TAMS65 -VT1: Notations and Formulas

— by Zhenxia Liu

1 Repetition of Probability Theory

1.1 Basic notations and formulas

X: random variable (stokastiska variabel);

Discrete random variable(diskret stokastisk variabel):
pmf = Probability mass function(sannolikhetsfunktion), $p_X(k) = p(k) := P(X = k)$.

Continuous random variable(kontinuerlig stokastisk variabel):
pdf = Probability density function(täthesfunktion), $f_X(x) = f(x)$.

cdf = Cumulative distribution function(fördelningsfunktion):

$$F(x) := P(X \leq x) = \begin{cases} \sum_{k \leq x} p_X(k) & \text{discrete r.v.} \\ \int_{-\infty}^{x} f_X(t)dt & \text{continuous r.v.} \end{cases}$$

Expectation/mean/expected value(väntevärde)

$$\mu = E(X) = \begin{cases} \sum k p_X(k), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f_X(x)dx, & \text{if } X \text{ is continuous;} \end{cases}$$

If $Y = g(X)$, then

$$E(Y) = \begin{cases} \sum_k g(k) p_X(k), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} g(x) f_X(x)dx, & \text{if } X \text{ is continuous;} \end{cases}$$

Variance(Varians): $\sigma^2 = V(X) = var(X) = E((X-\mu)^2) = E(X^2) - (E(X))^2$;

Standard deviation(Standardavvikelse): $\sigma = D(X) = \sqrt{V(X)}$;

If $X_1, ..., X_n$ are r.v.s and $c_0, c_1, ..., c_n$ are constants, then

$$E\left(c_0 + c_1 X_1 + \ldots + c_n X_n\right) = c_0 + c_1 E\left(X_1\right) + \ldots + c_n E\left(X_n\right)$$

If $X_1, ..., X_n$ are **independent(oberoende)**, then

$$V\left(c_0 + c_1 X_1 + \ldots + c_n X_n\right) = \sum_{i=1}^{n} c_i^2 V\left(X_i\right)$$

If X and Y are r.v.s and a, b, c are constants, then

$$V\left(aX + bY + c\right) = a^2 V(X) + 2ab \cdot cov(X, Y) + b^2 V(Y)$$

1.2 Several discrete r.v.

Binomial distribution(Binomialfördelning) $X \sim Bin(n, p)$

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for } k = 0, 1, \ldots, n$$

$E(X) = np \quad V(X) = np(1-p)$

Poisson distribution(Poissonfördelning) $X \sim Po(\lambda)$

$$p_X(k) = \frac{\mu^k}{k!} e^{-\mu}, \quad k = 0, 1, 2, \ldots.$$

$E(X) = \mu \quad V(X) = \mu$

Hypergeometric distribution(Hypergeometrisk fördelning) $X \sim Hyp(N, n, p)$

$$p_X(x) = \frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}} \quad \text{for} \quad 0 \leq x \leq Np \quad \text{and} \quad 0 \leq n - x \leq N(1-p)$$

$E(X) = np \quad V(X) = \frac{N-n}{N-1} np(1-p)$

1.3 Several continuous r.v.

Normal distribution(normalfördelning): $X \sim N(\mu, \sigma)$

$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Standard normal distribution $Z \sim N(0, 1)$ $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$.

If $X \sim N(\mu, \sigma)$, then $Y = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$.

If $X_1, ..., X_n$ are **independent(oberoende)** and each $X_i \sim N(\mu_i, \sigma_i)$, for any constants $d, c_1, ..., c_n$, then we have

$$d + \sum_{i=1}^{n} c_i X_i \sim N\left(d + \sum_{i=1}^{n} c_i \mu_i, \sqrt{\sum_{i=1}^{n} c_i^2 \sigma_i^2}\right)$$

Exponential distribution(Exponentialfördelning) $X \sim Exp(\frac{1}{\mu})$

$$f_X(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \text{ for } x > 0$$

$E(X) = \mu \quad V(X) = \mu^2$

Uniform distribution(Likformigfördelning) $X \sim U(a,b)$ or $X \sim Re(a,b)$

$$f_X(x) = \frac{1}{b-a} \text{ for } a < x < b$$

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

1.4 Central Limit Theorem (CLT)(Centrala gränsvärdessatsen)

Let $\{X_1, X_2, \dots, X_n\}$ be a sequence of independent and identically distributed random variables with expectation $E(X) = \mu$ and variance $V(X) = \sigma^2$. Then for large $n \geq 30$,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1). \tag{1}$$

• $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$.

• If the population is normal, then CLT holds for any n.

• Understand that $\mu = E(\bar{X})$ and $(\sigma/\sqrt{n})^2 = V(\bar{X})$.

2 Statistics Theory

2.1 Basics in Statistics

Population X;

Random sample (slumpmässigt stickprov): $X_1, \dots, X_n$ are independent and have the same distribution as the population X. Before observe/measure, $X_1, \dots, X_n$ are random variables, and after observe/measure, we use $x_1, \dots, x_n$ which are numbers (not random variables);

Observations(observationer): $x_1, \dots, x_n$.

Sample mean (stickprovsmedelvärde): Before observe/measure, $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, and after observe/measure, $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$;

Sample variance (Stickprovsvarians): Before observe/measure, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$, and after observe/measure, $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$;

Sample standard deviation (Stickprovsstandardavvikelse): Before observe/measure, $S = \sqrt{S^2}$, and after observe/measure, $s = \sqrt{s^2}$;

2.2 Point estimation

For a population X with an unknown parameter θ, and a random sample $\{X_1, \dots, X_n\}$:

Point Estimator(stickprovsvariabeln): $\hat{\Theta} = f(X_1, \dots, X_n)$, a random variable;

Point Estimate (punktskattning): $\hat{\theta} = f(x_1, \dots, x_n)$, a number;

Unbiased(Väntevärdesriktig): $E(\hat{\Theta}) = \theta$;

Effective (Effektiv): Two estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased, we say that $\hat{\Theta}_1$ is more effective than $\hat{\Theta}_2$ if $V(\hat{\Theta}_1) < V(\hat{\Theta}_2)$;

Consistent (Konsistent): A point estimator $\hat{\Theta} = g(X_1, \dots, X_n)$ is consistent if

$$\lim_{n\to\infty} P(|\hat{\Theta} - \theta| > \varepsilon) = 0, \text{ for any constant } \varepsilon > 0.$$

(This is called "*convergence in probability*").

Theorem: If $E(\hat{\Theta}) = \theta$ and $\lim_{n\to\infty} V(\hat{\Theta}) = 0$, then $\hat{\Theta}$ is consistent.

Method of moments (Momentmetoden)-MM: # of equations depends on # of unknown parameters,

$$E(X) = \bar{x}, \quad E(X^2) = \frac{1}{n}\sum_{i=1}^n x_i^2, \quad E(X^3) = \frac{1}{n}\sum_{i=1}^n x_i^3, \quad \dots\dots$$

Least square method (minsta-kvadrat-metoden)-LSM: The least square estimate $\hat{\theta}$ is the one minimizing

$$Q(\theta) = \sum_{i=1}^n (x_i - E(X))^2.$$

Maximum-likelihood method (Maximum-likelihood-metoden)-ML: The maximum-likelihood point estimate $\hat{\theta}$ is the one maximizing the likelihood function

$$L(\theta) = \begin{cases} \prod_{i=1}^n f(x_i;\theta), & \text{if } X \text{ is continuous,} \\ \prod_{i=1}^n p(x_i;\theta), & \text{if } X \text{ is discrete.} \end{cases}$$

Remark 1 on ML: In general, it is easier/better to maximize $\ln L(\theta)$;

Remark 2 on ML: If there are several random samples (say m) from independent populations with a same unknown parameter θ, then the maximum-likelihood estimate $\hat{\theta}$ is the one maximizing the likelihood function defined as $L(\theta) = L_1(\theta) \dots L_m(\theta)$, where $L_i(\theta)$ is the likelihood function from the i-th sample.

Estimates of population mean μ: point estimator $\hat{M} = \bar{X}$ and point estimate $\hat{\mu} = \bar{x}$.

Estimates of population variance σ^2:

• If there is only one random sample,

If μ is known(känt), point estimator $\hat{\Sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$ and point estimate $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$.

If μ is unknown(okänt), point estimator $\hat{\Sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ and point estimate $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$ Sample variance.

• If there are m samples from independent populations with unknown means and a same variance σ^2, then

$$\hat{\sigma}^2 = s^2 = \frac{(n_1-1)s_1^2 + \ldots + (n_m-1)s_m^2}{(n_1-1)+\ldots+(n_m-1)} \qquad \text{(unbiased)}$$

where n_i is the sample size of the i-th sample, and s_i^2 is the sample variance of the i-th sample.

Note that: MM and ML give a point estimate of σ^2 as follows

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{(NOT unbiased).}$$

An **adjusted/corrected(korrigerade)** point estimate would be the sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{(unbiased).}$$

Sample standard deviation $s = \sqrt{s^2}$ and $S = \sqrt{S^2}$.

Standard error(medelfelet) of a point estimate $\hat{\theta}$: $d(\hat{\theta})$ is an estimation of the standard deviation $D(\hat{\Theta})$.

2.3 Interval estimation - Confidence interval(Konfidensintervall) -CI

2.3.1 One random sample $\{X_1, \ldots, X_n\}$ from $N(\mu, \sigma)$

CI for μ

$$\text{the sampling distribution is } \begin{cases} \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1), \text{ if } \sigma \text{ is known} \\ \\ \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1), \text{ if } \sigma \text{ is unknown} \end{cases}$$

CI for σ^2, the sampling distribution is $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

Note: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

2.3.2 Two random samples $\{X_1, \ldots, X_{n_1}\}$ and $\{Y_1, \ldots, Y_{n_2}\}$ from independent distributions $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$, respectively.

$$\text{CI for } \mu_1 - \mu_2, \text{ the sampling distribution is } \begin{cases} \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1), \text{ if } \sigma_1 \text{ and } \sigma_2 \text{ are known;} \\ \\ \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2), \text{ if } \sigma_1 = \sigma_2 = \sigma \text{ is unknown;} \\ \\ \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \approx t(f), \text{ if } \sigma_1 \neq \sigma_2 \text{ both are unknown;} \\ \qquad \left[\text{degrees of freedom } f = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}\right] \end{cases}$$

CI for σ^2 : If $\sigma_1 = \sigma_2 = \sigma$, the distribution function is $\frac{(n_1+n_2-2)S^2}{\sigma^2} \sim \chi^2(n_1+n_2-2)$.

Note that: Unknown σ^2 can be estimated by the samples variance $s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$.

Remark: The idea of using **sampling distribution** to find confidence intervals is very important. There are a lot more different confidence intervals besides above. For instance, we consider two independent samples: $\{X_1, \ldots, X_{n_1}\}$ from $N(\mu_1, \sigma_1)$ and $\{Y_1, \ldots, Y_{n_2}\}$ from $N(\mu_2, \sigma_2)$. In this case, we can easily prove that

$$c_1\bar{X} + c_2\bar{Y} \sim N\left(c_1\mu_1 + c_2\mu_2, \quad \sqrt{\frac{c_1^2\sigma_1^2}{n_1} + \frac{c_2^2\sigma_2^2}{n_2}}\right).$$

Then CI for $c_1\mu_1 + c_2\mu_2$, the following sampling distribution is

- If σ_1 and σ_2 are known, $\frac{(c_1\bar{X}+c_2\bar{Y})-(c_1\mu_1+c_2\mu_2)}{\sqrt{\frac{c_1^2\sigma_1^2}{n_1}+\frac{c_2^2\sigma_2^2}{n_2}}} \sim N(0,1)$.

- If $\sigma_1 = \sigma_2 = \sigma$ is unknown, $\frac{(c_1\bar{X}+c_2\bar{Y})-(c_1\mu_1+c_2\mu_2)}{S\sqrt{\frac{c_1^2}{n_1}+\frac{c_2^2}{n_2}}} \sim t(n_1+n_2-2)$.

2.3.3 Confidence intervals from More random samples from independent $N(\mu_i, \sigma_i)$, $i = 1, \ldots, n$.

Assume that θ is a linear combination of μ_i.

CI for θ, the sampling distribution is

• If σ_i is known,

$$\frac{\widehat{\Theta} - \theta}{D(\widehat{\Theta})} \sim N(0,1)$$

• If $\sigma_1 = \ldots = \sigma_n = \sigma$ is unknown,

$$\frac{\widehat{\Theta} - \theta}{\widehat{D}(\widehat{\Theta})} \sim t(f), \text{ where } \widehat{D} = S \cdot constant$$

CI for σ^2, the sampling distribution is

$$\frac{fS^2}{\sigma^2} \sim \chi^2(f)$$

Note: f = degrees of freedom for S^2.

2.3.4 Confidence intervals from normal approximations.

$$X \sim Bin(n,p): \text{ Sampling distribution} \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1) \text{ for } n\hat{p}(1-\hat{p}) > 10.$$

$$X \sim Po(\mu): \text{ Sampling distribution} \frac{\bar{X} - \mu}{\sqrt{\frac{\mu}{n}}} \approx N(0,1) \text{ for } n\hat{\mu} > 15.$$

$$X \sim Hyp(N,n,p): \text{ Sampling distribution} \frac{\widehat{P} - p}{\sqrt{\frac{N-n}{N-1}\frac{p(1-p)}{n}}} \approx N(0,1) \text{ for } \frac{n}{N} \le 0.1 \text{ and } n\hat{p}(1-\hat{p}) \ge 10.$$

$$X \sim Exp(\frac{1}{\mu}): \text{ Sampling distribution } \frac{\bar{X} - \mu}{\mu/\sqrt{n}} \approx N(0,1) \text{ for } n \ge 30$$

Remark: Again there are more confidence intervals besides above. For instance, we consider two independent samples: X from $Bin(n_1, p_1)$ and Y from $Bin(n_2, p_2)$, with unknown p_1 and p_2. As we know

$$\hat{P}_1 \approx N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right) \text{ and } \hat{P}_2 \approx N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}}\right),$$

Therefore, to get CI for $p_1 - p_2$, we consider this sampling distribution $\frac{(\hat{P}_1-\hat{P}_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}} \approx N(0,1)$ for $n_1\hat{p}_1(1-\hat{p}_1) > 10$ and $n_2\hat{p}_2(1-\hat{p}_2) > 10$.

2.3.5 Confidence intervals from the ratio of two population variances σ_2^2/σ_1^2.

Suppose there are two samples $\{X_1, \ldots, X_{n_1}\}$ and $\{Y_1, \ldots, Y_{n_2}\}$ from independent $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$, respectively. Then $\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1)$ and $\frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$, the sampling distribution is

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1).$$

2.3.6 Large sample size ($n \ge 30$, population may be completely unknown).

If there is no information about the population(s), then we can apply Central Limit Theorem (usually with a large sample $n \ge 30$) to get an approximated normal distributions. Here are two examples:

Example 1: Let $\{X_1, \ldots, X_n\}, n \ge 30$, be a random sample from an unknown population, then (no matter what distribution the population is)

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1).$$

Example 2: Let $\{X_1, \ldots, X_{n_1}\}, n_1 \ge 30$, be a random sample from an unknown population, and $\{Y_1, \ldots, Y_{n_2}\}, n_2 \ge 30$, be a random sample from another unknown population which is independent from the first population, then (no matter what distributions the populations are)

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx N(0,1).$$

3 Hypothesis testing(hypotesprövning) -HT

3.1 *One sample* and the *general theory* of hypothesis testing

Population X with an unknown parameter θ,

$$H_0: \theta = \theta_0 \quad vs. \quad H_1: \theta < \theta_0, \text{ or } \theta > \theta_0, \text{ or } \theta \ne \theta_0$$

HT-1 Population X is not Normal(approximation) distribution and has Only one observation x.

HT-2 All types of populations for Confidence interval.

	H_0 is true	H_0 is false and $\theta = \theta_1$
reject H_0	(type I error or significance level) α	(power) $h(\theta_1)$
don't reject H_0	$1 - \alpha$	(type II error) $\beta(\theta_1) = 1 - h(\theta_1)$

Find sampling distributions from **section 2.3 Interval estimation**.

TS := "test statistic"; and C := "rejection region/critical region".

$$TS \in C \Leftrightarrow \text{reject } H_0$$

$$p\text{-value} < \alpha \Leftrightarrow \text{reject } H_0$$

4 Basic χ^2-test

4.1 Test on distribution

$$\begin{cases} H_0: & X \sim \text{ distribution (with or without unknown parameters);} \\ H_1: & X \nsim \text{ distribution} \end{cases}$$

The sampling distribution is

$$\sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \sim \chi^2(k - 1 - \#\text{of unknown parameters})$$

for $\sum p_i = 1$ and $np_i > 5$.

4.2 Test of Independence / Homogeneity

Suppose we have a data with r rows and k columns,

$$\begin{cases} H_0: & \text{the grouping of } r \text{ rows and the grouping of } k \text{ columns are independent;} \\ H_1: & \text{the grouping of } r \text{ rows and the grouping of } k \text{ columns are not independent.} \end{cases}$$

Equivalently,

$$\begin{cases} H_0: & \text{the distributions of } r \text{ rows in each column are the same} \\ H_1: & \text{the distributions of } r \text{ rows in each column are Not the same} \end{cases}$$

Then the sampling distribution

$$\sum_{j=1}^{k} \sum_{i=1}^{r} \frac{(N_{ij} - np_{ij})^2}{np_{ij}} \sim \chi^2((r-1)(k-1))$$

for $np_{ij} > 5$, where $p_{ij} = p_i \cdot q_j$ are the theoretical probabilities.