom{4}{2} \frac{1}{3 \cdot 4} + \binom{4}{3} \frac{1}{2 \cdot 3 \cdot 4} - \binom{4}{4} \frac{1}{1 \cdot 2 \cdot 3 \cdot 4} \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} = \frac{5}{8}
\end{aligned}
$$

Generally if we have n professors and n coats we get

$$
P(A) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots + (-1)^{n+1} \frac{1}{n!} = 1 - \sum_{i=0}^{n} \frac{(-1)^i}{i!}
$$

Questions

Let $(\Omega, \mathcal{A}, P)$ be a probability space. $A, B, C \in \mathcal{A}$ with $P(A) = 0.5, P(B) = 0.3, P(C) = 0.2, P(A \cap B) = 0.2$.

Which of the following statements are true or false?

t	f	
☐	☐	The probability that B does not happens is 0.7.
☐	☐	$P(A \cup B) = 0.6$
☐	☐	$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C)$
☐	☐	Probability that B happens but not A is 0.1
☐	☐	The probability that neither A nor B will occur is 0.4.
☐	☐	$P(B \cap C) \leq 0.2$

A Brief Review: Probability Space I

- Random experiment
  - all possible distinct outcomes are known a priori
  - the outcome is not known a priori
  - it can be repeated under identical conditions
- Sample Space Ω: Set which includes all possible distinct outcomes of a random experiment
  Ω can be finite (example: tossing a coin), countable infinite (example: waiting for the first 6 in throwing a dice) or uncountable (example: measuring the weight of a captured fly).
- Objective: Assign probabilities to events of interest
- Problem: In case of uncountable sets Ω it is impossible to assign probabilities to all subsets $A \subset \Omega$
- Solution: Probability measure is defined on a collection of $\mathcal{A}$ - called $\sigma-$Algebra - of subsets of Ω.

A Brief Review: Probability Space II

- A system of subsets in Ω $\mathcal{A}$ is called a $\sigma-$algebra if
  - $\Omega \in \mathcal{A}$
  - $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
  - $A_i \in \mathcal{A}, i = 1, 2, 3, ... \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
- In case of finite or countable finite Ω we can use the power set of $\mathcal{P}(\Omega)$ as $\sigma-$algebra $\mathcal{A}$ to define a probability measure P for all elements of Ω
- A probability measure P is a function $P : \mathcal{A} \rightarrow [0, 1]$ with
  1. $P(A) \geq 0$ for all $A \in \mathcal{A}$
  2. $P(\Omega) = 1$
  3. If $\{A_i \in \mathcal{A} \,|\, i \in I\}$ is a countable, pairwise disjoint collection of events then $P(\bigcup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$
- With these definition a random experiment can be described by the triple $(\Omega, \mathcal{A}, P)$ which is called a probability space.

Discrete Probability Spaces I

- **Laplace Model:**
  Suppose a random experiment with a finite sample space Ω and each of the possible outcomes of the random experiment are equally like:

$$P(A) = \frac{|A|}{|\Omega|}, \quad A \subseteq \Omega$$

- **Generally:**
  Suppose that Ω is nonempty and countable, and that g is a nonnegative real-valued function defined on Ω with $\sum_{x\in\Omega} g(x) = 1$ then

$$P(A) = \sum_{x\in A} g(x), \quad A \subseteq \Omega$$

defines a probability measure.

Discrete Probability Spaces II

Examples - Laplace Model:

$$\text{probability} = \frac{\text{number of favourable outcomes}}{\text{number of possible equally-likely outcomes}}$$

- What is the probability p of getting either a one or a six if you roll the dice?
- What is the probability p that a card drawn at random from deck of playing cards will be an ace?
- Bag with 20 cherries, 14 sweet and 6 sour: If you pick a cherry at random, what is the probability p that it will be sweet?

Discrete Probability Spaces III

Example: Parallel computer with n=5 distinguishable processors P_1, P_2, P_3, P_4, P_5 and k non distinguishable jobs

1. k=3: If every processor can get at most one job what is the probability that P_1, P_2, P_3 get 2 jobs and P_4, P_5 get the remaining job?
2. k=10: If every processor can get more than one job and 10 jobs should be partitioned to the processors what is the probability that
   2.1 P_1 gets no job
   2.2 every processor get at least one job
   2.3 exactly one processor gets no job
   2.4 P_1 get 3 jobs, P_2 get 2 jobs, P_3 get 2 jobs, P_4 gets 1 job and P_5 get 2 jobs

Discrete Probability Spaces IV

Question 1: 5 distinguishable processors, 3 non distinguishable jobs, every processor at most one job

- Since the jobs are non distinguishable 3 out of 5 processors get one job: $\binom{5}{3} = 10$ possible partitionings of the jobs
- $\binom{3}{2} \cdot \binom{2}{1} = 6$ favourable partitionings of the jobs
- $\Rightarrow p = \frac{6}{10}$

Discrete Probability Spaces V

Question 2: 5 distinguishable processors, 10 non distinguishable jobs, more than on job for every processor possible

- Every partitioning can be described as a word of length 14 with the 2 different characters (P, J where J occure 10 times and P occure 4 times.
- *JJPPJJJPJPJJJJ* is equivalent to the partitioning: P_1 2 jobs, P_2 no jobs, P_3 3 jobs, P_4 1 job, P_5 4 jobs $\Rightarrow \frac{14!}{10! \cdot 4!} = \binom{14}{10}$ possible partitionings
- Thus a partitioning could be interpreted as a multiset with k=10 elements out of the $\{P_1, P_2, P_3, P_4, P_5\}$ with n=5 elements, too. The number of different multisets is $\binom{n+k-1}{k} = \binom{14}{10}$.

Discrete Probability Spaces VI

- P_1 gets no job: 10 jobs must be partitioned to 4 processors, i.e. $\binom{13}{10}$ favourable partitions $\Rightarrow p = \frac{\binom{13}{10}}{\binom{14}{10}} = \frac{2}{7}$
- Every processor gets at least one job: at first every processor get one job then 5 jobs must be partitioned to 5 processors, i.e. $\binom{9}{5}$ favourable partitions $\Rightarrow p = \frac{\binom{9}{5}}{\binom{14}{10}} = \frac{18}{143}$
- Exactly one processor gets no job: at first choose the processor with no jobs then partition the 10 jobs to the remaining 4 processors that every processor get at least one job, i.e. $p = 5 \cdot \frac{\binom{9}{3}}{\binom{14}{10}} = \frac{60}{143}$
- P_1 get 3 jobs, P_2 get 2 jobs, P_3 get 2 jobs, P_4 gets 1 job and P_5 get 2 jobs: choose of one special partitioning in the set of all partitionings, i.e. $p = \frac{1}{\binom{14}{10}} = \frac{1}{1001}$

Discrete Probability Spaces VI

Example: Buffon's coin experiment[1]

- dropping a coin randomly on a floor covered with identically shaped tiles with side length 1
- event: the coin crosses a crack between tiles

What is $\Omega, \mathcal{A}$ and P?

Assumptions:

- The coin is a perfect circle with radius r.
- The cracks between tiles are line segments.

[1] compare http://www.fmi.uni-sofia.bg/vesta/Virtual_Labs/buffon/buffon1.html

Discrete Probability Spaces VII

Idea: Not the tiles where the coin crosses the crack are important. We are only interested at the event of crossing a crack.

$\Rightarrow$ Record the center of the coin relative to the center of the tile where the coin happens to fall.

$$\Omega = [-1/2, 1/2]^2 = \{(x, y) \mid -1/2 \leq x \leq 1/2, -1/2 \leq y \leq 1/2\}$$

Assume $r < 1/2$ otherwise the coin will allways touch a crack.

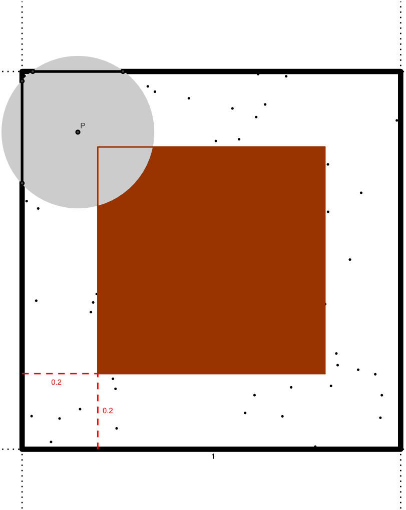
$\mathcal{A}$: Set generated by the set
$\{[a, b] \times [c, d] \mid a, b, c, d \in [-0.5, 0.5]\}$

Appropriate probability measure P: assume that (X, Y) is uniformly distributed on S, i.e.

$$P[(X, Y) \in A] = \frac{\text{area(A)}}{\text{area}(S)} \quad \text{for } A \in S.$$

$\Rightarrow$ Buffon_Coin_eng.gbb

Discrete Probability Spaces VIII



A coin with radius r is dropped on a tiled floor. The tiles are quadratic with length 1.
What is the probability, that the coins touches a crack between the tiles?
Is the centerpoint of the coin in the red square, the coin lies on a tile completely.
The sample space of the centerpoints of the coins is the black square.

$$p = \frac{\text{area red square}}{\text{area black square}} = 1 - (1 - 2r)^2 = 0.64$$

For n = 80 random drops of a coin the relative frequency is $r_{coin} = 0.64$

refresh

r = 0.2

n = 20

additional coin(s)

When does the coin touches a seam? Push the button "additional coin(s)" to drop n additional coins randomly. The last thrown coin and and the center points of all dropped coins are shown.

Change the radius r of the coin and the number n to see the influence of these paramters at p and r_{coin}.

Buffon_Coin_engl.gbb

What is the probability that the coin crosses a crack?

Questions

Two distinct fair dices are thrown.

Which of the following statements are true or false?

t	f	
☐	☐	The probability that the dice show a 2 and 5 is 1/36.
☐	☐	The probability that the dice show the same odd number is 1/6.
☐	☐	The probability that the dice show the same number is 1/6.
☐	☐	The probality that the scores of dices sum up to 6 is 5/36.
☐	☐	The probability that the dices show distinct numbers is 5/6.
☐	☐	The probability of no six is 5/6.

Conditional Probability and Independence I

Example: 3 tosses of a fair coin

- $\Omega = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$
- 2 or more heads in 3 tosses: $A = \{hhh, hht, hth, thh\}$
  $\Rightarrow P(A) = 1/2$
- What is the probability of A, given that the first toss lands head (event B)?
  $P(A|B) = 3/4$
  Notation: $P(A|B) =$ denotes the probability that event A will occur given that event B has occurred already.
- Probability of A, given that the first toss lands tail = $P(A|B^c) = 1/4$
- $P(A \cap B) = 3/8, P(B) = 1/2 : P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3}{4}$

Conditional Probability and Independence II

Definition: Let $(\Omega, \mathcal{A}, P)$ be a probablity space and $A, B \in \mathcal{A}$ with $P(B) > 0$. The conditional probability of A given B is

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

Multiplication Rule: $P(A \cap B) = P(A|B)P(B)$
Generally: Suppose that $A_1, A_2, ..., A_n$ is a sequence of events in a random experiment whose intersection have positive probabilities then

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)...P(A_n|A_1 \cap A_2 \cap ... \cap A_{n-1})$$

Definition: Let $(\Omega, \mathcal{A}, P)$ be a probablity space and $A, B \in \mathcal{A}$. If $P(A \cap B) = P(A)P(B)$ then A and B are called independent.

Conditional Probability and Independence III

Example from Mathematica Demonstrations:

- Consider drawing two balls with or without replacement from an urn containing blue and red balls.
- Random variable X: number of red balls
- possible values of X: {0 red balls,1 red ball,2 red balls}
- Urn sampling with or without replacement

Source: “Urn Sampling with or without Replacement”

http://demonstrations.wolfram.com/UrnSamplingWithOrWithoutReplacement/

Conditional Probability and Independence IV

Example: 2 programs P_1, P_2 should run on a workstation in multitasking.

- Consecutively every program runs a certain time period on the workstation.
- This will be continued cyclicly until both programs will be completed.
- The probability that P_1 will be finished after the first step ist 0.3.
- If P_1 not completed after the first step P_1 is completed after the second step with probability 0.7.
- If P_1 is not completed after the second step P_1 is completed after the third step with probability 0.8.

What is the probability that P_1 is not finished after 3 steps?

Conditional Probability and Independence V

Example: compare Heumann, Schomaker, p. 125, exercise 6.8

- There are epidemics which affect cows.
- Let the probability of the event that a cow has been transported by truck recently be 0.5.
- Let be the probability that a cow has been infected with a virus be 0.3.
- Let be the probability that a cow has been infected by a virus and has been transported by a truck recently be 0.2.

Conditional Probability and Independence VI

Find the following probabilities

1. a cow is infected or has been transported by a truck recently
2. a cow has been tranported by a truck recently and is not infected
3. a cow is not infected and has not been transported by a truck recently
4. a cow is infected given that it has been transported by a truck recently
5. a cow has not been transported by a truck recently given that the cow is infected.

Conditional Probability and Independence VII

- A: event that a cow has been transported by a truck recently
- B: a cow has been infected by a virus

If we take a random sample of 100 cows out of the underlying population we would expect the following frequencies:

truck	virus B	B^c	sum
A	20	30	50
A^c	10	40	50
sum	30	70	100

1. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{50+30-20}{100} = 0.6$
2. $P(A \cap B^c) = P(A) - P(A \cap B) = \frac{50-20}{100} = 0.3$
3. $P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cap B) = \frac{40}{100} = 0.4$
4. $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{20}{50} = 0.4$
5. $P(A^c|B) = \frac{P(A^c \cap B)}{P(B)} = \frac{10}{30} = 0.33$

Conditional Probability and Independence VIII

Example: Birthday Problem

- N students in a class.
- Probability that at least two students in the class with the same birthday?
- We do not regard variations in the distribution, such as leap years, twins, seasonal or weekday variations.
- We assume that the 365 possible birthdays are equally likely. Real-life birthday distributions are not uniform since not all dates are equally likely.

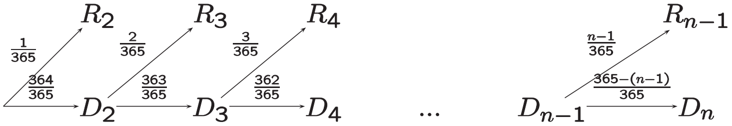
Conditional Probability and Independence IX

- Order the students in some arbitrary way.
- Go through the list of students birthday in that order and check whether or not each birthday is one that appeared previousley.
- If you find a repeat birthday in this process, stop.

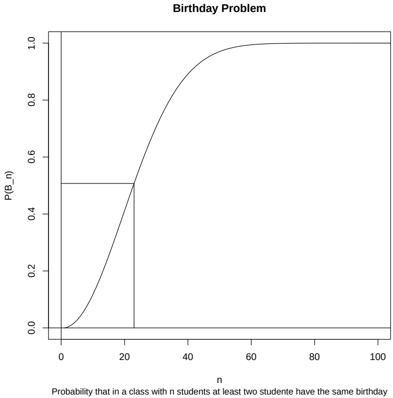
$R_j =$	the checking process stops with a repeat birthday at the jth student on the list
$D_j =$	the first j birthdays are different
$B_n =$	at least two students in the class have the same birthday

$$B_n = R_2 \cup R_3 \cup \ldots \cup R_n$$

$$P(B_n) = P(R_2) + P(R_3) + \ldots P(R_n) = 1 - P(D_n)$$



Conditional Probability and Independence X



n	P(D_n)	P(B_n)
2	0.99726	0.00274
10	0.88305	0.11695
20	0.58856	0.41144
23	0.49270	0.50730
30	0.29368	0.70632
50	0.02963	0.97037
70	0.00084	0.99916
365	0.00000	1.00000

$P(B_n)$ increases rapidly as n increases. The least n such that $P(B_n) > 0.5$ is $n = 23$.

Conditional Probability and Independence XI

Example: Hash-table

- Hash function with m different possible values
- Every value of the hash function has the same robability.
- If n objects should be stored in the hash table what is the probability that at least one collision ocures?

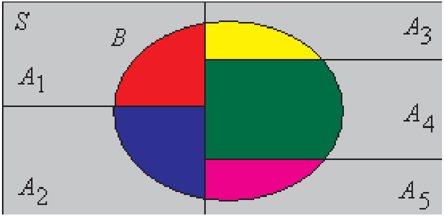
  $n \geq m$: at least one collision ocures.

  $n < m$: A = at least one collision, i.e. A^c = no collision
  The problem is equivalent to the birthday problem with m instead of 365 days.

$$P(A) = 1 - P(A^c) = 1 - (1 - \frac{1}{m})(1 - \frac{2}{m})...(1 - \frac{n-1}{m})$$

Conditional Probability and Independence XII

Suppose that $\{A_i | i \in I\}$ is a countable collection of events that partition the sample space S, and let B be another event.



$$\bigcup_{i \in I} A_i = \Omega,$$

A_i pairwise disjoint

Law of Total Probability:

$$P(B) = \sum_{i \in I} P(B \cap A_i) = \sum_{i \in I} P(B|A_i)P(A_i)$$

Bayes' Rule:

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i \in I} P(B|A_i)P(A_i)} \quad (j \in I)$$

Conditional Probability and Independence XIII

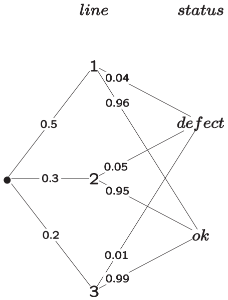
Example: A plant has 3 assembly lines that produces memory chips.

- Line 1 produces 50% of the chips and has a defective rate of 4%.
- Line 2 produces 30% of the chips and has a defective rate of 5%.
- Line 3 produces 20% of the chips and has a defective rate of 1%.
- A chip is chosen at random from the plant.

  What is the probability that the chip is defective?

  Given that the chip is defective, what is the conditional probability that the chips is produced at line i?

Conditional Probability and Independence XIV



1. Application of the law of total probability $\Rightarrow$ probability of a defective chip

$$
\begin{aligned}
P(\text{defect}) &= P(\text{defect}|\text{line 1})P(\text{line 1}) + \\
&\quad P(\text{defect}|\text{line 2})P(\text{line 2}) + \\
&\quad P(\text{defect}|\text{line 3})P(\text{line 3}) \\
&= 0.037
\end{aligned}
$$

2. Application of Bayes Rule $\Rightarrow$ conditional probabilities for each line given that the chip is defective

$$
\begin{aligned}
P(\text{line i}|\text{defect}) &= \frac{P(\text{defect}|\text{line i})P(\text{line i})}{P(\text{defect})} \\
&= \begin{cases} \frac{20}{37} & i = 1 \\ \frac{15}{37} & i = 2 \\ \frac{2}{37} & i = 3 \end{cases}
\end{aligned}
$$

Questions

In a population 30% of all persons are overweight, 20% of all persons suffer from high blood pressure and 10% of all person are overweight and suffer from high blood pressure.

Which of the following statements are true or false?

t	f	
☐	☐	The probability that a person suffers from high blood pressure if he is overweight is 1/2.
☐	☐	The probability that a person is not overweight and has no high blood pressure is 0.4.
☐	☐	50% of all persons with high blood pressure are overweight.
☐	☐	2/3 of all overweight persons have no high blood pressure.
☐	☐	Only 1/7 of all non overweight persons have high blood pressure.

Random Variables and Distributions I

compare Online Statistics → V 8, 12 and VII 1-6, Virtual Lab of Probability and Statistics → Distributions: 1, 2, 6 and Expected Value: 1, 2

Random Variable: Function that associates an outcome of a random experiment. Typically the the values of the function are real numbers.

Random Variables and Distributions II

- **Example:** A coin is tossed ten times.
  - Random variable X = number of tails
  - possible values of X: 0, 1, ..., 10
  - X is a discrete random variable.

- **Example:** A light bulb is burned until it burns out.
  - Random variable Y = lifetime in hours.
  - possible values of Y: any positive real value
  - Y is a continuous random variable.

Random Variables and Distributions III

Formal Definition:

- Random experiment with sample space S and σ-algebra of events $\mathcal{A}$.
- T a set with $\sigma-$algebra $\mathcal{B}$ of admissible subsets.

$(X : S \to T)$ is called a **random variable**, if

$$\{s \in S | X(s) \in B\} \in \mathcal{A} \quad \text{for every} \quad B \in \mathcal{B}$$

Remarks:

- Interpretation: X = measurement of interest in the context of the random experiment
- If the outcome of the random experiment is $s \in S$, X takes on the value $X(s)$. Thus the value of X cannot be predicted.
- **Notation:** $\{X \in B\} = \{s \in S | X(s) \in B\}$
- Probability space $(S, \mathcal{A}, P)$: for every $B \in \mathcal{B}$ the probability $P(X \in B)$ is defined.
- Typically the range of a random variable is $\mathbb{R}$ or a subset of $\mathbb{R}$.

Random Variables and Distributions IV

Example: Throwing two dices, Sum of two dice

$X : S \to T; X((i,j)) = i + j$ with $S = \{(i,j)|1 \leq i,j \leq 6\}$, $T = \mathbb{N}$

Die 1	Die 2	X	Die 1	Die 2	X	Die 1	Die 2	X
1	1	2	3	1	4	5	1	6
1	2	3	3	2	5	5	2	7
1	3	4	3	3	6	5	3	8
1	4	5	3	4	7	5	4	9
1	5	6	3	5	8	5	5	10
1	6	7	3	6	8	5	6	11
2	1	3	4	1	5	6	1	7
2	2	4	4	2	6	6	2	8
2	3	5	4	3	7	6	3	9
2	4	6	4	4	8	6	4	10
2	5	7	4	5	9	6	5	11
2	6	8	4	6	10	6	6	12

$\Rightarrow$

i	P(X=i)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Probability that the sum of the two dice will be

- 6: $P(X = 6) = 5/36$
- greater than 9: $P(X > 9) = 1 - P(X \leq 9) = 6/36$

Random Variables and Distributions V

Definition: X real-valued random variable on $(\Omega, \mathcal{A}, P)$.

$$F : \mathbb{R} \to [0,1], F(x) = P(X \leq x) \qquad (x \in \mathbb{R})$$

is called the distribution function of X.

Properties:

- $\lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1$
- F is monoton increasing
- F is continuous from the right
- $P(a < X \leq b) = F(b) - F(a)$
- $P(X > a) = 1 - F(a)$
- $P(X = a) = F(a) - \lim_{x > 0, x \to 0} F(a - x)$

Random Variables and Distributions VI

Independence: The random variables X, Y are called independent, if

$$P(X \le x, Y \le y) = P(X \le x) \cdot P(Y \le y) \quad \text{for all } x, y \in \mathbb{R}$$

Example:

- Two cards are chosen from a card deck with 32 cards.
- X = value of the first card
- Y = value of the second card.
- X, Y are independent, if the cards are chosen with replacement, and dependent if the cards are chosen without replacement.

Questions

Which of the following statements are true or false?

t	f	
☐	☐	The possible values of the random variable, which counts the occurency of sixes in rolling a die ten times are 0 and 6.
☐	☐	The distribution function of a real valued random variable is a strictly monotonous function.
☐	☐	If F is the distribution function of the random variable X we allways have $P(a < X \leq b) = F(b) - F(a)$ for all $a, b \in \mathbb{R}$ with $a \leq b$.
☐	☐	The number of throws of getting a 6 the first time a die is rolled and the number of throws of getting a 6 the second time are independent.
☐	☐	The number of throws of getting a 6 the first time a die is rolled and the sum of the numbers seen on the first and second trial is 8 are not independent.

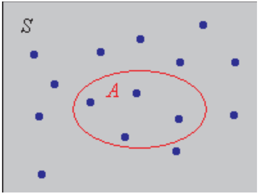
Discrete Distributions I

- Random experiment with probability space $(\Omega, \mathcal{A}, P)$
- A random variable X for the experiment that takes values in a countable set S is said to have a **discrete distribution.**
- **discrete probability density function of X:** function $f : S \to \mathbb{R}$ defined by

$$f(x) = P(X = x), \quad x \in S$$

The blue dots represent points of positive probability.

$$P(X \in A) = \sum_{x \in A} f(x), \quad A \subseteq S$$



Discrete Distributions II

- **Discrete Uniform Distribution:** An element X is chosen at random from a finite set S. All outcomes are equally likely, i.e. $f(x) = \frac{1}{|S|}, x \in S$.

  Examples: tossing a fair coin, rolling a fair die
- **Bernoulli Trial:** Random variable X considers the occurence of a certain event A. 1 denotes occurence of A (success) while 0 denotes non-occurence of A (failure). $p = P(X = 1)$ is the probability of success.

  Example: Rolling a fair die, A = success, i.e. 6, $p = 1/6$

Random Variables and Distributions III

Binomial Distribution - B(n,p):

- n independent repeated bernoulli trials $X_1, X_2, ..., X_n$
- $X = X_1 + X_2 + ... + X_n$ counts the number of successes in n trials
- $p = P(X_i = 1), 1 \leq i \leq n$.
- probability density function: $k \in \{0, 1, ..., n\}$.

$$P(\text{k successes in n trials}) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Example: Probability p of getting exactly 2 times a 6 in rolling a fair die 10 times

$$p = \binom{10}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^8 \approx 0.29071$$

Random Variables and Distributions IV

Hypergeometric Distribution - H(n,M,N):

- Population with of N objects
- M of the objects are type 1 and N-M are type 0.
- Sample of n objects is chosen at random (without replacement).
- X = number of type 1 objects in the sample.
- probability density function: $k \in \{0, 1, ..., M\}$

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

Remark: In case of sampling with replacement, we get a binomial distribition:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{with } p = \frac{M}{N}$$

Thus the hypergeometric distribution can be approximated by a binomial distribution for large N.

Random Variables and Distributions V

Example: An urn contains 30 red balls, 15 black balls and 5 white balls. 10 balls are drawn randomly one after the other from the urn.
What is the probability that

1. of the 10 drawn balls 6 are red.
2. of the 10 drawn balls 6 are red, 3 are black and 1 is white
3. the 10th ball is the first white ball in the sample.

Random Variables and Distributions VI

X = number of red balls in the sample

- **with replacement:**
  - 10 independent identical repetitions of the same random experiment (drawing a ball out of the urn)
  - Probability to get a red ball = 3/5
  - $X \sim B(n = 10, p = \frac{30}{50})$

$$P(X = 6) = \binom{10}{6} \left(\frac{3}{5}\right)^6 \left(\frac{2}{5}\right)^4$$

Random Variables and Distributions VII

- **without replacement:**
  - The composition of the ball changes after drawing one ball out of the urn
  - Every outcome of drawing ten balls out of the urn of the possible $\binom{50}{10}$ drawing is equally like.
  - From the 30 red balls 6 must be chosen and from the remaining 30 balls 4 must be chosen.
  - $X \sim H(n = 10, M = 30, N = 50)$

$$P(X = 6) = \frac{\binom{30}{6}\binom{20}{4}}{\binom{50}{10}}$$

Random Variables and Distributions VIII

X_r = number of red balls, X_b = number of black balls, X_w = number of white balls

- **with replacement:**
  - 10 independent identical repetitions of the same random experiment (drawing a ball out of the urn).
  - probabilities of drawing a specific ball are: $p_r = \frac{30}{50}$ for a red ball, $p_b = \frac{15}{50}$ for a black ball and $p_w = \frac{5}{50}$ for a white ball.
  - Only the final number of the colors and not the order of the colors in the sample is relevant.

$$P(X_r = 6, X_b = 3, X_w = 1)$$
$$= \frac{10!}{6! \cdot 3! \cdot 1!} \cdot \left(\frac{30}{50}\right)^6 \cdot \left(\frac{15}{50}\right)^3 \cdot \left(\frac{5}{50}\right)^1$$

  - **multinomial distribution**

Random Variables and Distributions IX

- **without replacement:**
  - The composition of the balls changes after drawing one ball out of the urn.
  - Every outcome of drawing ten balls out of the urn of the possible $\binom{50}{10}$ drawing is equally like.
  - From the 30 red balls 6 must be chosen, from the 15 black balls 3 must be chosen and from the 5 white balls 1 must be chosen.
  - Distribution of $\{X_r, X_b, X_w\}$: generalisation of the hypergeometric distribution:

$$P(X_r = 6, X_b = 3, X_w = 1) = \frac{\binom{30}{6} \cdot \binom{15}{3} \cdot \binom{5}{1}}{\binom{50}{10}}$$

Random Variables and Distributions X

X = number of the draw with the first white ball

- **with replacement:**
  - 10 independent identical repetitions of the same random experiment (drawing a ball out of the urn).
  - Probability to draw a white ball: $p = \frac{5}{50}$.
  - If the ten'th is the first white ball in the first nine drawing are no white balls but the next drawn ball is a white.

$$P(X = 10) = (1 - \frac{5}{50})^9 \cdot \frac{5}{50}$$

  - **geometric distribution**

Random Variables and Distributions XI

- **without replacement:**
  - The composition of the balls changes after drawing one ball out of the urn.
  - Every outcome of drawing ten balls out of the urn of the possible $\binom{50}{10}$ drawing is equally like.
  - The first 9 balls are not white. Thus there must 9 not white balls out of 45 not white balls be chosen. Since the next ball must be a white ball:

$$P(X = 10) = \frac{\binom{45}{9} \cdot \binom{5}{0}}{\binom{50}{9}} \cdot \frac{5}{41}$$

Questions

A fair die is rolled 5 times.
Which of the following statements are true or false?

t	f	
☐	☐	P(first die = 3 and second die = 5)=1/36
☐	☐	P(3 times a 1)=3/6
☐	☐	P(first 6 in the fifth throw)=$(\frac{5}{6})^4 \cdot \frac{1}{6}$
☐	☐	P(2 times a 2 and 1 time a 5) = $\frac{5!}{2!1!2!} \cdot (\frac{1}{6})^2 \cdot \frac{1}{6} \cdot (\frac{4}{6})^2$
☐	☐	P(only numbers less equal 2) = 1/3

Quantile I

Definition:

- X random variable with distribution function F
- $p \in (0, 1)$
- Quantile of order p for the distribution F = $\tilde{x}_p$ is defined by

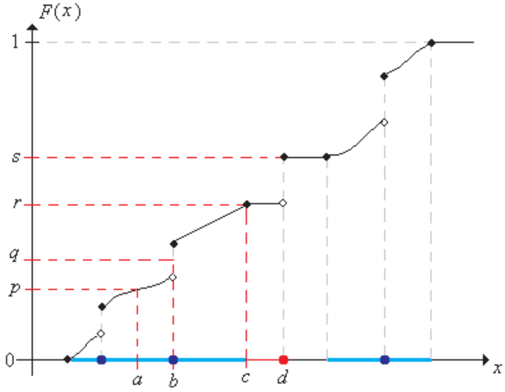
$$F(x) = \begin{cases} \leq p & \text{if } x < \tilde{x}_p \\ \geq p & \text{if } x \geq \tilde{x}_p \end{cases}$$

Remark:

- Quantile of order p is a value where the graph of the distribution function crosses (or jumps over) p.
- In case of a strictly monotonously increasing distribution function the quantiles are uniquely defined.

Quantile II

For example, in the picture below, a is the unique quantile of order p and b is the unique quantile of order q. On the other hand, the quantiles of order r form the interval $[c, d)$, and moreover, d is a quantile for all orders in the interval $[r, s]$.



). of Virtual Prob.)

Expectation, Variance I

Definition: The **expectation** or **expected value** $E(X)$ of a discrete random variable X is

$$E(X) = \sum_{all\ x} xP(X = x)$$

Remark:

- E(X) = average of all possible values of X, weighted by their probabilities.
- Sometimes E(X) is called mean of X.
- Compare: average $\bar{x}$ of a sample $(x_1, ..., x_n)$

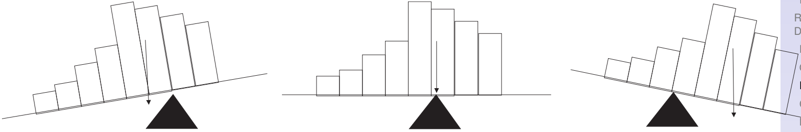
$$\bar{x} = (x_1 + ... + x_n)/n = \sum_{all\ x} xP_n(x)$$

where $P_n(x)$ is the relative frequency of x.

Expectation, Variance II

Remark: Pitman 3.2., p. 162

Interpretation: The mean is the center of gravity of a histogram



Expectation, Variance III

Example Rolling a die:

$X =$ number produced by rolling a fair die

$$E(X) = 1 \cdot P(X = 1) + \ldots + 6 \cdot P(X = 6) = \frac{1}{6} + \ldots + \frac{6}{6} = \frac{7}{2}$$

- Assumming a large number of independent rolls, we expect intuitively that each of the relative frequencies of the numbers is likely to be very close to 1/6.
- $\bar{x} = 1 \cdot P_n(X = 1) + \ldots + 6 \cdot P_n(X = 6)$ is assumed to be close to $E(X) = 7/2$.

Expectations as a Long-Run Average: If probabilities for values of X are approximate long-run frequencies, then $E(X)$ is approximately the long-run average value of X.

Remark: A precise formulation is given in the **law of large numbers**.

Expectation, Variance IV

Properties:
For any two random variables X and Y

$$E(X + Y) = E(X) + E(Y)$$

no matter whether X and Y are independent or not.
Example: Let $T = X_1 + X_2$ be the sum of numbers from 2 fair dice

$$E(T) = E(X_1 + X_2) = E(X_1) + E(X_2) = 7$$

Expectation, Variance V

Let G be any real-valued function defined on the set of possible values of X. Then

- $E(g(X)) = \sum_{all\ x} g(x)P(X = x)$
- Typically, $E(g(X)) \neq g(E(X))$
  Example: X number from a fair die

$$E(X^2) = \frac{1+4+9+16+25+36}{6} = \frac{91}{6} \neq \frac{49}{4} = (E(X))^2$$

- But the expectation of a linear function of X is determined by:

$$E(aX + b) = aE(X) + b \quad \text{for all } a, b \in \mathbb{R}$$

Expectation, Variance VI

If X and Y are independent then

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

Remark:

- Not true in general for dependent variables
- or example, if $X = Y$ then $E(X \cdot Y) = E(X^2)$ and $E(X) \cdot E(Y) = (E(X))^2$ but from above typically $E(X^2) \neq (E(X))^2$.

Expectation, Variance VII

Variance and Standard Deviation

Definition: The variance of X, denoted $Var(X)$, is the expected value of the squared deviation of X from the expected value $E(X)$:

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

Remark:

- Standard deviation of X = square root of the variance of X:
$$\sqrt{Var(X)}.$$
- The variance of a random variable gives an idea of how widely spread the values of the random variable are likely to be. It gives an impression of how closely concentrated round the expected value the distribution is.

Expectation, Variance VIII

Remarks:

- A useful computational formula is:

$$Var(X) = E(X^2) - (E(X))^2$$

- For example, let X be the number on a fair die:

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= (1^2 + 2^2 + ... + 6^2)/6 - (7/2)^2 = 35/12 \end{aligned}$$

- $Var(aX + b) = a^2 \cdot Var(X)$ for all $a, b \in \mathbb{R}$

Expectation, Variance IX

Addition Rule for Variances

If X and Y are independent random variables, then

$$Var(X+Y) = Var(X) + Var(Y)$$

Remark: In contrast to expectations, variances do not always add for dependent random variables. For example, if $X = Y$, then

$$Var(X+Y) = Var(2X) = 4\,Var(X)$$

while

$$Var(X) + Var(Y) = Var(X) + Var(X) = 2\,Var(X)$$

Expectation, Variance X

Some Expected Values and Variance:

- X uniformly distributed on $\{a, a+1, ..., b\}$:

$$E(X) = \frac{a+b}{2}, Var(X) = \frac{(b-a+1)^2-1}{12}$$

- X binomialy distributed (n repeated Bernoulli trials with success probability p):

$$E(X) = np, Var(X) = np(1-p)$$

- X hypergeometricly distributed (n objects chosen from a population of m objects with r objects of type 1 and m-r object of type 0):

$$E(X) = \frac{nr}{m} = np \quad \text{with } p = \frac{r}{m}, Var(X) = n\frac{m-n}{m-1}p(1-p)$$

Questions

An urn contains 2 red and 8 black balls. Two balls are randomly drawn from the urn. The random variable X counts the number of red balls.

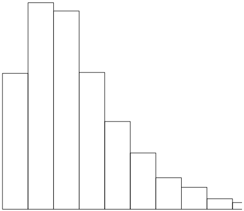
Which of the following statements are true or false?

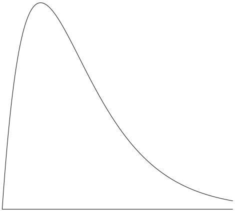
t	f	
☐	☐	50% quantile of X = 1
☐	☐	E(X) can be interpreted as the long-run average of the number of red balls, if we repeat the random experiment many times.
☐	☐	E(X)=0.4
☐	☐	$E(X^2) = (E(X))^2$
☐	☐	Var(X)=0.32

Continuous Distributions I

compare http://www.milefoot.com/math/stat/rv-contpdf.htm

- Instead of random variables restricted to integer values they will now be able to take on any value in some interval of real numbers.
- Graphically: Instead of discrete bars of a frequency diagram we will have a (possibly piecewise) continuous function.





Continuous Distributions II

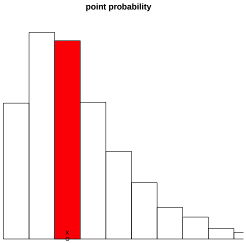
Comparison of Discrete and Continuous graphs

- discrete case: probabilities were given by a probability distribution function $P(X = x)$ graphically displayed by using its value as the area of the corresponding bar.
- continuous case: the probability $P(X = x)$ is displaced by the infinitesimal probability $f(x)dx$ of the event that X falls in an infinitesimal interval of length dx, for example $x \leq X \leq x + dx$
- $f(x)$ is called the probability density function
- Probabilities are determined by the areas under the curve $f(x)$.

Continuous Distributions III

point probability:

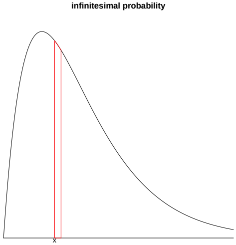
Discrete Distributions



$$P(X = x) = P(x)$$

P(x) is the probability that X has value x.

Continous Distributions



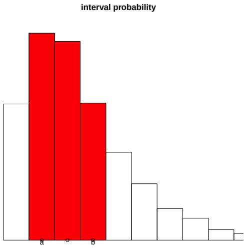
$$P(X \in dx) = f(x)dx$$

The density f(x) gives the probability per unit length for values near x.

Continuous Distributions IV

interval probability:

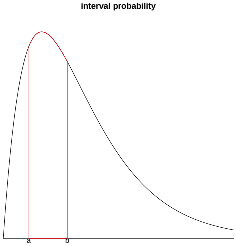
Discrete Distributions



$$P(a < X \leq b) = \sum_{a<x\leq b} P(x)$$

relative area under a histogram between a and b

Continous Distributions



$$P(X \in dx) = f(x)dx$$

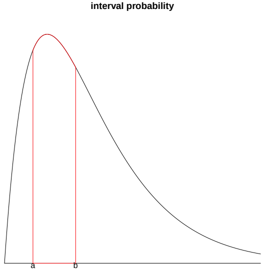
area under the graph of f(x) between a and b

Continuous Distributions V

Definition: A real valued random variable X is called continuously distributed with density f, if the distribution function F can be defined with a nonnegative function $f : \mathbb{R} \to \mathbb{R}$ by

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)dt, \quad x \in \mathbb{R}$$

$$P(a < X \leq b) = \int_{a}^{b} f(t)dt$$



Continuous Distributions VI

Properties:

- $\int_{-\infty}^{\infty} f(t)dt = 1$
- $P(X = x) = 0$ for all x
- $E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$, $\mathrm{Var}(X) = E((X - E(X))^2)$
- same properties for expectation and variance as for discrete random variables
- independency analagously defined as for discrete random variables

Continuous Distributions VII

Uniform Distribution

An element X is chosen at random from a real, restricted interval (a, b). The probability of outcomes in a certain interval $(x, y) \subseteq (a, b)$ depends on the relative length only:

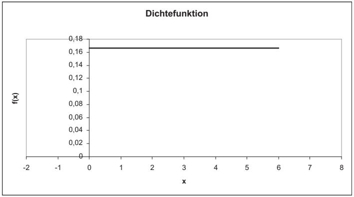
$$P(x < X < y) = \frac{\text{length}(x, y)}{\text{length}(a, b)} = \frac{y - x}{b - a}$$

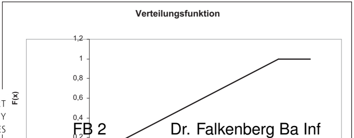
$$\begin{aligned} f(x) &= \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & else \end{cases} \\ E(X) &= \frac{a+b}{2} \\ Var(X) &= \frac{(b-a)^2}{12} \end{aligned}$$

Continuous Distributions VIII

Remarks:

- Generalization of discrete uniform distributions
- Every interval in (a,b) with the same length has the same probability.
- If buses arrive at a given bus stop every 6 minutes, and you arrive at the bus stop at a random time, the time you wait for the next bus to arrive could be described by a uniform distribution over the interval from 0 to 6.





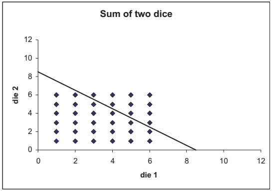
Continuous Distributions IX

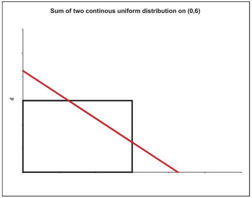
Sum of two continuous uniform distribution

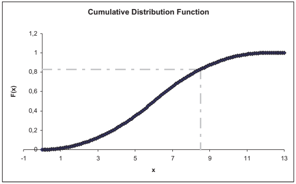
Let X_1, X_2 two independent on (0,6) continuously uniformly distributed random variables and $X = X_1 + X_2$. Analogously to the discrete case

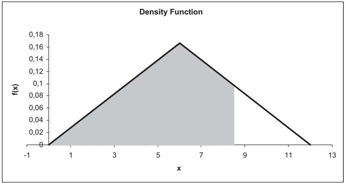
$$F(x) = P(X \leq x) = \frac{\text{area of possible values below } x_1 + x_2 = x}{\text{area of possible values}}$$

Continuous Distributions X









$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{0.5x^2}{36} & \text{if } 0 < x \leq 6 \\ \frac{36-0.5(12-x)(12-x)}{36} & \text{if } 6 < x \leq 12 \\ 1 & \text{if } x > 12 \end{cases}$$

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{x}{36} & \text{if } 0 < x \leq 6 \\ \frac{12-x}{36} & \text{if } 6 < x \leq 12 \\ 0 & \text{if } x > 12 \end{cases}$$

Questions

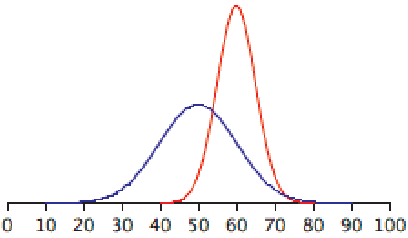
Let random variable X be uniformly distributed on the interval $[0, 10]$.
Which of the following statements are true or false?

t	f	
☐	☐	$P(X = 3) = 0$
☐	☐	$P(1 < X \leq 7.5) = 0.65$
☐	☐	For every subinterval I in $[0, 10]$ with length d we have $P(X \in I) = d/10$.
☐	☐	The density function of f of X is given by $f(x) = \begin{cases} 1 & 0 \leq x \leq 10 \\ 0 & \text{else} \end{cases}$
☐	☐	Var(X)=1/12
☐	☐	$P(X > 7.5 \mid X > 5) = 0.5$

Normal Distributions I

- Most important and most widely used continuous distributions in statistics.
- Densities bell shaped and symmetric with relatively more values at the center of the distribution and relatively few in the tails.
- Defined by two parameters: mean μ and standard deviation σ.

blue: mean = 50, standard deviation = 10
red: mean = 60 , standard deviation = 5

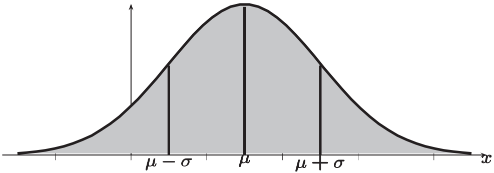


Normal Distributions II

Properties

- Density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



- f symmetric: $f(\mu + x) = f(\mu - x)$
- maximum at $x = \mu$
- inflection points at $\mu \pm \sigma$
- $E(X) = \mu$, $Var(X) = \sigma^2$

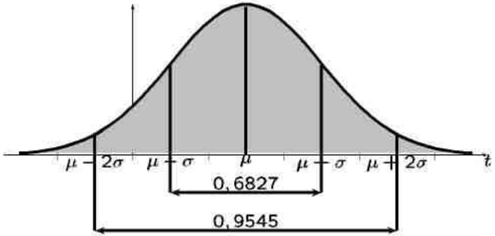
Normal Distributions III

Standard Normal Distribution

- Normal distribution with $\mu = 0$ and $\sigma = 1$
- Distribution function: $\Phi(.)$
- Areas of the normal distribution resp. values of $\Phi(x)$ are often represented by tables.
- $X \sim N(\mu, \sigma^2) \Rightarrow Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$
- $P\left(\dfrac{X - \mu}{\sigma} \leq \dfrac{x - \mu}{\sigma}\right) = \Phi\left(\dfrac{x - \mu}{\sigma}\right)$
- $P(X \leq \tilde{x}_p) = p \Leftrightarrow \tilde{x}_p = \mu + \sigma u_p$ with $\Phi(u_p) = p$.

Normal Distributions IV

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$



Let X be a normal distributed random variable, then

$$\begin{aligned}
P(\mu - \sigma \leq X \leq \mu + \sigma) &= 68,27\% \\
P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= 95,44\% \\
P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= 99,73\% \\
P(\mu - 4\sigma \leq X \leq \mu + 4\sigma) &= 99,99\%
\end{aligned}$$

Questions

Assume the speed of vehicles along a stretch of a highway has an approximately normal distribution with a mean of 71 mph and a standard deviation of 8 mph.

1. The current speed limit is 65 mph. What is the proportion of vehicles less than or equal to the speed limit?
2. What proportion of the vehicles would be going more than 70 mph?
3. What proportion of the vehicles would be going less than 70 mph and more than 50 mph?
4. A new speed limit will be initiated such that approximately 10% of vehicles will be over the speed limit. What is the new speed limit based on this criterion?
5. In what way do you think the actual distribution of speeds differs from a normal distribution?

Normal Distributions V

Prooperties:

- **Invariance under linear transformations**
  $X \sim N(\mu, \sigma^2)$, a and b constants with$a \neq 0$,
  $\Rightarrow aX + b \sim N(a\mu + b, a^2\sigma^2)$.
- **Invariance relative to sums of independent variables**
  $X_i \sim N(\mu_i, \sigma_i^2)$ for $i \in \{1, 2\}$, X_1 and X_2 independent
  $\Rightarrow X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
  Generalisation: sum of n independent, normal variables is normal.

Question

Online Statistics VII Exercise 5: Questionnaire to assess women's and men's attitudes toward using animals in research.

- One question: whether animal research is wrong to be answered on a 7-point scale.
- Assumption:
  - mean for women = 5
  - mean for men = 4
  - standard deviation for both groups = 1.5
  - scores normally distributed

If 12 women and 12 men are selected randomly, what is the probability that the mean of the women will be more than 1.5 points higher than the mean of the men?

Central Limit Theorem I

compare Virtual Lab of Probability and Statistics $\rightarrow$ Random Samples: 6

- Fundamental theorem of probability
- Roughly: distribution of the sum of a large number of independent, identically distributed variables approximately normal, regardless of the underlying distribution
- Extrmely important
- Reason that many statistical procedures work

Central Limit Theorem II

Example: Throwing n fair dice k-times

- X_i number of die i: X_i discrete uniform distributed on $\{1, \ldots, 6\}$
- X_i, X_j pairwise independent
- $Y_n = X_1 + \ldots + X_n$ sum of numbers in one throw

$\Rightarrow$

$$
\begin{aligned}
E(X_i) &= \tfrac{7}{2}, \text{Var}(X) = \tfrac{35}{12} \\
E(Y_n) &= \tfrac{7}{2} \cdot n, \text{Var}(Y_n) = \tfrac{35}{12} \cdot n \\
P(Y_n = i) &= P(Y_{n-1} + X_n = i); i = n, \ldots, 6n \\
&= \tfrac{1}{6} \cdot (P(Y_{n-1} = i-1) + \ldots + P(Y_{n-1} = i-6))
\end{aligned}
$$

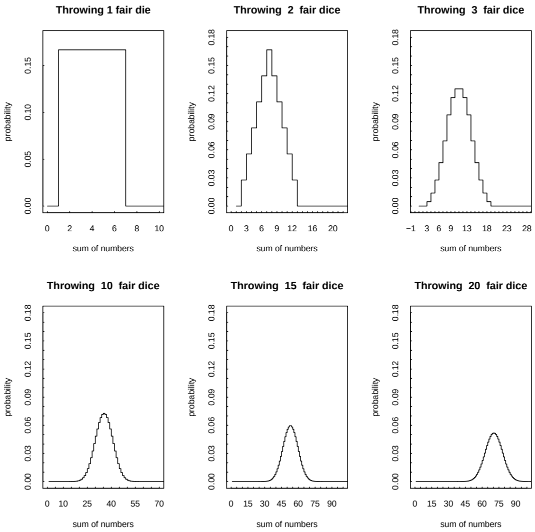
Central Limit Theorem III

In general:

- Basic experiment and a random variable X
- Mean and standard deviation of X: $\mu \neq 0$ and σ
- Repeat the experiment over and over: $X_1, X_2, X_3, \ldots$
  Sequence of independent random variables, each with the same distribution as X. Let
  $Y_n = X_1 + X_2 + \ldots + X_n$
- Notice $E(Y_n) = n\mu$ and $Var(Y_n) = n\sigma^2$
- $Var(Y_n) \to \infty$, $E(Y_n) \to \infty$
- Y_n itself does not have a limiting distribution
- Consider not Y_n itself, but the standard score of Y_n:

$$Z_n = \frac{Y_n - n\mu}{\sqrt{n}\sigma}$$

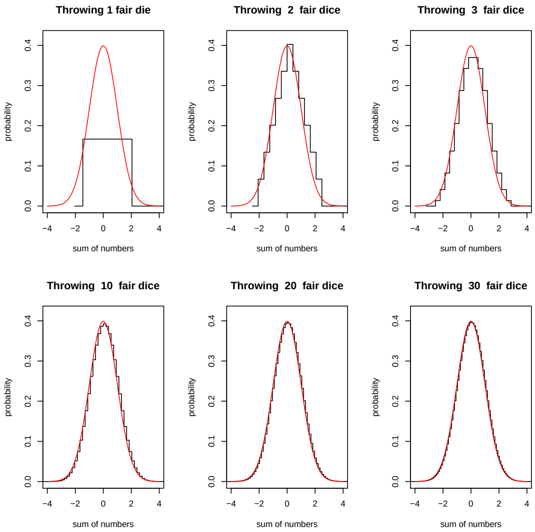
Central Limit Theorem IV



$Var(Y_n) \to \infty$, $E(Y_n) \to \infty$

Y_n itself does not have a limiting distribution

Central Limit Theorem V



It appears as if standardizing Y_n results in a limiting distribution. Illustrating the Central Limit Theorem with sums of Bernoulli random variables Source: Wolfram Demonstrations Project http://demonstrations.wolfram.com

Central Limit Theorem VI

Central limit theorem:

The distribution of Z_n converges to the standard normal distribution as n increases to infinity.

Remarks:

- We can approximate the distribution of certain statistics, even if we know very little about the underlying sampling distribution.
- If n is "large", then the distribution of Y_n (or equivalently the sample mean) is approximately normal.
- Of course, the term "large" is relative.
  - Roughly, the more "abnormal" the basic distribution, the larger n must be for normal approximations to work well.
  - Rule of thumb: sample size n of at least 30 will suffice, for many distributions smaller n will be sufficient

Central Limit Theorem VII

Limit theorem of de Moivre Laplace:

$$\lim_{n\to\infty} P\left(\frac{Y_n - np}{\sqrt{np(1-p)}} \leq x\right) = \Phi(x), \qquad x \in \mathbb{R}$$

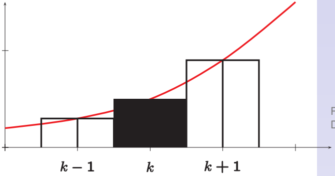
Remark: : Binomial distribution of the number of successes Y_n in n independent Bernoulli trials with probability p of success on each trial is approximately normal distribution if n is large.

Central Limit Theorem VIII

Normal Approximation of a Discrete Distribution:

- If $X \in \mathbb{N}$ then the partial sum $Y_n \in \mathbb{N}$
- Approximation of discrete distribution by continuous one
- $\{k - 0.5 < Y_n \leq k + 0.5\}$, $\{Y_n = k\}$ are equivalent

Continuity correction:

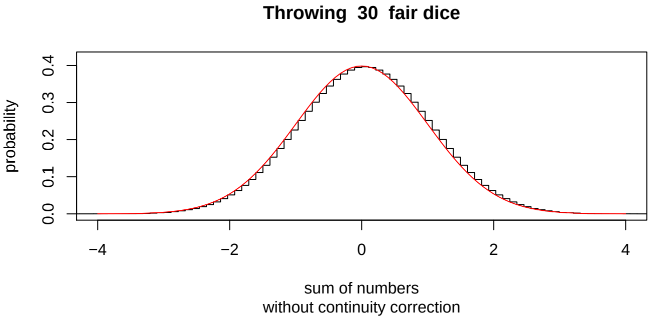


$$P(Y_n = k) \approx \Phi\left(\frac{k + 0.5 - n\mu}{\sqrt{n} \cdot \sigma}\right) - \Phi\left(\frac{k - 0.5 - n\mu}{\sqrt{n} \cdot \sigma}\right)$$

Extension the continuity correction using the additivity of probability:

$$P(Y_n \leq m) = \sum_{k \leq m} P(Y_n = k) \approx \Phi\left(\frac{m + 0.5 - n\mu}{\sqrt{n} \cdot \sigma}\right)$$

Central Limit Theorem IX



Central Limit Theorem X

The binomial distribution of the number of successes X in n independent Bernoulli trials with probability p of success on each trial can be approximated by a normal distribution if n is large:

$$P(X \leq m) \approx \Phi\left(\frac{m + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

Rule of thumb: approximation acceptable, if $np(1-p) \geq 9$

Content