Dataanalyse - Lecture 1: Probability

Ege Rubak - rubak@math.aau.dk
http://www.math.aau.dk/∼rubak/teaching/2010/nano4

8. februar 2010

Terminology

Sample space, U : The set of all possible outcomes.
Outcome, u : An element in the sample space, $u \in U$.
Event, A : A collection of outcomes, $A \subseteq U$.

Examples

Throw a die

Sample space : $U = \{1, 2, 3, 4, 5, 6\}$
An outcome : Get a 1, since $1 \in U$
An event : Get an even number, since $\{2, 4, 6\} \subseteq U$

Throw 2 dices

Sample space : $U = \{(m, n) : n, m = 1, 2, 3, 4, 5, 6\}$
$= \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), \ldots\}$
An outcome : Get two 6, since $(6, 6) \in U$
An event : Get two equal numbers, since
$\{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\} \subseteq U$

Terminology

Sample space, U : The set of all possible outcomes.

Outcome, u : An element in the sample space, $u \in U$.

Event, A : A collection of outcomes, $A \subseteq U$.

Examples

Throw a die

Sample space : $U = \{1, 2, 3, 4, 5, 6\}$

An outcome : Get a 1, since $1 \in U$

An event : Get an even number, since $\{2, 4, 6\} \subseteq U$

Throw 2 dices

Sample space : $U = \{(m, n) : n, m = 1, 2, 3, 4, 5, 6\}$
$= \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), \ldots\}$

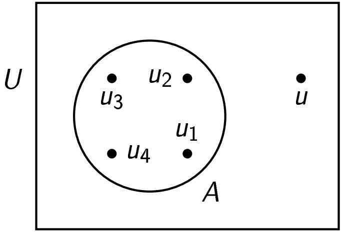
An outcome : Get two 6, since $(6, 6) \in U$

An event : Get two equal numbers, since
$\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\} \subseteq U$

Terminology

Sample space, U : The set of all possible outcomes.
Outcome, u : An element in the sample space, $u \in U$.
Event, A : A collection of outcomes, $A \subseteq U$.

We illustrate this with set drawings (Venn diagrams):



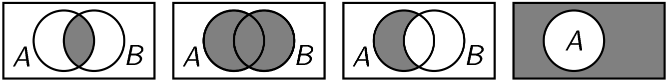
Here $A = \{u_1, u_2, u_3, u_4\}$.

Events

- If A and B are events, we can make new events:

Event	Notation
Both A and B occur.	$A \cap B$ (or A, B)
A or B (or both) occur.	$A \cup B$ (or $A + B$)
A occur, but B doesn't.	$A \setminus B$
A doesn't occur.	$\complement A$

- With set drawings:



$A \cap B$ $\quad$ $A \cup B$ $\quad$ $A \setminus B$ $\quad$ $\complement A$

Probabilities

- Let U be a sample space.
- P is a probability distribution on U, if for alle events A and B
  1. $0 \leq P(A) \leq 1$.
  2. $P(\emptyset) = 0$.
  3. $P(U) = 1$.
  4. $P(A \cup B) = P(A) + P(B)$, if $A \cap B = \emptyset$.
- $P(A)$ is the probability of observing A.

Conditional probability

- Let B be an event with $P(B) > 0$.
- The conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A \text{ if we know } B \text{ has occured})$$

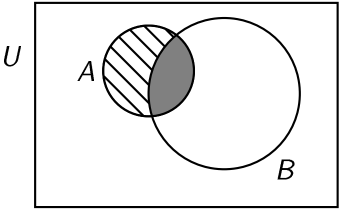
- We let B be a new (and smaller) sample space.

Conditional probability

- Let B be an event with $P(B) > 0$.
- The conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A \text{ if we know } B \text{ has occured})$$

- We let B be a new (and smaller) sample space.



Independence

- Two events A and B are independent if

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

- If $P(A) > 0$ and $P(B) > 0$:

$$P(A) = P(A|B) \quad \text{and} \quad P(B) = P(B|A)$$

- We don't get any information about A by knowing B (and vice versa).

Independence

- Two events A and B are independent if

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

- If $P(A) > 0$ and $P(B) > 0$:

$$P(A) = P(A|B) \quad \text{and} \quad P(B) = P(B|A)$$

- We don't get any information about A by knowing B (and vice versa).

Independence

- Two events A and B are independent if

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

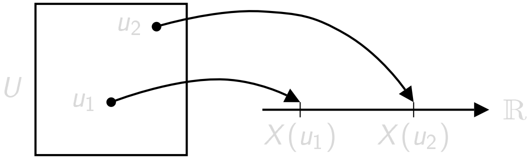
- If $P(A) > 0$ and $P(B) > 0$:

$$P(A) = P(A|B) \quad \text{and} \quad P(B) = P(B|A)$$

- We don't get any information about A by knowing B (and vice versa).

Stocastic variables

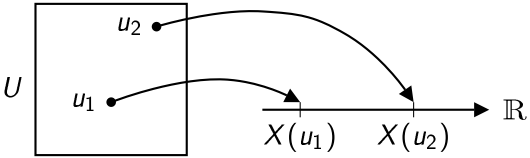
- Also called a random variable.
- Heuristic: A variable that takes different values with different probabilities.
- Rigorous: A function whoose set of definition is a sample space with a probability distribution



- X is discrete if it only takes countably many values.
- X is continuous if it takes uncountably many values.

Stocastic variables

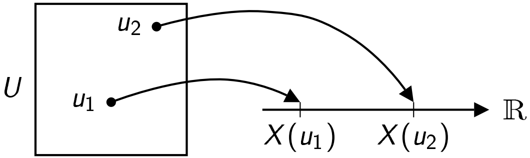
- Also called a random variable.
- Heuristic: A variable that takes different values with different probabilities.
- Rigorous: A function whoose set of definition is a sample space with a probability distribution



- X is discrete if it only takes countably many values.
- X is continuous if it takes uncountably many values.

Stocastic variables

- Also called a random variable.
- Heuristic: A variable that takes different values with different probabilities.
- Rigorous: A function whoose set of definition is a sample space with a probability distribution



- X is discrete if it only takes countably many values.
- X is continuous if it takes uncountably many values.

Stocastic variables

- Examples:

Experiment	Stocastic variable	Type
Throw a die	# eyes	discrete
Throw a die	$\sum$ eyes	discrete
Weigh a person	Weight	continuous
Measure men in DK	height	continuous

Distributions

- If X is a discrete stocastic variable, the probability function for X:

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

- If Y is a continuous stocastic variable, $f(y)$ is the density function for Y if

$$P(a \leq Y \leq b) = \int_a^b f(y)\,\mathrm{d}y$$

- Note:
  - $P(Y = c) = 0$ for any number c.
  - The "$\leq$" can be replaced by "$<$".

Distributions

- If X is a discrete stocastic variable, the probability function for X:

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

- If Y is a continuous stocastic variable, $f(y)$ is the density function for Y if

$$P(a \leq Y \leq b) = \int_a^b f(y)\,\mathrm{d}y$$

- Note:
  - $P(Y = c) = 0$ for any number c.
  - The "$\leq$" can be replaced by "$<$".

Distributions

- If X is a discrete stocastic variable, the probability function for X:

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

- If Y is a continuous stocastic variable, $f(y)$ is the density function for Y if

$$P(a \leq Y \leq b) = \int_a^b f(y)\,\mathrm{d}y$$

- Note:
  - $P(Y = c) = 0$ for any number c.
  - The "$\leq$" can be replaced by "$<$".

Characteristics

Two numbers are especially interesting for a stocastic variable X

- Mean value/**E**xpected value:

$$E(X)$$

- Variance – the expected deviation from the mean value:

$$\text{Var}(X) = E\big((X - E(X))^2\big) = E(X^2) - E(X)^2$$

Characteristics

Two numbers are especially interesting for a discrete stocastic variable X with probability function $f(x)$

- Mean value/**E**xpected value:

$$E(X) = \sum_{\text{outcome}} x f(x)$$

- Variance – the expected deviation from the mean value:

$$\begin{aligned} \mathrm{Var}(X) &= E\big((X - E(X))^2\big) = E(X^2) - E(X)^2 \\ &= \sum_{\text{outcome}} (x - E(X))^2 f(x) \\ &= \sum_{\text{outcome}} x^2 f(x) - E(X)^2 \end{aligned}$$

Characteristics

Two numbers are especially interesting for a continuous stocastic variable X with density function $f(x)$

- Mean value/**E**xpected value:

$$E(X) = \int_{\text{outcome}} x f(x)\,\mathrm{d}x$$

- Variance – the expected deviation from the mean value:

$$\begin{aligned}\mathrm{Var}(X) &= E\big((X - E(X))^2\big) = E(X^2) - E(X)^2 \\ &= \int_{\text{outcome}} (x - E(X))^2 f(x)\,\mathrm{d}x \\ &= \int_{\text{outcome}} x^2 f(x)\,\mathrm{d}x - E(X)^2\end{aligned}$$

Several stocastic variables

- X and Y are discrete stocastic variables with probability functions $f_X(x)$ and $f_Y(y)$, respectively.
- (X, Y) is a new stocastic variable.
- Joint probability function of X and Y, $f_{(X,Y)}(x, y)$:

$$P(a \leq X \leq b, c \leq Y \leq d) = \sum_{x=a}^{b} \sum_{y=c}^{d} f_{(X,Y)}(x, y).$$

- Marginal density:

$$f_X(x) = \sum_{y} f_{(X,Y)}(x, y) \quad , \quad f_Y(y) = \sum_{x} f_{(X,Y)}(x, y).$$

Several stocastic variables

- X and Y are discrete stocastic variables with probability functions $f_X(x)$ and $f_Y(y)$, respectively.
- (X, Y) is a new stocastic variable.
- Joint probability function of X and Y, $f_{(X,Y)}(x, y)$:

$$P(a \leq X \leq b, c \leq Y \leq d) = \sum_{x=a}^{b} \sum_{y=c}^{d} f_{(X,Y)}(x, y).$$

- Marginal density:

$$f_X(x) = \sum_{y} f_{(X,Y)}(x, y) \quad , \quad f_Y(y) = \sum_{x} f_{(X,Y)}(x, y).$$

Several stocastic variables

- X and Y are discrete stocastic variables with probability functions $f_X(x)$ and $f_Y(y)$, respectively.
- (X, Y) is a new stocastic variable.
- Joint probability function of X and Y, $f_{(X,Y)}(x, y)$:

$$P(a \leq X \leq b, c \leq Y \leq d) = \sum_{x=a}^{b} \sum_{y=c}^{d} f_{(X,Y)}(x, y).$$

- Marginal density:

$$f_X(x) = \sum_{y} f_{(X,Y)}(x, y) \quad , \quad f_Y(y) = \sum_{x} f_{(X,Y)}(x, y).$$

Several stocastic variables

- X and Y are continuous stocastic variables with densities $f_X(x)$ and $f_Y(y)$, respectively.
- (X, Y) is a new stocastic variable.
- Joint density of X and Y, $f_{(X,Y)}(x,y)$:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{(X,Y)}(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

- Marginal density:

$$f_X(x) = \int f_{(X,Y)}(x,y)\,\mathrm{d}y \quad , \quad f_Y(y) = \int f_{(X,Y)}(x,y)\,\mathrm{d}x.$$

Independence

- The conditional distribution of X given Y, $f_{X|Y}(x|y)$:

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)}$$

- X and Y are independent if

$$f_X(x) = f_{X|Y}(x|y)$$

or

$$f_{(X,Y)}(x,y) = f_X(x)f_Y(y).$$

Independence

- The conditional distribution of X given Y, $f_{X|Y}(x|y)$:

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)}$$

- X and Y are independent if

$$f_X(x) = f_{X|Y}(x|y)$$

or

$$f_{(X,Y)}(x,y) = f_X(x)f_Y(y).$$

Distribution function

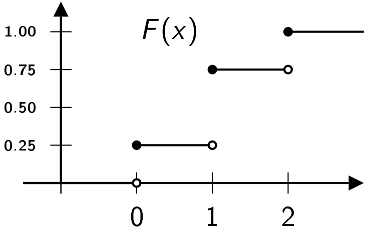
- X is a stocastic variable with function $f(x)$
- The cumulative distribution function for X:

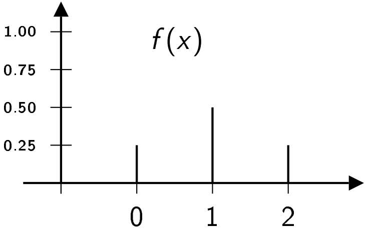
$$F_X(x) = P(X \leq x)$$

Distribution function

- X is a discrete stocastic variable with probability function $f(x)$
- The cumulative distribution function for X:

$$F_X(x) = P(X \leq x) = \sum_{y \leq x} f(y)$$



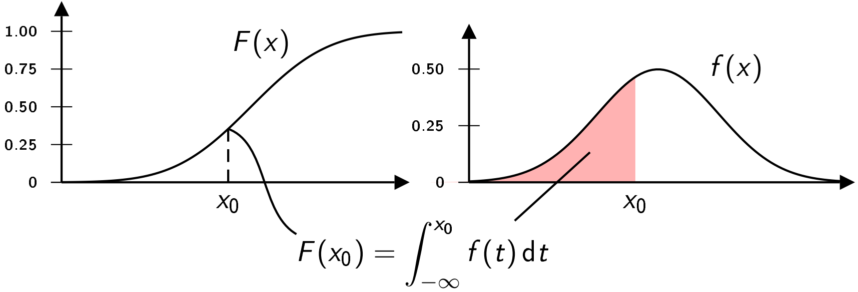


Note that $F_X(x)$ is defined for *all* $x \in \mathbb{R}$.

Distribution function

- X is a continuous stocastic variable with density function $f(x)$
- The cumulative distribution function for X:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

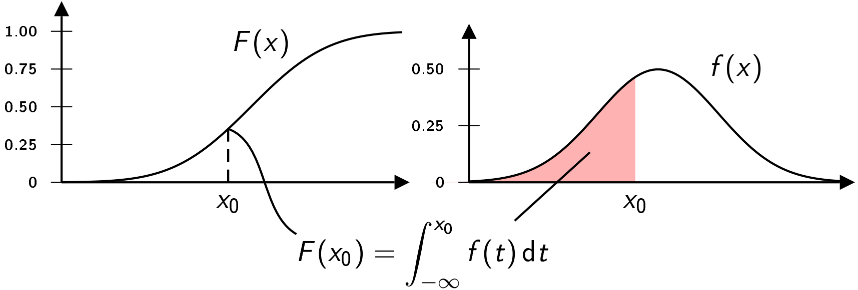


Note that $F'(x) = f(x)$

Distribution function

- X is a continuous stocastic variable with density function $f(x)$
- The cumulative distribution function for X:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$



Note that $F'(x) = f(x)$

Fractiles

- X is a stocastic variable with distribution function $F_X(x)$.
- For $0 < p < 1$ the p fractile is the number(s) x_p where

$$F(x_p) = p.$$

- 50% fractile is called the median og 2. quartil.
- 25% and 75% fractiles are called 1. and 3. quartile, respectively.
- Fractiles for the normal distribution

Fractiles

- X is a stocastic variable with distribution function $F_X(x)$.
- For $0 < p < 1$ the p fractile is the number(s) x_p where

$$F(x_p) = p.$$

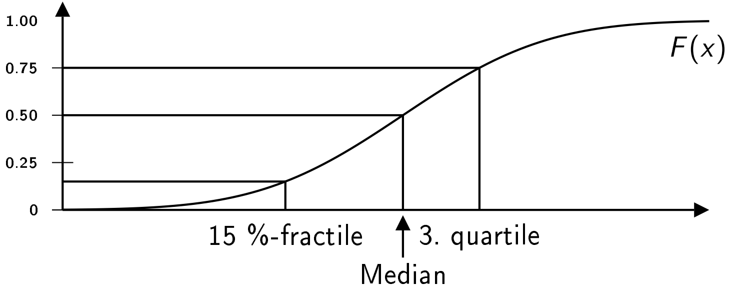
- 50% fractile is called the median og 2. quartil.
- 25% and 75% fractiles are called 1. and 3. quartile, respectively.
- Fractiles for the normal distribution

Fractiles

- X is a stocastic variable with distribution function $F_X(x)$.
- For $0 < p < 1$ the p fractile is the number(s) x_p where

$$F(x_p) = p.$$

- 50% fractile is called the median og 2. quartil.
- 25% and 75% fractiles are called 1. and 3. quartile, respectively.
- Fractiles for the normal distribution



Binomial distribution

- X is Bernoulli-distributed with parameter p if $X \in \{0, 1\}$ with $P(X = 1) = p$ og $P(X = 0) = 1 - p$.
- Let $X_1, X_2, \dots, X_n$ be independent and Bernoulli distributed with parameter p.
- $S = X_1 + X_2 + \cdots + X_n$ is binomial distributed with parameters n and p (Notation: $X \sim B(n, p)$).
- Properties:
  - S takes values in $\{0, 1, 2, \dots, n\}$.
  - S has probability function

$$P(S = k) = p^k (1 - p)^{n-k}.$$

Binomial distribution

- X is Bernoulli-distributed with parameter p if $X \in \{0, 1\}$ with $P(X = 1) = p$ og $P(X = 0) = 1 - p$.
- Let $X_1, X_2, \ldots, X_n$ be independent and Bernoulli distributed with parameter p.
- $S = X_1 + X_2 + \cdots + X_n$ is binomial distributed with parameters n and p (Notation: $X \sim B(n, p)$).
- Properties:
  - S takes values in $\{0, 1, 2, \ldots, n\}$.
  - S has probability function

$$P(S = k) = p^k (1 - p)^{n-k}.$$

Binomial distribution

- X is Bernoulli-distributed with parameter p if $X \in \{0, 1\}$ with $P(X = 1) = p$ og $P(X = 0) = 1 - p$.
- Let $X_1, X_2, \dots, X_n$ be independent and Bernoulli distributed with parameter p.
- $S = X_1 + X_2 + \cdots + X_n$ is binomial distributed with parameters n and p (Notation: $X \sim B(n, p)$).
- Properties:
  - S takes values in $\{0, 1, 2, \dots, n\}$.
  - S has probability function

$$P(S = k) = p^k (1 - p)^{n-k}.$$

Binomial distribution

- X is Bernoulli-distributed with parameter p if $X \in \{0, 1\}$ with $P(X = 1) = p$ og $P(X = 0) = 1 - p$.
- Let $X_1, X_2, \ldots, X_n$ be independent and Bernoulli distributed with parameter p.
- $S = X_1 + X_2 + \cdots + X_n$ is binomial distributed with parameters n and p (Notation: $X \sim B(n, p)$).
- Properties:
  - S takes values in $\{0, 1, 2, \ldots, n\}$.
  - S has probability function

$$P(S = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Binomial distribution

- X is Bernoulli-distributed with parameter p if $X \in \{0, 1\}$ with $P(X = 1) = p$ og $P(X = 0) = 1 - p$.
- Let $X_1, X_2, \ldots, X_n$ be independent and Bernoulli distributed with parameter p.
- $S = X_1 + X_2 + \cdots + X_n$ is binomial distributed with parameters n and p (Notation: $X \sim B(n, p)$).
- Properties:
  - S takes values in $\{0, 1, 2, \ldots, n\}$.
  - S has probability function

number of ways
to choose k of n

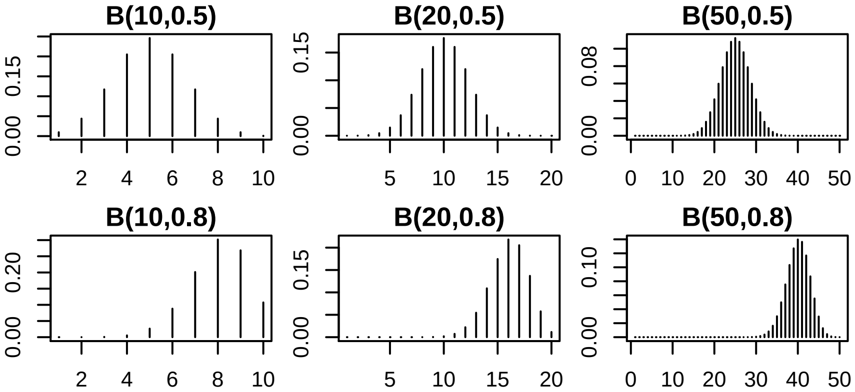
$$P(S = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Binomial distribution

- For $X \sim B(n, p)$:

$$E(S) = n \cdot p \quad \text{og} \quad \text{Var}(S) = n \cdot p \cdot (1 - p)$$

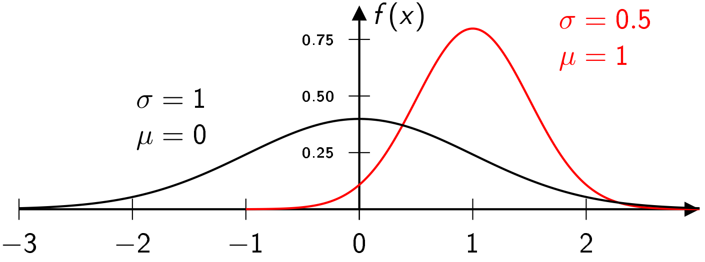
- Probability function with different parameters.



Normal distribution

- X is normal distributed with mean μ and variance σ^2 (notation: $X \sim N(\mu, \sigma^2)$) if is has density function

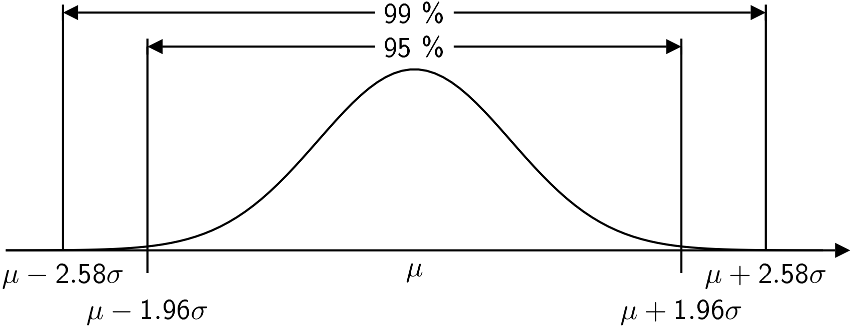
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$



- The normal distribution is the most important distribution you'll encounter!

Normal distribution

- $N(0,1)$ is the standard normal distribution.
- If $X \sim N(0,1)$ and $Y \sim \mu + \sigma X$, then $Y \sim N(\mu, \sigma^2)$.
- We have:



Normal distribution

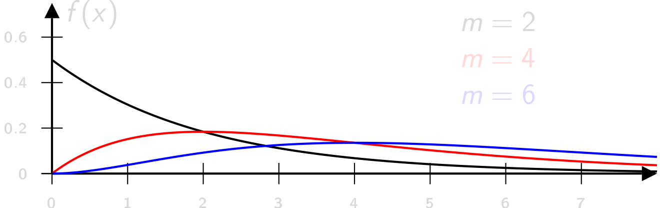
Important features:

- Is determined by its mean value and variance.
- If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent:

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

χ^2 distribution

- $X_1, X_2, \ldots, X_m$ are independent, standard normal distributed.
- The sum
$$X = \sum_{i=1}^{m} X_i^2 = X_1^2 + X_2^2 + \cdots + X_m^2$$
is χ^2 distributed with m degrees of freedom (notation: $X \sim \chi^2(m)$).
- Note:
  - X is positive.
  - $E(X) = m$.
  - Density function $f(x)$ for X:

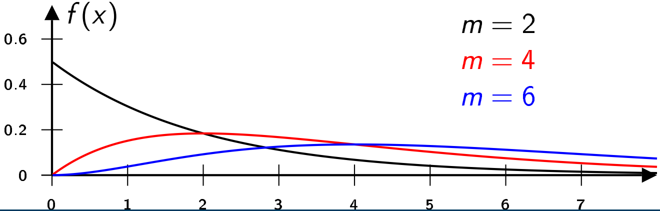


χ^2 distribution

- $X_1, X_2, \ldots, X_m$ are independent, standard normal distributed.
- The sum

$$X = \sum_{i=1}^{m} X_i^2 = X_1^2 + X_2^2 + \cdots + X_m^2$$

  is χ^2 distributed with m degrees of freedom (notation: $X \sim \chi^2(m)$).
- Note:
  - X is positive.
  - $E(X) = m$.
  - Density function $f(x)$ for X:



Central limit theorem

Theorem (Central limit theorem)

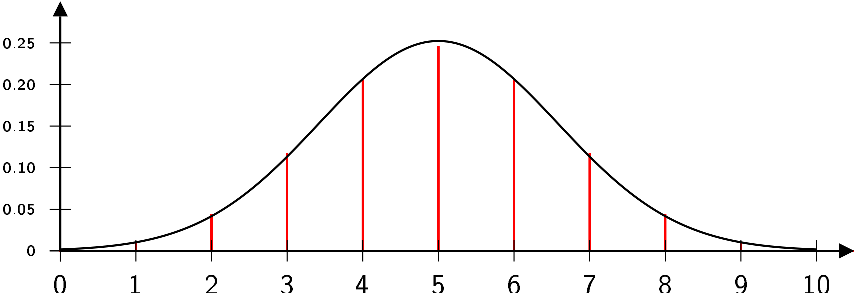
The sum of many independent, identically distributed (iid) variables is approximately normal distributed.

Note:

- The result is true no matter what distribution the terms in the sum follow.
- If each term is normal distributed, we can skip “approximately”.
- “many” can be as few as 10.

Central limit theorem – example

If n is big, $B(n,p)$ is approx. $N(np, np(1-p))$.



Online examples:
http://bcs.whfreeman.com/ips4e/cat_010/applets/CLT-Binomial.html
http://www.stat.sc.edu/~west/javahtml/CLT.html