

TAMS24: Notations and Formulas

— by Xiangfeng Yang

1 Basic notations and definitions

X: random variable (stokastiska variabel);

Mean (Väntevärde):

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

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

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

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);

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

$$E\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i E(X_i),$$

$$V\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 V(X_i), \text{ if } X_1, \dots, X_n \text{ are independent (oberoende);}$$

If $X \sim N(\mu, \sigma)$, then $\frac{X-\mu}{\sigma} \sim N(0, 1)$;

If X_1, \dots, X_n are independent and $X_i \sim N(\mu_i, \sigma_i)$, then

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

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

Estimator (Stickprovsvariabeln): $\hat{\Theta} = g(X_1, \dots, X_n)$, a random variable;

Estimate (Punktskattning): $\hat{\theta} = g(x_1, \dots, x_n)$, 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 $V(\hat{\Theta}_1) < V(\hat{\Theta}_2)$;

Binomial distribution $X \sim Bin(N, p)$: there are N independent and identical trials, each trial has a probability of success p , and X = the number of successes in these N trials. The random variable $X \sim Bin(N, p)$ has a probability function (sannolikhetsfunktion)

$$p(k) = P(X = k) = \binom{N}{k} p^k (1-p)^{N-k};$$

Exponential distribution $X \sim Exp(1/\mu)$: when we consider the waiting time/lifetime... The random variable $X \sim Exp(1/\mu)$ has a density function (täthetsfunktion)

$$f(x) = \frac{1}{\mu} e^{-x/\mu}, \quad x \geq 0.$$

2 Point estimation

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

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

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

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

(This is called “convergence in probability”).

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

Least square method (minsta-kvadrat-metoden): 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): The maximum-likelihood 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 different 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 population.

Estimates of population variance σ^2 : If there is only one population with an unknown mean, then method of moments and maximum-likelihood method, in general, give an estimate of σ^2 as follows

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

An adjusted (or corrected) estimate would be the sample variance

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

If there are m different populations with unknown means and a same variance σ^2 , then an adjusted (or corrected) ML estimate is

$$s^2 = \frac{(n_1 - 1)s_1^2 + \dots + (n_m - 1)s_m^2}{(n_1 - 1) + \dots + (n_m - 1)} \quad (\text{unbiased})$$

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

Standard error (medelfelet) of an estimator $\hat{\Theta}$: \sim is an estimate of the standard deviation $D(\hat{\Theta})$.

3 Interval estimation

<p>One sample $\{X_1, \dots, X_n\}$ from $N(\mu, \sigma)$</p>	$I_\mu = \begin{cases} \bar{x} \mp \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}, & \text{if } \sigma \text{ is known; } \left[\text{fact } \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \right] \\ \bar{x} \mp t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}}, & \text{if } \sigma \text{ is unknown; } \left[\text{fact } \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1) \right] \end{cases}$ $I_{\sigma^2} = \left(\frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)} \right); \left[\text{fact } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \right]$ <p>Unknown σ^2 can be estimated by the sample variance $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$</p>
<p>Two samples $\{X_1, \dots, X_{n_1}\}$ from $N(\mu_1, \sigma_1)$; $\{Y_1, \dots, Y_{n_2}\}$ from $N(\mu_2, \sigma_2)$; $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ are independent</p>	$I_{\mu_1 - \mu_2} = \begin{cases} (\bar{x} - \bar{y}) \mp \lambda_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, & \text{if } \sigma_1 \text{ and } \sigma_2 \text{ are known;} \\ \left[\text{fact } \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1) \right] \\ (\bar{x} - \bar{y}) \mp t_{\alpha/2}(n_1 + n_2 - 2) \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, & \text{if } \sigma_1 = \sigma_2 = \sigma \text{ is unknown;} \\ \left[\text{fact } \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2) \right] \\ \approx (\bar{x} - \bar{y}) \mp t_{\alpha/2}(f) \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, & \text{if } \sigma_1 \neq \sigma_2 \text{ both are unknown;} \\ \left[\text{fact } \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx t(f) \right] \\ \left[\text{degrees of freedom } f = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \right] \end{cases}$ $I_{\sigma^2} = \left(\frac{(n_1 + n_2 - 2)s^2}{\chi_{\frac{\alpha}{2}}^2(n_1 + n_2 - 2)}, \frac{(n_1 + n_2 - 2)s^2}{\chi_{1-\frac{\alpha}{2}}^2(n_1 + n_2 - 2)} \right), \text{ if } \sigma_1 = \sigma_2 = \sigma;$ $\left[\text{fact } \frac{(n_1 + n_2 - 2)S^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2) \right]$ <p>Unknown σ^2 can be estimated by the samples variance $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$</p>
<p>m samples:</p>	<p>The unknown $\sigma_1^2 = \dots = \sigma_m^2 = \sigma^2$ can be estimated by $s^2 = \frac{(n_1 - 1)s_1^2 + \dots + (n_m - 1)s_m^2}{(n_1 - 1) + \dots + (n_m - 1)}$.</p>

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

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

- If σ is known, then fact $\frac{(c_1 \bar{X} + c_2 \bar{Y}) - (c_1 \mu_1 + c_2 \mu_2)}{\sigma \sqrt{\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2}}} \sim N(0, 1)$. So we can find $I_{c_1 \mu_1 + c_2 \mu_2}$;
- If σ is unknown, then fact $\frac{(c_1 \bar{X} + c_2 \bar{Y}) - (c_1 \mu_1 + c_2 \mu_2)}{S \sqrt{\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2}}} \sim t(n_1 + n_2 - 2)$. So we can find $I_{c_1 \mu_1 + c_2 \mu_2}$.

3.1 Confidence intervals from normal approximations.

$$X \sim \text{Bin}(N, p) : I_p = \hat{p} \mp \lambda_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}, \text{ fact } \frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{N}}} \approx N(0, 1).$$

(we require that $N\hat{p} > 10$ and $N\hat{p}(1-\hat{p}) > 10$)

$$X \sim \text{Hyp}(N, n, p) : I_p = \hat{p} \mp \lambda_{\alpha/2} \sqrt{\frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \hat{p}(1-\hat{p})}, \text{ fact } \frac{\hat{P} - p}{\sqrt{\frac{N-n}{N-1} \cdot \frac{1}{n} \cdot \hat{P}(1-\hat{P})}} \approx N(0, 1).$$

$$X \sim \text{Po}(\mu) : I_\mu = \bar{x} \mp \lambda_{\alpha/2} \sqrt{\frac{\bar{x}}{n}}, \text{ fact } \frac{\bar{X} - \mu}{\sqrt{\frac{\bar{X}}{n}}} \approx N(0, 1).$$

(we require that $n\bar{x} > 15$)

$$X \sim \text{Exp}\left(\frac{1}{\mu}\right) : \bullet I_\mu = \left(\frac{\bar{x}}{1 + \frac{\lambda_{\alpha/2}}{\sqrt{n}}}, \frac{\bar{x}}{1 - \frac{\lambda_{\alpha/2}}{\sqrt{n}}} \right), \text{ fact } \frac{\bar{X} - \mu}{\mu/\sqrt{n}} \approx N(0, 1),$$

$$\bullet I_\mu = \bar{x} \mp \lambda_{\alpha/2} \frac{\bar{x}}{\sqrt{n}}, \text{ fact } \frac{\bar{X} - \mu}{\bar{X}/\sqrt{n}} \approx N(0, 1).$$

(we require that $n \geq 30$)

Remark: Again there are more confidence intervals besides above. For instance, we consider two independent samples: X from $\text{Bin}(N_1, p_1)$ and Y from $\text{Bin}(N_2, p_2)$, with unknown p_1 and p_2 . As we know

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

so $\hat{P}_1 - \hat{P}_2 \approx N \left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right)$. Therefore, fact is $\frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \approx N(0, 1)$,

$$I_{p_1 - p_2} = (\hat{p}_1 - \hat{p}_2) \mp \lambda_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

3.2 Confidence intervals from the ratio of two population variances.

$$SS_{TOT} = SS_R + SS_E, \text{ and } R^2 = \frac{SS_R}{SS_{TOT}}.$$

*** σ^2 is estimated as $\hat{\sigma}^2 = S^2 = \frac{SS_E}{n-k-1}$.

*** For the Hypothesis testing: $H_0 : \beta_1 = \dots = \beta_k = 0$ vs $H_1 : \text{at least one } \beta_j \neq 0$,

$$\begin{cases} \frac{SS_R/k}{SS_E/(n-k-1)} \sim F(k, n-k-1) \\ \text{TS} = \frac{SS_R/k}{SS_E/(n-k-1)} \\ C = (F_\alpha(k, n-k-1), +\infty). \end{cases}$$

*** We know $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, thus if we denote

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} h_{00} & h_{01} & \dots & h_{0k} \\ h_{10} & h_{11} & \dots & h_{1k} \\ \vdots & \vdots & & \vdots \\ h_{k1} & h_{k2} & \dots & h_{kk} \end{pmatrix},$$

then $\hat{\beta}_j \sim N(\beta_j, \sigma\sqrt{h_{jj}})$ and $\frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{h_{jj}}} \sim N(0, 1)$. But σ is generally unknown, therefore

$$\frac{\hat{\beta}_j - \beta_j}{S\sqrt{h_{jj}}} \sim t(n-k-1), \quad [s\sqrt{h_{jj}} \text{ is sometimes denoted as } d(\hat{\beta}_j) \text{ or } se(\hat{\beta}_j)].$$

Confidence interval of β_j is: $I_{\beta_j} = \hat{\beta}_j \mp t_{\alpha/2}(n-k-1) \cdot s\sqrt{h_{jj}}$;

Hypothesis testing $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$ has

$$\begin{cases} \text{TS} = \frac{\hat{\beta}_j}{s\sqrt{h_{jj}}} \\ C = (-\infty, -t_{\alpha/2}(n-k-1)) \cup (t_{\alpha/2}(n-k-1), +\infty). \end{cases}$$

Rewrite simple and multiple linear regressions as follows:

$$\begin{aligned} Y &= \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon, \quad \varepsilon \sim N(0, \sigma), \quad (\text{the model}); \\ \mu &= E(Y) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k, \quad (\text{the mean}); \\ \hat{\mu} &= \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k, \quad (\text{the estimated line}). \end{aligned}$$

For a given/fixed $\mathbf{x} = (1, x_1, \dots, x_k)'$, the scalar $\hat{\mu}$ is an estimate of unknown μ (and Y). Then we can talk about 'accuracy' of this estimate in terms of confidence intervals (and prediction intervals).

Confidence interval of μ : $I_\mu = \hat{\mu} \mp t_{\alpha/2}(n-k-1) \cdot s \cdot \sqrt{\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$.

Prediction interval of Y : $I_Y = \hat{\mu} \mp t_{\alpha/2}(n-k-1) \cdot s \cdot \sqrt{1 + \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$.

Suppose we have two models:

$$\begin{cases} \text{Model 1: } Y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon; \\ \text{Model 2: } Y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \beta_{k+1} x_{k+1} + \dots + \beta_{k+p} x_{k+p} + \varepsilon, \end{cases}$$

and we want to test $H_0 : \beta_{k+1} = \dots = \beta_{k+p} = 0$ vs $H_1 : \text{at least one } \beta_{k+i} \neq 0$,

$$\begin{cases} \frac{(SS_E^{(1)} - SS_E^{(2)})/p}{SS_E^{(2)}/(n-k-p-1)} \sim F(p, n-k-p-1) \\ \text{TS} = \frac{(SS_E^{(1)} - SS_E^{(2)})/p}{SS_E^{(2)}/(n-k-p-1)} \\ C = (F_\alpha(p, n-k-p-1), +\infty). \end{cases}$$

Variable selection. If we have a response variable y with possibly many predictors x_1, \dots, x_k , then how to choose appropriate x 's (some x 's are useful to Y , and some are not):

Step 1: $\text{corr}([x_1, \dots, x_k], y)$, choose a maximal correlation (say x_i), $Y = \beta_0 + \beta_i x_i + \varepsilon$, test if $\beta_i = 0$?

Step 2: do regression $Y = \beta_0 + \beta_i x_i + \beta_* x_* + \varepsilon$ for $* = 1, \dots, i-1, i+1, \dots, k$, choose a minimal SS_E (say x_j), $Y = \beta_0 + \beta_i x_i + \beta_j x_j + \varepsilon$, test if $\beta_j = 0$?

Step 3: repeat Step 2 until the last test for $\beta = 0$ is not rejected.

7 Basic χ^2 -test

Suppose we want to test $\begin{cases} H_0 : X \sim \text{distribution (with or without unknown parameters)}; \\ H_1 : X \not\sim \text{distribution (with or without unknown parameters)}. \end{cases}$

Then $\begin{cases} \text{fact is : } \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i} \sim \chi^2(k-1 - \# \text{of unknown parameters}); \\ \text{TS} = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}; \\ C = (\chi_\alpha^2(k-1 - \# \text{of unknown parameters}), +\infty). \end{cases}$

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

$\begin{cases} H_0 : \text{different rows have a same pattern (in terms of columns)}; \\ H_1 : \text{different rows have different patterns (in terms of columns)}. \end{cases}$

Equivalently,

$\begin{cases} H_0 : \text{rows and columns are independent}; \\ H_1 : \text{rows and columns are not independent}. \end{cases}$

Then

$\begin{cases} \text{fact is : } \sum_{j=1}^k \sum_{i=1}^r \frac{(N_{ij} - np_{ij})^2}{np_{ij}} \sim \chi^2((r-1)(k-1)); \\ \text{TS} = \sum_{j=1}^k \sum_{i=1}^r \frac{(N_{ij} - np_{ij})^2}{np_{ij}}; \\ C = (\chi_\alpha^2((r-1)(k-1)), +\infty), \end{cases}$

where $p_{ij} = p_i \cdot q_j$ are the theoretical probabilities.