

LIST OF FORMULAS FOR STK1100 AND STK1110

(Version of 11. November 2015)

1. Probability

Let $A, B, A_1, A_2, \dots, B_1, B_2, \dots$ be events, that is, subsets of a sample space Ω .

a) Axioms:

A probability function P is a function from subsets of the sample space Ω to real numbers, satisfying

$$\begin{aligned}P(\Omega) &= 1 \\P(A) &\geq 0 \\P(A_1 \cup A_2) &= P(A_1) + P(A_2) \quad \text{if } A_1 \cap A_2 = \emptyset \\P\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} P(A_i) \quad \text{if } A_i \cap A_j = \emptyset \text{ for } i \neq j\end{aligned}$$

b) $P(A') = 1 - P(A)$

c) $P(\emptyset) = 0$

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

e) The addition law of probability/ the sum rule:

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

f) Conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{if } P(B) > 0$$

g) Total probability:

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i) \quad \text{if } \bigcup_{i=1}^n B_i = \Omega \text{ and } B_i \cap B_j = \emptyset \text{ for } i \neq j$$

h) Bayes' Rule:

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)} \quad \text{under same conditions as in g)}$$

i) A and B are (statistically) independent events if $P(A \cap B) = P(A)P(B)$

j) A_1, \dots, A_n are (statistically) independent events if

$$P(A_{i_1} \cap \dots \cap A_{i_m}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_m})$$

for any subset of indexes i_1, i_2, \dots, i_m

k) The product rule:

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) \\ = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}) \end{aligned}$$

2. Combinatorics

a) Two operations that can be done in respectively n and m different ways can be combined in $n \cdot m$ ways.

b) The number of ordered subsets of r elements drawn with replacement from a set of n elements is n^r

c) The number of ordered subsets of r elements drawn without replacement from a set of n elements is $n(n-1) \cdots (n-r+1)$

d) Number of permutations of n elements is $n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$

e) The number of unordered subsets of r elements drawn from a set of n elements is

$$\binom{n}{r} = \frac{n(n-1) \cdots (n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$

f) Number of ways a set of n elements can be divided into r subsets with n_i elements in the i th subset is

$$\binom{n}{n_1 \ n_2 \ \cdots \ n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

3. Probability distributions

- a) For a random variable X (discrete or continuous), $F(x) = P(X \leq x)$ is the cumulative distribution function (cdf).
- b) For a discrete random variable X which can take the values x_1, x_2, x_3, \dots we have

$$p(x_j) = P(X = x_j)$$
$$F(x) = \sum_{x_j \leq x} p(x_j)$$

$p(x_j)$ is a point probability if

$$p(x_j) \geq 0 \quad \text{for all } j$$
$$\sum_j p(x_j) = 1$$

- c) For a continuous random variable X we have

$$P(a < X < b) = \int_a^b f(x) dx$$
$$F(x) = \int_{-\infty}^x f(u) du$$
$$f(x) = F'(x)$$

$f(x)$ is a probability density function if

$$f(x) \geq 0$$
$$\int_{-\infty}^{\infty} f(x) dx = 1$$

- d) For two random variables X and Y (discrete or continuous) the joint cumulative distribution function is $F(x, y) = P(X \leq x, Y \leq y)$
- e) For discrete random variables X and Y which can take the values x_1, x_2, \dots and y_1, y_2, \dots respectively, we have

$$p(x_i, y_j) = P(X = x_i, Y = y_j)$$
$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j)$$

$p(x_i, y_j)$ is a joint point probability if it fullfills the same conditions as in b)

f) For continuous random variables X and Y we have

$$P((X, Y) \in A) = \int \int_A f(u, v) dv du$$

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du$$

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

$f(x, y)$ is a joint probability density function if it fullfills the same conditions as in c)

g) Marginal point probabilities:

$$p_X(x_i) = \sum_j p(x_i, y_j) \quad (\text{for } X)$$

$$p_Y(y_j) = \sum_i p(x_i, y_j) \quad (\text{for } Y)$$

h) Marginal probability densities:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (\text{for } X)$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad (\text{for } Y)$$

i) Independence:

The random variables X and Y are independent if

$$p(x_i, y_j) = p_X(x_i)p_Y(y_j) \quad (\text{discrete})$$

$$f(x, y) = f_X(x)f_Y(y) \quad (\text{continuous})$$

j) Conditional point probabilities:

$$p_{X|Y}(x_i|y_j) = \frac{p(x_i, y_j)}{p_Y(y_j)} \quad (\text{for } X \text{ given } Y = y_j)$$

$$p_{Y|X}(y_j|x_i) = \frac{p(x_i, y_j)}{p_X(x_i)} \quad (\text{for } Y \text{ given } X = x_i)$$

Assuming $p_Y(y_j) > 0$ and $p_X(x_i) > 0$, respectively. Conditional point probabilities can be treated as regular point probabilities.

k) Conditional probability densities:

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} \quad (\text{for } X \text{ given } Y = y)$$

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} \quad (\text{for } Y \text{ given } X = x)$$

Assuming $f_Y(y) > 0$ and $f_X(x) > 0$, respectively. Conditional probability densities can be treated as regular probability densities.

4. Expectation

a) The expected value of a random variable X is defined as

$$E(X) = \sum_j x_j p(x_j) \quad (\text{discrete})$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad (\text{continuous})$$

b) For a real function $g(X)$ of a random variable X , the expected value is

$$E[g(X)] = \sum_j g(x_j) p(x_j) \quad (\text{discrete})$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx \quad (\text{continuous})$$

c) $E(a + bX) = a + bE(X)$

d) For a real function $g(X, Y)$ of two random variables X and Y , the expected value is

$$E[g(X, Y)] = \sum_i \sum_j g(x_i, y_j) p(x_i, y_j) \quad (\text{discrete})$$

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dy dx \quad (\text{continuous})$$

e) If X and Y are independent $E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$

f) If X and Y are independent $E(XY) = E(X) \cdot E(Y)$

g) $E\left(a + \sum_{i=1}^n b_i X_i\right) = a + \sum_{i=1}^n b_i E(X_i)$

h) Conditional expectation:

$$E(Y|X = x_i) = \sum_j y_j p_{Y|X}(y_j|x_i) \quad (\text{discrete})$$

$$E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy \quad (\text{continuous})$$

5. Variance and standard deviation

a) The variance and standard deviation of a random variable X are defined as

$$V(X) = E[(X - \mu)^2]$$
$$\text{sd}(X) = \sqrt{V(X)}$$

b) $V(X) = E(X^2) - (E(X))^2$

c) $V(a + bX) = b^2 V(X)$

d) If X_1, \dots, X_n are independent we have

$$V\left(a + \sum_{i=1}^n b_i X_i\right) = \sum_{i=1}^n b_i^2 V(X_i)$$

e)

$$V\left(a + \sum_{i=1}^n b_i X_i\right) = \sum_{i=1}^n b_i^2 V(X_i) + \sum_{i=1}^n \sum_{j \neq i} b_i b_j \text{Cov}(X_i, X_j)$$

f) Chebyshev's inequality:

Let X be a random variable with $\mu = E(X)$ and $\sigma^2 = V(X)$.

For all $t > 0$ we have

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$$

6. Covariance and correlation

a) Let X and Y be random variables with $\mu_X = E(X)$, $\sigma_X^2 = V(X)$, $\mu_Y = E(Y)$ and $\sigma_Y^2 = V(Y)$. The covariance and correlation of X and Y is then defined as

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$
$$\rho = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- b) $\text{Cov}(X, X) = V(X)$
- c) $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$
- d) X, Y independent $\Rightarrow \text{Cov}(X, Y) = 0$
- e)

$$\text{Cov} \left(a + \sum_{i=1}^n b_i X_i, c + \sum_{j=1}^m d_j Y_j \right) = \sum_{i=1}^n \sum_{j=1}^m b_i d_j \text{Cov}(X_i, Y_j)$$

- f) $-1 \leq \text{Corr}(X, Y) \leq 1$ and $\text{Corr}(X, Y) = \pm 1$ if and only if there exists two numbers a, b such that $Y = a + bX$ (except, eventually, on areas of zero probability)

7. Moment generating functions

- a) For a random variable X (discrete or continuous) the moment generating function is $M_X(t) = E(e^{tX})$
- b) If the moment generating function $M_X(t)$ exists for t in an open interval containing 0, then it uniquely determines the distribution of X .
- c) If the moment generating function $M_X(t)$ exists for t in an open interval containing 0, then all moments of X exist, and we can find the r th moment by $E(X^r) = M_X^{(r)}(0)$
- d) $M_{a+bX}(t) = e^{at} M_X(bt)$
- e) If X and Y are independent: $M_{X+Y}(t) = M_X(t)M_Y(t)$

8. Some discrete probability distributions

- a) Binomial distribution:

Point probability: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n$

Moment generating function: $M_X(t) = (1-p + pe^t)^n$

Expectation: $E(X) = np$

Variance : $V(X) = np(1-p)$

Approximation 1: $Z = \frac{X - np}{\sqrt{np(1-p)}}$ is approximately normally distributed

when np and $n(1-p)$ both are sufficiently big (at least 10)

Approximation 2: X is approximately Poisson distributed with parameter $\lambda = np$ when n is big and p is small

Sum rule: $X \sim \text{binomial}(n, p)$, $Y \sim \text{binomial}(m, p)$
and X, Y independent $\Rightarrow X + Y \sim \text{binomial}(n + m, p)$

b) Geometric distribution:

Point probability: $P(X = k) = (1 - p)^{k-1}p \quad k = 1, 2, \dots$

Moment generating function: $M_X(t) = e^t p / [1 - (1 - p)e^t]$

Expectation: $E(X) = 1/p$

Variance: $V(x) = (1 - p)/p^2$

Sum rule: If X is geometrically distributed with probability p then
 $X - 1$ is negative binomial $(1, p)$. Then if X and Y are
geometrically distributed with same p and independent then
 $X + Y - 2$ is negative binomial $(2, p)$

c) Negative binomial distribution:

Point probability: $P(X = k) = \binom{k+r-1}{r-1} p^r (1 - p)^k \quad k = 0, 1, 2, \dots$

Moment generating function: $M_X(t) = \{p/[1 - (1 - p)e^t]\}^r$

Expectation: $E(X) = r(1 - p)/p$

Variance: $V(X) = r(1 - p)/p^2$

Sum rule: $X \sim \text{negative binomial}(r_1, p)$,
 $Y \sim \text{negative binomial}(r_2, p)$
and X, Y independent
 $\Rightarrow X + Y \sim \text{negative binomial}(r_1 + r_2, p)$

d) Hypergeometric distribution:

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

Expectation: $E(X) = n \cdot \frac{M}{N}$

Variance: $V(X) = n \frac{M}{N} (1 - \frac{M}{N}) \frac{N-n}{N-1}$

Approximation: X is approximately binomial $(n, \frac{M}{N})$
when n is much smaller than N

e) Poisson distribution:

Point probability: $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, \dots$

Moment generating function: $M_X(t) = e^{\lambda(e^t - 1)}$

Expectation: $E(X) = \lambda$

Variance: $V(X) = \lambda$

Approximation: $Z = \frac{X - \lambda}{\sqrt{\lambda}}$ is approximately normally distributed
when λ is sufficiently big (at least 10)

Sum rule: $X \sim \text{Poisson}(\lambda_1), Y \sim \text{Poisson}(\lambda_2)$
and X, Y independent $\Rightarrow X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$

e) Multinomial distribution:

Point probability: $P(N_1 = n_1, \dots, N_r = n_r) = \frac{n!}{n_1! \dots n_r!} p_1^{n_1} \dots p_r^{n_r}$

Here $\sum_{i=1}^r p_i = 1$ and $\sum_{i=1}^r n_i = n$

Marginal distribution: $N_i \sim \text{binomial}(n, p_i)$

9. Some continuous probability distributions

a) Normal distribution:

Density: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty$

Moment generating function: $M_X(t) = e^{\mu t} e^{\sigma^2 t^2/2}$

Expectation: $E(X) = \mu$

Variance: $V(X) = \sigma^2$

Transformation: $X \sim N(\mu, \sigma^2) \Rightarrow a + bX \sim N(a + b\mu, b^2\sigma^2)$
 $X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$

Sum rule: $X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2),$
 X, Y independent
 $\Rightarrow X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

b) Exponential distribution:

Density: $f(x) = \lambda e^{-\lambda x} \quad x > 0$

Moment generating function: $M_X(t) = \lambda/(\lambda - t)$ for $t < \lambda$

Expectation: $E(X) = 1/\lambda$

Variance: $V(X) = 1/\lambda^2$

Sum rule: $X \sim \exp(\lambda), Y \sim \exp(\lambda), X$ and Y independent
 $\Rightarrow X + Y \sim \text{gamma}(2, 1/\lambda)$

c) Gamma distribution:

Density: $f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad x > 0$

Gamma function: $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$
 $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
 $\Gamma(n) = (n - 1)!$ when n is an integer
 $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$

Moment generating function: $M_X(t) = [1/(1 - \beta t)]^\alpha$

Expectation: $E(X) = \alpha\beta$

Variance: $V(X) = \alpha\beta^2$

Sum rule: $X \sim \text{gamma}(\alpha, \beta)$, $Y \sim \text{gamma}(\delta, \beta)$,
 X and Y independent $\Rightarrow X + Y \sim \text{gamma}(\alpha + \delta, \beta)$

d) Chi-squared distribution:

Density: $f(v) = \frac{1}{2^{n/2}\Gamma(n/2)} v^{(n/2)-1} e^{-v/2}$ $v > 0$
 n degrees of freedom

Expectation: $E(V) = n$

Variance: $V(V) = 2n$

Sum rule: $U \sim \chi_n^2$, $V \sim \chi_m^2$, U and V independent
 $\Rightarrow U + V \sim \chi_{n+m}^2$

Result: $Z \sim N(0, 1) \Rightarrow Z^2 \sim \chi_1^2$

e) Student's t -distribution:

Density: $f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)} (1 + \frac{t^2}{n})^{-(n+1)/2}$ $-\infty < t < \infty$
 n degrees of freedom

Expectation: $E(T) = 0$ ($n \geq 2$)

Variance: $V(T) = n/(n - 2)$ ($n \geq 3$)

Result: $Z \sim N(0, 1)$, $U \sim \chi_n^2$, Z, U independent $\Rightarrow Z/\sqrt{U/n} \sim t_n$

f) Binormal distribution:

Density:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right\}$$

Marginal distribution: $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$

Correlation: $\text{Corr}(X, Y) = \rho$

Conditional distribution: Given $X = x$, Y is normally distributed with
expectation $E(Y|X = x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$
and variance $V(Y|X = x) = \sigma_Y^2 (1 - \rho^2)$

10. One normally distributed sample

If X_1, X_2, \dots, X_n are independent and $N(\mu, \sigma^2)$ distributed then we have that:

- a) $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ are independent
- b) $\bar{X} \sim N(\mu, \sigma^2/n)$
- c) $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$
- d) $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

11. Two normally distributed samples

Let X_1, X_2, \dots, X_n be independent and $N(\mu_X, \sigma^2)$ distributed, and Y_1, Y_2, \dots, Y_m independent and $N(\mu_Y, \sigma^2)$ distributed. The two samples are independent of each other. Let \bar{X}, \bar{Y}, S_X^2 and S_Y^2 be defined as in 10a). Then we have that:

- a) $S_p^2 = [(n-1)S_X^2 + (m-1)S_Y^2]/(m+n-2)$ is a weighted estimator for σ^2
- b) $\bar{X} - \bar{Y} \sim N(\mu_X - \mu_Y, \sigma^2(\frac{1}{n} + \frac{1}{m}))$
- c) $(n+m-2)S_p^2/\sigma^2 \sim \chi_{m+n-2}^2$
- d) $\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{m+n-2}$

12. Regression analysis

Assume $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$; $i = 1, 2, \dots, n$; where x_{ij} -s are given numbers and ϵ_i -s are independent and $N(0, \sigma^2)$ distributed. Then we have that:

- a) The least squares estimators for β_0 and β_1 are

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} \quad \text{and} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- b) The estimators in a) are normally distributed and unbiased, and

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{and} \quad \text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- c) Let $\text{SSE} = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$. Then $S^2 = \text{SSE}/(n-2)$ is an unbiased estimator for σ^2 , and $(n-2)S^2/\sigma^2 \sim \chi_{n-2}^2$

13. Multiple linear regression

Assume $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$; $i = 1, 2, \dots, n$; where x_{ij} -s are given numbers and ϵ_i -s are independent and $N(0, \sigma^2)$ distributed. The model can be written in matrix form as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$ and $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^T$ are n -, n - and $(k+1)$ -dimensional vectors, and $\mathbf{X} = \{x_{ij}\}$ (with $x_{i0} = 1$) is a $n \times (k+1)$ -dimensional matrix. Then:

1. The least squares estimator for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$.
2. Let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \dots, \hat{\beta}_k)^T$. Then $\hat{\beta}_j$ -s are normally distributed and unbiased, and

$$\text{Var}(\hat{\beta}_j) = \sigma^2 c_{jj} \quad \text{og} \quad \text{Cov}(\hat{\beta}_j, \hat{\beta}_l) = \sigma^2 c_{jl}$$

where c_{jl} is element (j, l) in the $(k+1) \times (k+1)$ matrix $\mathbf{C} = (\mathbf{X}^T \mathbf{X})^{-1}$.

3. Let $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$, og let $SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$. Then $S^2 = SSE/[n - (k+1)]$ is an unbiased estimator for σ^2 , and $[n - (k+1)]S^2/\sigma^2 \sim \chi_{n-(k+1)}^2$. Also, S^2 and $\hat{\boldsymbol{\beta}}$ are independent.
4. Let $S_{\hat{\beta}_j}^2$ be the variance estimator for $\hat{\beta}_j$ we get by replacing σ^2 with S^2 in the formula for $\text{Var}(\hat{\beta}_j)$ (in b). Then $(\hat{\beta}_j - \beta_j)/S_{\hat{\beta}_j} \sim t_{n-(k+1)}$.