Topics in statistics III/IV (MATH3361/4071) Michaelmas term, 2020-2021

Handout 6: Inferential tools based on asympt. efficient statistics

Lecturer & author: Georgios Karagiannis georgios.karagiannis@durham.ac.uk
References: [2, 1]

Notation 1. Let $X, X_1, X_2, ..., X_n$ be a sequence of independent random samples generated from a distribution f_θ labeled by a d-dimensional parameter $\theta \in \Theta \subset \mathbb{R}^d$, and admitting PDF $f(\cdot|\theta)$.

Assumption 2. *Assume that the conditions from the Cramer Theorem 19 (Handout 4) are satisfied.*

Notation 3. Consider that $\hat{\theta}_n$ is the MLE of θ.

- Although the methods below use the MLE $\hat{\theta}_n$, in fact, any asymptotic equivalent estimator $\clubsuit_n$ of the MLE can be used; e.g., the one-step-estimators with moment estimator initial guess.

  Recall that: asymptotic equivalent $\clubsuit_n - \hat{\theta}_n \xrightarrow{P} 0 \implies$ asymptotic efficient $\sqrt{n}(\clubsuit_n - \theta_0) \xrightarrow{D} \mathrm{N}(0, \mathcal{I}(\theta_0)^{-1})$

- All the tools below are asymptotic equivalent, as you can imagine. However, for smaller samples it seems that the Likelihood ration is the most powerful.

Note 4. We present 3 types of HT and CI. Regarding the hypothesis test the rational is depicted in Figure 1.

1 Likelihood ratio based

1.1 Pivotal statistic

Summary 5. Let the log likelihood ratio statistic be

$$W_{\mathrm{LR}}(\theta) = -2(\ell_n(\theta) - \ell_n(\hat{\theta}_n))$$

If θ_0 is the true value of θ then

$$W_{\mathrm{LR}}(\theta_0) \xrightarrow{D} \chi^2_d$$

and this is used as a Pivotal value.

Theorem 6. *Let $X_1, X_2, ..., X_n$ be independent random samples generated from a distribution f_θ labeled by a d-dimensional parameter $\theta \in \Theta \subset \mathbb{R}^d$, and admitting PDF $f(\cdot|\theta)$. Assume the conditions from the Cramer Theorem 19 (Handout 4) are satisfied, and that θ_0 is the true value, then*

$$W_{LR}(\theta_0) = -2(\ell_n(\theta_0) - \ell_n(\hat{\theta}_n)) \xrightarrow{D} \chi^2_d$$

it is where $\hat{\theta}_n$ is the MLE of θ.

Proof. It is given as a homework.[1] □

[1] Layout of the proof:

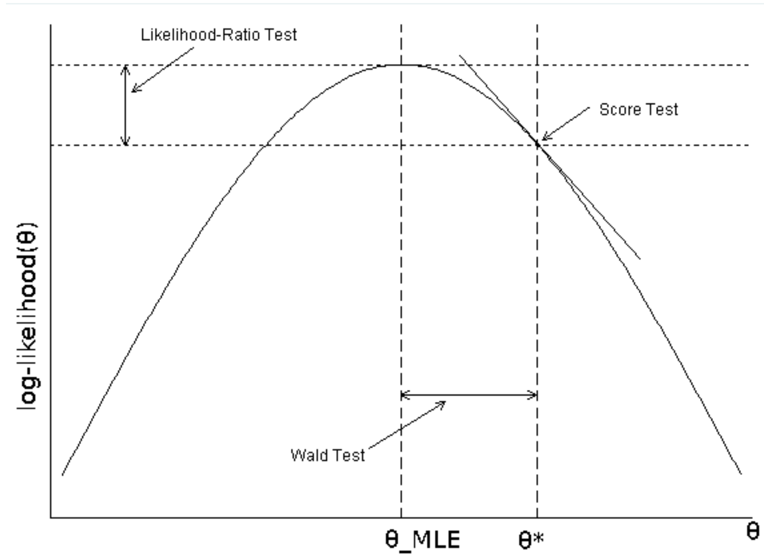

Figure 1: Comparison, in 1D , of:
Likelihood ratio : $W_{\text{LR}}(\theta_0) = -2(\ell_n(\theta_0) - \ell_n(\hat{\theta}_n))$
Wald : $W_{\text{Wald}}(\theta_0) = n(\theta_0 - \hat{\theta}_n)^2 \mathcal{I}(\theta_0) = -(\theta_0 - \hat{\theta}_n)^2 \text{E}(\ddot{\ell}_n(\theta_0))$
Score statistic : $W_{\text{Score}}(\theta_0) = n\dot{\ell}_n(\theta_0)/\mathcal{I}(\theta_0) = -\dot{\ell}_n(\theta_0)/\text{E}(\ddot{\ell}_n(\theta_0))$

1.2 Hypothesis test

Proposition 7. *Consider Hypothesis test*

$$H_0 : \theta = \theta_* \quad vs. \quad H_1 : \theta \neq \theta_*$$

Hence the rejection area, at sig. level a, is

$$\begin{aligned} RA(X_{1:n}) &= \{X_{1:n} : W_{LR}(\theta_*) \geq \chi^2_{d,1-a}\} \\ &= \{X_{1:n} : -2(\ell_n(\theta_*) - \ell_n(\hat{\theta}_n)) \geq \chi^2_{d,1-a}\} \end{aligned}$$

1.3 Confidence intervals

Proposition 8. *The $(1-a)$ confidence interval for θ is*

$$\begin{aligned} CI(\theta) &= \{\theta \in \Theta : W_{LR}(\theta) \leq \chi^2_{d,1-a}\} \\ &= \{\theta \in \Theta : -2(\ell_n(\theta) - \ell_n(\hat{\theta}_n)) \leq \chi^2_{d,1-a}\} \end{aligned}$$

produced by inverting the $RA(x_{1:n})$

Hint-1 Expand $\ell_n(\theta_0)$ around $\hat{\theta}_n$ by Taylor expansion

Hint-2 Prove that

$$W_{\text{LR}}(\theta_0) \xrightarrow{a.s} n(\theta_0 - \hat{\theta}_n)^T \mathcal{I}(\theta_0)(\theta_0 - \hat{\theta}_n) \tag{1}$$

Hint-3 Prove that $W_{\text{LR}}(\theta_0) \xrightarrow{D} \chi^2_d$

1.4 Comments

1. Regarding the HT, the comparison relies on the distance of the log-likelihood ratio $\ell_n(\theta_*)$ and $\ell_n(\hat{\theta}_n)$. The larger the distance is, the biggest doubt about the H_0 based on my data. See Figure 1.

2. It is more powerful than the other 2 tests, and hence preferable if it can by practically evaluated. The other 2 were derived possibly because ages ago people did not have computers and wanted to use something computationally cheaper.

2 Wald based

2.1 Pivotal statistic

Summary 9. We can use as a (Wald) pivotal statistics

$$W_{\mathrm{W}}(\theta_0) = n(\hat{\theta}_n - \theta_0)^T \mathcal{I}(\theta_0)(\hat{\theta}_n - \theta_0) \xrightarrow{D} \chi^2_d \tag{2}$$

$$W'_{\mathrm{W}}(\theta_0) = n(\hat{\theta}_n - \theta_0)^T \mathcal{I}(\hat{\theta}_n)(\hat{\theta}_n - \theta_0) \xrightarrow{D} \chi^2_d \tag{3}$$

$$W''_{\mathrm{W}}(\theta_0) = (\hat{\theta}_n - \theta_0)^T \mathcal{J}_n(\hat{\theta}_n)(\hat{\theta}_n - \theta_0) \xrightarrow{D} \chi^2_d \tag{4}$$

Any of them can be used to construct Hypothesis test, and Confidence intervals.

Remark 10. (Wald) pivotal statistics in (2, 3, and 4) are asymptotically equivalent for large samples (this is obvious by construction). However, the order of preference is $W_{\mathrm{W}}(\theta_0)$, $W'_{\mathrm{W}}(\theta_0)$, $W''_{\mathrm{W}}(\theta_0)$ when the sample size is not that large, however the proof is out of scope.

Remark 11. The asymptotic distributions of (Wald) pivotal statistics in (2, 3, and 4) can be derived based on the From the Cramer Theorem 19 (Handout 4). As consequences of Cramer Theorem, I get

$$\sqrt{n}\mathcal{I}(\theta_0)^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathrm{N}(0, I) \tag{5}$$

$$\sqrt{n}\mathcal{I}(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathrm{N}(0, I) \tag{6}$$

$$\mathcal{J}_n(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathrm{N}(0, I) \tag{7}$$

(the proof is in Lemma 20). Then by using Slutsky Theorem, and the fact that that $\sum_{i=1}^d z_i^2 \sim \chi^2_d$ when $z_i \sim \mathrm{N}(0,1)$ for $i = 1, ..., d$, I get the equations 2, 3, and 4.

2.2 Hypothesis test

Proposition 12. *Consider Hypothesis test*

$$H_0 : \theta = \theta_* \quad vs. \quad H_1 : \theta \neq \theta_*$$

Hence the rejection area, at sig. level a, is

$$RA(X_{1:n}) = \{X_{1:n} : W_{Wald}(\theta_0) \geq \chi^2_{d,1-a}\} = \{X_{1:n} : n(\hat{\theta}_n - \theta_*)^T \mathcal{I}(\theta_*)(\hat{\theta}_n - \theta_*) \geq \chi^2_{d,1-a}\}$$

etc...

2.3 Confidence intervals

Proposition 13. *The* $(1-a)$ *confidence interval for* θ *is*

$$\mathit{CI}(\theta) = \{\theta \in \Theta : W_{\mathit{Wald}}(\theta) \le \chi^2_{d,1-a}\} = \{\theta \in \Theta : n(\hat{\theta}_n - \theta)^T \mathcal{I}(\theta)(\hat{\theta}_n - \theta) \le \chi^2_{d,1-a}\}$$

etc... produced by inverting the $RA(x_{1:n})$

2.4 Comment

1. The Wald pivotal statistics are asymptotically equivalent to the LR one.
2. Regarding the HT, the comparison relies on the distance of the θ_* and $\hat{\theta}_n$, calibrated by the Information matrix (Fisher or Observed). The larger the distance is, the biggest doubt about the H_0 based on my data. See Figure 1.
3. Wald type of HT, CI are less expensive than the likelihood ratio ones because they require the computation of the expensive likelihood less number of times.

3 Score type tools

3.1 Pivotal statistic

Definition. The Score statistic is defined as

$$U(\theta) = [\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_n(\theta)]^T = [\sum_{i=1}^{d} \underbrace{\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X_i|\theta)}_{=\Psi(X_i,\theta)}]^T$$

where I put $\cdot^T$ because $U(\theta)$ is a $d \times 1$ vector.

Proposition 14. *The asymptotic distribution is*

$$\frac{1}{\sqrt{n}}U(\theta) \xrightarrow{D} N(0, \mathcal{I}(\theta))$$

which results as in Example/Proposition 16 (Handout 4). Then similar to above

$$\frac{1}{\sqrt{n}}\mathcal{I}(\theta)^{-1/2}U(\theta) \xrightarrow{D} N(0, I) \tag{8}$$

$$\frac{1}{\sqrt{n}}\mathcal{I}(\hat{\theta}_n)^{-1/2}U(\theta) \xrightarrow{D} N(0, I) \tag{9}$$

$$\mathcal{J}_n(\hat{\theta}_n)^{-1/2}U(\theta) \xrightarrow{D} N(0, I) \tag{10}$$

Therefore the following pivotal statistics can be used

$$W_{Score}(\theta_0) = \frac{1}{n}U(\theta)^T\mathcal{I}(\theta)^{-1}U(\theta) \xrightarrow{D} \chi^2_d \tag{11}$$

$$W'_{Score}(\theta_0) = \frac{1}{n}U(\theta)^T\mathcal{I}(\hat{\theta}_n)^{-1}U(\theta) \xrightarrow{D} \chi^2_d \tag{12}$$

$$W''_{Score}(\theta_0) = U(\theta)^T\mathcal{J}_n(\hat{\theta}_n)^{-1}U(\theta) \xrightarrow{D} \chi^2_d \tag{13}$$

with criterion the above preference order and their tractability.

3.2 Hypothesis test

Proposition 15. *Consider Hypothesis test*

$$H_0 : \theta = \theta_* \quad vs. \quad H_1 : \theta \neq \theta_*$$

Hence the rejection area, at sig. level a, is

$$\begin{aligned} RA(X_{1:n}) &= \{X_{1:n} : W_{Score}(\theta_0) \geq \chi^2_{d,1-a}\} \\ &= \{X_{1:n} : \frac{1}{n}U(\theta_*)^T\mathcal{I}(\theta_*)^{-1}U(\theta_*) \geq \chi^2_{d,1-a}\} \end{aligned} \tag{14}$$

etc...

3.3 Confidence intervals

Proposition 16. *The* $(1 - a)$ *confidence interval for* θ *is*

$$\begin{aligned} CI(\theta) &= \{\theta \in \Theta : W_{Score}(\theta) \leq \chi^2_{d,1-a}\} \\ &= \{\theta \in \Theta : \frac{1}{n}U(\theta)^T\mathcal{I}(\theta)^{-1}U(\theta) \leq \chi^2_{d,1-a}\} \end{aligned} \tag{15}$$

etc... produced by inverting the $RA(X_{1:n})$.

3.4 Comments

1. The Score pivotal statistics are asymptotically equivalent to the LR one.

2. Regarding the HT, the comparison relies on the slop of the log-likelihood at θ_* (aka the $U(\theta_*)$) calibrated by the curvature (Hessian matrix) at θ_*. The larger/steeper the slope, the bigger the distance from the peak (MLE $\hat{\theta}_n$), hence the biggest doubt about the H_0 based on my data. See Figure 1.

3. Score type of HT, CI are less expensive than the likelihood ratio ones because they require the computation of the expensive likelihood less number of times.

4. Score statistic type of HT and CI are computational convenient, in situations when the practitioner wants to calculate the HT or CI for parameter ϕ, which is a function of parameter θ whose Score type HT or CI have already been calculated.

   Let function $g : \mathbb{R}^d \to \mathbb{R}^k$ with $\phi = g(\theta)$. The score statistic

$$U^*(\phi) = [\frac{\mathrm{d}}{\mathrm{d}\phi}\ell_n(\theta)]^T = [\sum_{i=1}^{d} \frac{\mathrm{d}}{\mathrm{d}\phi} f(X_i|\phi)]^T$$

   is such that

$$U^*(\phi) = [\frac{\mathrm{d}}{\mathrm{d}\phi}\ell_n(\phi)]^T = [\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_n(\theta)\frac{\mathrm{d}\theta}{\mathrm{d}\phi}]^T = [\frac{\mathrm{d}\theta}{\mathrm{d}\phi}]^T U(\theta)$$

   has expected value

$$\mathrm{E}(U^*(\phi)) = \mathrm{E}([\frac{\mathrm{d}\theta}{\mathrm{d}\phi}]^T U(\theta)) = 0$$

   has variance

$$\mathrm{var}(U^*(\phi)) = \mathrm{var}([\frac{\mathrm{d}\theta}{\mathrm{d}\phi}]^T U(\theta)) = [\frac{\mathrm{d}\theta}{\mathrm{d}\phi}]^T \mathcal{I}(\theta)[\frac{\mathrm{d}\theta}{\mathrm{d}\phi}]$$

   and hence has asymptotic distribution

$$\frac{1}{\sqrt{n}}U^*(\phi) \xrightarrow{D} \mathrm{N}(0, \underbrace{[\frac{\mathrm{d}\theta}{\mathrm{d}\phi}]^T \mathcal{I}(\theta)[\frac{\mathrm{d}\theta}{\mathrm{d}\phi}]}_{=\mathcal{I}^*(\phi)})$$

   Hence, one can derive HT and CI as in (14) and (15) by substituting properly [2]. Notice that if the score HT and CI for θ are available then the score HT and CI for the transformation $\phi = g(\theta)$ can by computed by avoiding to recompute the expensive likelihood function

Example 17. Let random sample $x_1, ..., x_n \overset{IID}{\sim} \mathrm{Poi}(\theta)$, $\theta > 0$ with PDF

$$f(x|\theta) = \frac{\theta^x e^{-\theta}}{x!} 1(x > 0)$$

For the hypothesis test

$$\mathrm{H}_0 : \theta = \theta_* \quad \text{vs.} \quad \mathrm{H}_1 : \theta \neq \theta_*$$

Calculate

1. the log-likelihood ratio RA at a sig. level
2. the Wald's type RA at a sig. level (the 3 of them)
3. the Score's type RA at a sig. level (the 3 of them)

[2]please write down the derivation

Solution. Ok before that, let's calculate all the quantities required.

$$
\begin{aligned}
\log f(x|\theta) &\propto x\log(\theta)-\theta && ;; && 0=\dot{\ell}(\theta)|_{\theta=\hat{\theta}} \implies \hat{\theta}=\bar{x}\\
\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(x|\theta) &= \frac{x}{\theta}-1 && ;; && \ddot{\ell}(\theta)=-n\bar{x}\frac{1}{\theta^2}\\
\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(x|\theta) &= -\frac{x}{\theta^2} && ;; && \mathcal{J}_n(\theta)=-\ddot{\ell}(\theta)=n\bar{x}\frac{1}{\theta^2}\\
\mathcal{I}(\theta) &= -\mathrm{E}(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(x|\theta))=\frac{1}{\theta} && ;; && \mathcal{J}_n(\hat{\theta})=\frac{n}{\bar{x}}\\
\ell(\theta) &= n\bar{x}\log(\theta)-n\theta && ;; && U(\theta)=\dot{\ell}(\theta)^T=n\bar{x}\frac{1}{\theta}-n\\
\dot{\ell}(\theta) &= n\bar{x}\frac{1}{\theta}-n && ;; &&
\end{aligned}
$$

1. It is

$$
\begin{aligned}
\mathrm{CI}(\theta) &= \{\theta\in(0,\infty):-2(\ell_n(\theta)-\ell_n(\hat{\theta}_n))\le\chi^2_{1,1-a}\}\\
&= \{\theta\in(0,\infty):-2((n\bar{x}\log(\frac{\theta}{\bar{x}})-n(\theta-\bar{x}))\le\chi^2_{1,1-a}\}
\end{aligned}
$$

well, here we do not have the condition $n\theta-n\bar{x}$ like in the contingency tables ...

2. It is

$$
\begin{aligned}
\mathrm{RA}(x_{1:n}) &= \{x_{1:n}:n(\hat{\theta}_n-\theta_*)^T\mathcal{I}(\theta_*)(\hat{\theta}_n-\theta_*)\ge\chi^2_{1,1-a}\}\\
&= \{x_{1:n}:n(\bar{x}-\theta_*)^2\frac{1}{\theta_*}\ge\chi^2_{1,1-a}\}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{RA}(x_{1:n}) &= \{x_{1:n}:n(\hat{\theta}_n-\theta_*)^T\mathcal{I}(\hat{\theta}_n)(\hat{\theta}_n-\theta_*)\ge\chi^2_{1,1-a}\}\\
&= \{x_{1:n}:n(\bar{x}-\theta_*)^2\frac{1}{\bar{x}}\ge\chi^2_{1,1-a}\}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{RA}(x_{1:n}) &= \{x_{1:n}:(\hat{\theta}_n-\theta_*)^T\mathcal{J}_n(\hat{\theta}_n)(\hat{\theta}_n-\theta_*)\ge\chi^2_{1,1-a}\}\\
&= \{x_{1:n}:(\bar{x}-\theta_*)^2\frac{n}{\bar{x}}\ge\chi^2_{1,1-a}\}
\end{aligned}
$$

where the latter two CI coinside, however this is just a coinsidence...

3. It is

$$
\begin{aligned}
\mathrm{RA}(x_{1:n}) &= \{x_{1:n}:\frac{1}{n}U(\theta_*)^T\mathcal{I}(\theta_*)^{-1}U(\theta_*)\ge\chi^2_{1,1-a}\}\\
&= \{x_{1:n}:\frac{1}{n}(n\bar{x}\frac{1}{\theta_*}-n)^2(\frac{1}{\theta_*})^{-1}\ge\chi^2_{1,1-a}\}\\
&= \{x_{1:n}:n(\bar{x}\frac{1}{\theta_*}-1)^2\theta_*\ge\chi^2_{1,1-a}\}
\end{aligned}
$$

and

$$\begin{aligned}\mathrm{RA}(x_{1:n}) &= \{x_{1:n} : \frac{1}{n}U(\theta_*)^T\mathcal{I}(\hat{\theta}_n)^{-1}U(\theta_*) \geq \chi^2_{1,1-a}\} \\ &= \{x_{1:n} : n(\bar{x}\frac{1}{\theta_*} - 1)^2\hat{\theta}_n \geq \chi^2_{1,1-a}\}\end{aligned}$$

and

$$\begin{aligned}\mathrm{RA}(x_{1:n}) &= \{x_{1:n} : \frac{1}{n}U(\theta_*)^T\mathcal{J}_n(\hat{\theta}_n)^{-1}U(\theta_*) \geq \chi^2_{1,1-a}\} \\ &= \{x_{1:n} : n(\bar{x}\frac{1}{\theta_*} - 1)^2\hat{\theta}_n \geq \chi^2_{1,1-a}\}\end{aligned}$$

where the latter two CI coincide, however it is just a coincidence...

Confidence intervals can be computed by inverting the RA, based on the theory learnt in Concepts in Stats 2 (Term 1).

Example 18. Consider an M way contingency table $(n_{i,j})$ generated by a Poisson sampling scheme. Consider it is modeled by a log-Linear model with link function

$$\log(\mu) = X^T\beta \tag{16}$$

in the vectorized form, where vector $\beta \in \mathbb{R}^d$ contains the unknown coefficients. Consider that identifiability contains have be considered in (16). Show that

1. the asymptotic distribution of the MLEs $\hat{\beta}_n$ is such that

$$(X\mathrm{diag}(\mu_n)X^T)^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{D} \mathrm{N}(0, I) \tag{17}$$

   where $\hat{\mu}_n = \exp(X^T\hat{\beta}_n)$, and β_0 is the true value of β.

2. An $(1-a)100\%$ asymptotic confidence interval for β_0 is

$$\{(n_{i,j}) \,:\, (\hat{\beta}_n - \beta_0)^T(X\mathrm{diag}(\hat{\mu}_n)X^T)(\hat{\beta}_n - \beta_0) \leq \chi^2_{d,1-a}\}$$

Solution.

1. Since $\hat{\beta}_n$ is an MLE then it is asymptotically Normal as

$$\mathcal{J}_n(\hat{\theta}_n)^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{D} \mathrm{N}(0, I) \tag{18}$$

   It is

$$\mathcal{J}_n = -\frac{\mathrm{d}^2}{\mathrm{d}\beta^2}\ell_n(\beta)|_{\beta=\hat{\beta}_n} = -\frac{\mathrm{d}^2}{\mathrm{d}\beta^2}\sum_{i=1}^{n}\log(\mathrm{Poi}(n|\mu(\beta)))|_{\beta=\hat{\beta}_n}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\beta_j}\ell_n(\underset{\sim}{\mu}(\beta)) &= -\sum_i n_i X_{i,j} + \sum_i \exp(\sum_j X_{i,j}\beta_j)X_{i,j} \\ \frac{\mathrm{d}^2}{\mathrm{d}\beta_j\mathrm{d}\beta_k}\ell_n(\underset{\sim}{\mu}(\beta)) &= \sum_i \exp(\sum_j X_{i,j}\beta_j)X_{i,j}X_{i,k}\end{aligned}$$

and

$$\mathcal{J}_n = X^T \text{diag}(\hat{\mu}_n) X$$

Therefore from (18)

$$\underbrace{(X^T \text{diag}(\hat{\mu}_n) X)^{1/2} (\hat{\beta}_n - \beta_0)}_{=Z_n} \xrightarrow{D} \text{N}(0, I)$$

2. I use the statistic

$$T_n = (\hat{\beta}_n - \beta_0)^T (X^T \text{diag}(\hat{\mu}_n) X)(\hat{\beta}_n - \beta_0) = Z_n^T Z_n = \sum_{i=1}^{d} Z_{n,i}$$

where $T_n \sim \chi^2_d$ as summation of d standard normal variables. Then

$$1 - a = P_{\chi^2_d}(T_n < q)$$

where $q = \chi^2_{d,1-a}$.

Exercise sheet

Exercise #31
Exercise #32
Exercise #33
Exercise #34
Exercise #35

References

[1] Tom M Apostol. *Mathematical analysis; 2nd ed.* Addison-Wesley Series in Mathematics. Addison-Wesley, Reading, MA, 1974. URL `https://cds.cern.ch/record/105425`.

[2] Robert J Serfling. *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons, 2009.

A Appendix

Fact 19. *Cholesky decomposition: Every symmetric, positive definite matrix Σ can be decomposed into a product of a unique lower triangular matrix L and its transpose, i.e. $\Sigma = LL^T$.*

- *The following mathematical procedure can be used to compute the Cholesky factor L of Σ.*

  for $i = 1, ...d$

  for $j = 1, .., d$

$$L_{i,j} = \begin{cases} \sqrt{\Sigma_{i,i} - \sum_{k=1}^{i-1} L_{i,k}^2} & if i = j \\ \frac{1}{L_{j,j}}(\Sigma_{i,i} - \sum_{k=1}^{i-1} L_{i,k}L_{j,k}) & if i > j \\ 0 & if i < j \end{cases}$$

  – *the computations evolve row-wise, i.e. $L_{1,1} \to L_{2,1} \to L_{2,2} \to L_{3,1} \to L_{3,2} \to L_{3,3}$ etc...*

Lemma 20. *(Which will be used as a Proposition later on) Show that given that the assumptions [C.1-C.5] of Theorem 19 are satisfied, and that $\mathcal{I}(\theta)$ and $\mathcal{J}_n(\theta)$ are continuous on θ, then*

$$\sqrt{n}\mathcal{I}(\theta_0)^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, I) \tag{19}$$

$$\sqrt{n}\mathcal{I}(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, I) \tag{20}$$

$$\mathcal{J}_n(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, I) \tag{21}$$

where $A^{1/2}$ denotes the lower triangular matrix of the Cholesky decomposition of A; i.e., $A = A^{1/2}(A^{1/2})^T$.

Solution.

- Eq 19 results from Cramer Theorem, and the properties of covariance matrix.
- Eq. 20 results by using Cramer Theorem and Slutsky theorems. Precisely, because $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$, Slutsky implies $\mathcal{I}(\hat{\theta}_n) \xrightarrow{a.s.} \mathcal{I}(\theta_0)$ which implies $\mathcal{I}(\hat{\theta}_n)^{1/2}\mathcal{I}(\theta_0)^{-1/2} \xrightarrow{a.s.} I$. Therefore, by Slutsky

$$\underbrace{\mathcal{I}(\hat{\theta}_n)^{1/2}\mathcal{I}(\theta_0)^{-1/2}\sqrt{n}\mathcal{I}(\theta_0)^{1/2}(\hat{\theta}_n - \theta_0)}_{=\sqrt{n}\mathcal{I}(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0)} \xrightarrow{D} \underbrace{I \times \mathrm{N}(0, I)}_{=\mathrm{N}(0,I)}$$

- Eq. 21 results by using the USLLN and Slutsly theorems. So I just need to show that

$$\frac{1}{n}\mathcal{J}_n(\hat{\theta}_n) \xrightarrow{a.s.} \mathcal{I}(\theta_0)$$

  Set $U(x, \theta) = -\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log(f(x|\theta))$, and $\mathcal{I}(\theta) = \mathrm{E}(U(x, \theta))$. Then

$$|\frac{1}{n}\sum_{i=1}^{n}\underbrace{\left(-\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log(f(x_i|\hat{\theta}_n))\right)}_{U(x_i,\hat{\theta}_n)} - \mathcal{I}(\theta_0)| \le |\frac{1}{n}\sum_{i=1}^{n}U(x_i, \hat{\theta}_n) - \mathcal{I}(\hat{\theta}_n)| + |\mathcal{I}(\hat{\theta}_n) - \mathcal{I}(\theta_0)| \tag{22}$$

$$\le \sup_{|\hat{\theta}_n - \theta_0| \le \delta} |\frac{1}{n}\sum_{i=1}^{n}U(x_i, \theta) - \mathcal{I}(\theta)| + |\mathcal{I}(\hat{\theta}_n) - \mathcal{I}(\theta_0)| \tag{23}$$

The first term converges to zero because the assumptions of the USLLN are satisfied. The second term converges to zero because $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ and hence $\mathcal{I}(\hat{\theta}_n) \xrightarrow{a.s.} \mathcal{I}(\theta_0)$ by using Slutsky theorem.

So by Slutsy $(\frac{1}{n}\mathcal{J}_n(\hat{\theta}_n))^{1/2}\mathcal{I}(\theta_0)^{-1/2} \xrightarrow{a.s.} I$, and by Slutsky again

$$\underbrace{(\frac{1}{n}\mathcal{J}_n(\hat{\theta}_n))^{1/2}\mathcal{I}(\theta_0)^{-1/2}I(\theta_0)^{1/2}(\hat{\theta}_n - \theta_0)}_{=\mathcal{J}_n(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0)} \xrightarrow{D} \underbrace{I \times \mathrm{N}(0, I)}_{=\mathrm{N}(0,I)}$$