

Handout 6: Tools for inference under the presence of nuisance parameters

Lecturer & author: Georgios Karagiannis

georgios.karagiannis@durham.ac.uk

References: [4, (Ch. 4)], [5, (Ch. 8)], [2, (Ch. 10)], [3, (Ch. 4)], [1, (Ch. 22)]

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

Note 2. Assume the unknown d -dimensional parameter θ is partitioned as $\theta = (\psi, \phi)^\top$, by a d_ψ -dimensional $\psi \in \Psi \subset \mathbb{R}^{d_\psi}$, and d_ϕ -dimensional $\phi \in \Phi \subset \mathbb{R}^{d_\phi}$. Obviously $d = d_\psi + d_\phi$.

Definition 3. Given a statistical model $\{X_i \stackrel{\text{IID}}{\sim} f(\cdot|\theta)\}$ where $\theta = (\psi, \phi)$, if we are interested in learning the sub-parameter (or parameter function) $\psi = \psi(\theta)$, but we do not care about $\phi = \phi(\theta)$, the sub-parameter (or parameter function) ψ is called the parameter of interest, and the sub-parameter (or parameter function) ϕ is called the nuisance parameter.

Example 4. To motivate, consider the LR hypothesis test for comparing between two nested the log-linear models, $[X, YZ]$ and $[XY, YZ]$. Given a statistical model $\{X_i \stackrel{\text{IID}}{\sim} \text{Poi}(\mu(\lambda))\}$, what we did was:

$$\begin{aligned} \begin{cases} H_0 : [X, YZ] \\ H_1 : [XY, YZ] \end{cases} &\iff \begin{cases} H_0 : \log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ} \\ H_1 : \log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ} + \lambda_{jk}^{XY} \end{cases} \\ &\iff \begin{cases} H_0 : \lambda_{jk}^{XY} = 0, \text{ and any } \lambda, \lambda_i^X, \lambda_j^Y, \lambda_k^Z, \lambda_{jk}^{YZ} \in \mathbb{R} \\ H_1 : \lambda_{jk}^{XY} \neq 0, \text{ and any } \lambda, \lambda_i^X, \lambda_j^Y, \lambda_k^Z, \lambda_{jk}^{YZ} \in \mathbb{R} \end{cases} \iff \begin{cases} H_0 : \psi = \psi_*, \text{ and } \forall \phi \\ H_1 : \psi \neq \psi_*, \text{ and } \forall \phi \end{cases} \end{aligned}$$

where $\theta = \lambda = (\psi, \phi)$ is the unknown parameter, $\psi = \psi(\lambda) = (\lambda_{jk}^{XY})_{\forall i,j}$ is the parameter of interest, $\phi = \phi(\lambda) = (\lambda, \lambda_i^X, \lambda_j^Y, \lambda_k^Z, \lambda_{jk}^{YZ})_{\forall i,j}$ is the nuisance parameter, and $\psi_* = 0$, is the test value. This LR test does not actually fall in the category of the original likelihood ratio test in (Handout 6) which considers $H_0 : \theta = \theta_*$ vs $H_0 : \theta \neq \theta_*$ because we do not infer about parameters $\phi = (\lambda, \lambda_i^X, \lambda_j^Y, \lambda_k^Z, \lambda_{jk}^{YZ})_{\forall i,j}$ which just cause inconvenience.

Note 5. To learn ψ from the data $X_i \stackrel{\text{IID}}{\sim} f(\cdot|\theta = (\psi, \phi))$, as well as consider uncertainty about the unknown ϕ , I need to construct appropriate pivotal quantities $Q(\psi, X_{1:n})$ for ψ as well as compute their sampling distribution which should not depend on the unknown nuisance ϕ . One can derive such statistics by “profiling out” ϕ and constructing corresponding Likelihood ratio, Wald, or Score statistics whose asymptotic distribution can be easily derived.

Definition 6. Given a likelihood $L_n(\theta)$ the profile likelihood $L_{n,p}(\psi)$ of ψ is

$$L_{n,p}(\psi) = \sup_{\forall \phi} L_n(\underbrace{\psi, \phi}_{=\theta}) = L_n(\psi, \hat{\phi}_\psi)$$

where $\hat{\phi}_\psi$ denotes the MLE of ϕ as if ψ was a known parameter constant: i.e.

$$\hat{\phi}_\psi = \arg \sup_{\forall \phi \in \Phi} L_n(\psi, \phi)$$

Definition 7. The profile log-likelihood $\ell_{n,p}(\psi)$ of ψ , as

$$\ell_{n,p}(\psi) = \log(L_{n,p}(\psi)) = \log(L_n(\psi, \hat{\phi}_\psi)) = \ell_n(\psi, \hat{\phi}_\psi)$$

Note 8. Once the profile log-likelihood $L_{n,p}(\psi)$ of ψ is specified, then we can perform inference (point estimation, CI, HT, etc...) as usual but using $L_{n,p}(\psi)$.

1 Point estimation via profile maximum likelihood

Summary 9. The MLE $\hat{\psi} = \hat{\psi}(x_1, \dots, x_n)$ of ψ by profiling out ϕ is the

$$\hat{\psi} = \arg \sup_{\forall \psi \in \Psi} \ell_{n,p}(\psi) = \arg \sup_{\forall \psi \in \Psi} \ell_n(\psi, \hat{\phi}_\psi)$$

It can be found as follows:

1. Pretend that ψ is a known parameter and compute the MLE of ϕ

$$\hat{\phi}_\psi = \arg \sup_{\forall \phi \in \Phi} L_n(\psi, \phi)$$

e.g. as a root of the ML equations

$$0 = \frac{d}{d\phi} \ell_n(\psi, \phi)|_{\phi=\hat{\phi}_\psi}$$

2. Compute the profile MLE $\hat{\psi}$ (using the profile likelihood) as

$$\hat{\psi} = \arg \sup_{\forall \psi \in \Psi} \ell_{n,p}(\psi) = \arg \sup_{\forall \psi \in \Psi} \ell_n(\psi, \hat{\phi}_\psi)$$

e.g. as a root of the profile ML equations

$$0 = \frac{d}{d\psi} \ell_{n,p}(\psi)|_{\psi=\hat{\psi}} \quad \text{or equiv.} \quad 0 = \frac{d}{d\psi} \ell_n(\psi, \hat{\phi}_\psi)|_{\psi=\hat{\psi}}$$

Note 10. It can be seen that $(\hat{\psi}, \hat{\phi}_{\hat{\psi}})$ are the standard MLE: $(\hat{\psi}, \hat{\phi}) = \arg \sup_{\forall \psi, \phi} L_n(\psi, \phi)$; as

$$\sup_{\forall \psi} L_{n,p}(\psi) = \sup_{\forall \psi} \left(\sup_{\forall \phi} L_n(\psi, \phi) \right) = \sup_{\forall \psi, \phi} L_n(\psi, \phi)$$

Proposition 11. Assume the assumptions of Cramer Theorem 19 (Handout 4) are satisfied. Let $\theta_0 = (\psi_0, \phi_0)$ be the real value of θ . Then the profile MLE $\hat{\psi}$ is strongly consistent $\hat{\psi} \xrightarrow{as} \psi_0$, and its asymptotic distribution is such that

$$\sqrt{n}(\hat{\psi} - \psi_0) \xrightarrow{D} N\left(0, \left[\mathcal{I}_{11}(\theta_0) - \mathcal{I}_{21}(\theta_0)\mathcal{I}_{22}^{-1}(\theta_0)\mathcal{I}_{21}^\top(\theta_0)\right]^{-1}\right) \quad (1.1)$$

where $\{\mathcal{I}_{11}(\theta_0), \mathcal{I}_{21}(\theta_0), \mathcal{I}_{22}(\theta_0)\}$ is a partition of the Fisher Information matrix as

$$\mathcal{I}(\theta_0) = \begin{bmatrix} \mathcal{I}_{11}(\theta_0) & \mathcal{I}_{21}^\top(\theta_0) \\ \mathcal{I}_{21}(\theta_0) & \mathcal{I}_{22}(\theta_0) \end{bmatrix}.$$

Proof. That it is strongly consistent can be proven by considering a projection matrix $P = [I, 0]$ and applying Slutsky rules as $P[\hat{\psi}, \hat{\phi}]^\top \xrightarrow{as} P[\psi, \phi]^\top$. Regarding the asymptotic distribution, from Cramer theorem, it is

$$\sqrt{n} \begin{bmatrix} \hat{\psi} - \psi_0 \\ \hat{\phi} - \phi_0 \end{bmatrix} \xrightarrow{D} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathcal{I}_{11}(\theta_0) & \mathcal{I}_{21}^\top(\theta_0) \\ \mathcal{I}_{21}(\theta_0) & \mathcal{I}_{22}(\theta_0) \end{bmatrix}^{-1}\right)$$

Then the marginal distribution of ψ can be derived by the use of Delta method with transformation function $g(\theta) = [I, 0]\theta$ for $\psi = g(\theta)$, and the identity

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} [A - BD^{-1}C]^{-1} & -A^{-1}B[D - CA^{-1}B]^{-1} \\ -D^{-1}C[A - BD^{-1}C]^{-1} & [D - CA^{-1}B]^{-1} \end{bmatrix}$$

□

Remark 12. Note that if I applied directly Cramer theorem for $\hat{\psi}$ then

$$\sqrt{n}(\hat{\psi} - \psi_0) \xrightarrow{D} N\left(0, [\mathcal{I}_{11}(\theta_0)]^{-1}\right) \quad (1.2)$$

which would lead to overconfident inference because comparing with the asymptotic variance in (1.1), it is

$$\begin{aligned} & \left[\mathcal{I}_{11}(\theta_0) - \mathcal{I}_{21}(\theta_0)\mathcal{I}_{22}^{-1}(\theta_0)\mathcal{I}_{21}^\top(\theta_0)\right]^{-1} - [\mathcal{I}_{11}(\theta_0)]^{-1} \\ & \stackrel{(*)}{=} \left[\mathcal{I}_{21}(\theta_0) [\mathcal{I}_{11}(\theta_0)]^{-1}\right]^\top [D - CA^{-1}B]^{-1} \left[\mathcal{I}_{21}(\theta_0) [\mathcal{I}_{11}(\theta_0)]^{-1}\right] \geq 0 \end{aligned}$$

which is semi- positive definite. Here (*) by Woodbury matrix identity. This is reasonable as in (1.2) I ignored uncertainty about ϕ and the fact I used data to learn ϕ as well.

2 Popular pivotal statistics for CI & HT

Note 13. Due to the presence of nuisance parameter ϕ in the statistical model, we can resort to asymptotic pivotal statistics for ψ by profiling out ϕ from the original Likelihood ratio, Score, and Walds' pivots.

2.1 The Walds' pivotal statistic

Definition 14. Let $\theta_0 = (\psi_0, \phi_0)$ be the real value of θ . The Wald statistic is defined as

$$W_{W,p}(\psi) = n(\hat{\psi}_n - \psi)^T \left[\mathcal{I}_{11}(\theta_0) - \mathcal{I}_{21}(\theta_0) \mathcal{I}_{22}^{-1}(\theta_0) \mathcal{I}_{21}^T(\theta_0) \right] (\hat{\psi}_n - \psi) \quad (2.1)$$

Definition 15. Other, more tractable variations of the Wald statistic are

$$W'_{W,p}(\psi) = n(\hat{\psi}_n - \psi)^T \left[\mathcal{I}_{11}(\hat{\theta}_n) - \mathcal{I}_{21}(\hat{\theta}_n) \mathcal{I}_{22}^{-1}(\hat{\theta}_n) \mathcal{I}_{21}^T(\hat{\theta}_n) \right] (\hat{\psi}_n - \psi) \quad (2.2)$$

$$W''_{W,p}(\psi) = (\hat{\psi}_n - \psi)^T \left[\mathcal{J}_{n;11}(\hat{\theta}_n) - \mathcal{J}_{n;21}(\hat{\theta}_n) \mathcal{J}_{n;22}^{-1}(\hat{\theta}_n) \mathcal{J}_{n;21}^T(\hat{\theta}_n) \right] (\hat{\psi}_n - \psi) \quad (2.3)$$

Proposition 16. Assume the conditions from the Cramer Theorem 19 (Handout 4) are satisfied. Let $\theta_0 = (\psi_0, \phi_0)$ be the real value of $\theta = (\psi, \phi)$, with $\psi \in \mathbb{R}^{d_\psi}$, $\phi \in \mathbb{R}^{d_\phi}$, and $d = d_\psi + d_\phi$. Then $W_{W,p}(\psi_0) \xrightarrow{D} \chi_{d_\psi}^2$, $W'_{W,p}(\psi_0) \xrightarrow{D} \chi_{d_\psi}^2$, and $W''_{W,p}(\psi_0) \xrightarrow{D} \chi_{d_\psi}^2$ and they are all asymptotically equivalent.

Proof. The asymptotic equivalence can be proved by showing $W'_{W,p}(\psi_0) - W_{W,p}(\psi_0) \xrightarrow{p} 0$ for each pair. The Asymptotic distribution can be produced from (1.1) and Slutsky rules. \square

Proposition 17. Given a statistical model $\{X_i \stackrel{IID}{\sim} f(\cdot|\theta)\}$ where $\theta = (\psi, \phi)$, the Wald Hypothesis test for

$$H_0 : \psi = \psi_* \quad \text{vs.} \quad H_1 : \psi \neq \psi_*$$

has a rejection area, at sig. level α ,

$$RA(X_{1:n}) = \{X_{1:n} : W_{Wald}(\psi_0) \geq \chi_{d_\psi, 1-\alpha}^2\} \quad (2.4)$$

Similar is the rejection area produced by $W'_{W,p}(\psi)$ and $W''_{W,p}(\psi)$.

Proposition 18. Given a statistical model $\{X_i \stackrel{IID}{\sim} f(\cdot|\theta)\}$ where $\theta = (\psi, \phi)$, the $(1 - \alpha)$ confidence region for ψ is

$$CI(\psi) = \{\theta \in \Theta : W_{Wald}(\psi) \leq \chi_{d_\psi, 1-\alpha}^2\} \quad (2.5)$$

produced by inverting the $RA(x_{1:n})$. Similar is the confidence set produced by $W'_{W,p}(\psi)$ and $W''_{W,p}(\psi)$.

3 Score pivotal statistic

Definition. The profile score statistic is defined as

$$U_p(\psi) = \frac{d}{d\theta} \ell(\psi, \phi) \Big|_{(\psi, \hat{\phi}_\psi)} \quad (3.1)$$

Proposition 19. *The asymptotic distribution of the profile score statistic, given that θ is the real value of theta, is*

$$\frac{1}{\sqrt{n}}U_p(\psi) \xrightarrow{D} N\left(0, \left[\mathcal{I}_{11}(\theta) - \mathcal{I}_{21}(\theta)\mathcal{I}_{22}^{-1}(\theta)\mathcal{I}_{21}^\top(\theta)\right]\right) \quad (3.2)$$

Proof. The proof is available in [3, (Ch. 4)]. □

Definition 20. Let $\theta_0 = (\psi_0, \phi_0)$ be the real value of θ . The following score pivotal statistic is produced from the score statistic:

$$W_{\text{Score},p}(\psi) = \frac{1}{n} \left[\dot{\ell}_{n,p}(\psi) \right]^\top \left[\mathcal{I}_{11}(\theta_0) - \mathcal{I}_{21}(\theta_0)\mathcal{I}_{22}^{-1}(\theta_0)\mathcal{I}_{21}^\top(\theta_0) \right]^{-1} \dot{\ell}_{n,p}(\psi) \quad (3.3)$$

Definition 21. Other, more tractable variations of the Wald statistic are

$$W'_{\text{Score},p}(\psi) = \frac{1}{n} \left[\dot{\ell}_{n,p}(\psi) \right]^\top \left[\mathcal{I}_{11}(\hat{\theta}_n) - \mathcal{I}_{21}(\hat{\theta}_n)\mathcal{I}_{22}^{-1}(\hat{\theta}_n)\mathcal{I}_{21}^\top(\hat{\theta}_n) \right]^{-1} \dot{\ell}_{n,p}(\psi) \quad (3.4)$$

$$W''_{\text{Score},p}(\psi) = \left[\dot{\ell}_{n,p}(\psi) \right]^\top \left[\mathcal{J}_{n;11}(\hat{\theta}_n) - \mathcal{J}_{n;21}(\hat{\theta}_n)\mathcal{J}_{n;22}^{-1}(\hat{\theta}_n)\mathcal{J}_{n;21}^\top(\hat{\theta}_n) \right]^{-1} \dot{\ell}_{n,p}(\psi) \quad (3.5)$$

Proposition 22. *Assume the conditions from the Cramer Theorem 19 (Handout 4) are satisfied. Let $\theta_0 = (\psi_0, \phi_0)$ be the real value of $\theta = (\psi, \phi)$, with $\psi \in \mathbb{R}^{d_\psi}$, $\phi \in \mathbb{R}^{d_\phi}$, and $d = d_\psi + d_\phi$. Then $W_{\text{Score},p}(\psi_0) \xrightarrow{D} \chi_{d_\psi}^2$, $W'_{\text{Score},p}(\psi_0) \xrightarrow{D} \chi_{d_\psi}^2$, and $W''_{\text{Score},p}(\psi_0) \xrightarrow{D} \chi_{d_\psi}^2$ and they are all asymptotically equivalent.*

Proof. The asymptotic equivalence can be proved by showing $W'_{\text{Score},p}(\psi_0) - W_{\text{Score},p}(\psi_0) \xrightarrow{p} 0$ for each pair. The asymptotic distribution can be produced from Proposition 19 and Slutsky rules. The proof is available in [3, (Ch. 4)] □

Proposition 23. *Given a statistical model $\{X_i \stackrel{\text{IID}}{\sim} f(\cdot|\theta)\}$ where $\theta = (\psi, \phi)$, the Score Hypothesis test*

$$H_0 : \psi = \psi_* \quad \text{vs.} \quad H_1 : \psi \neq \psi_*$$

has a rejection area, at sig. level α ,

$$RA(X_{1:n}) = \{X_{1:n} : W_{\text{Score},p}(\psi_0) \geq \chi_{d_\psi, 1-\alpha}^2\} \quad (3.6)$$

Similar is the rejection area produced by $W'_{\text{Score},p}(\psi_0)$ and $W''_{\text{Score},p}(\psi_0)$.

Proposition 24. *Given a statistical model $\{X_i \stackrel{\text{IID}}{\sim} f(\cdot|\theta)\}$ where $\theta = (\psi, \phi)$, the $(1 - \alpha)$ Score confidence interval for ψ is*

$$CI(\psi) = \{\psi \in \Psi : W_{\text{Score},p}(\psi) \leq \chi_{d_\psi, 1-\alpha}^2\} \quad (3.7)$$

produced by inverting the $RA(X_{1:n})$. Similar is the confidence set based on $W'_{\text{Score},p}(\psi_0)$ and $W''_{\text{Score},p}(\psi_0)$.

3.1 Likelihood ratio (LR) pivotal statistic

Note 25. To profile out ϕ from the likelihood ratio statistic, it would be reasonable to modify the original likelihood ratio to use profiled likelihoods suitably

$$W_{LR,p}(\psi_*) = -2 \log \left(\frac{L_{n,p}(\psi_*)}{\sup_{\psi \neq \psi_*} L_{n,p}(\psi)} \right) = -2 \log \left(\frac{L_n(\psi_*, \hat{\phi}_{\psi_*})}{\sup_{\psi \neq \psi_*, \forall \phi} L_n(\psi, \phi)} \right) = -2(\ell_n(\psi_*, \hat{\phi}_{\psi_*}) - \ell_n(\hat{\theta}_n))$$

where $\hat{\theta}_n = (\hat{\psi}_n, \hat{\phi}_n)$ is the MLE of $\theta = (\psi, \phi)$.

Definition 26. Given a statistical model $\{X_i \stackrel{IID}{\sim} f(\cdot|\theta)\}$ where $\theta = (\psi, \phi)$, the log likelihood ratio statistic at ψ is

$$W_{LR,p}(\psi) = -2 \left(\ell_{n,p}(\psi) - \ell_{n,p}(\hat{\psi}_n) \right) = -2 \left(\ell_n(\psi, \hat{\phi}_{\psi_n}) - \ell_n(\hat{\psi}_n, \hat{\phi}_{\hat{\psi}_n}) \right) \quad (3.8)$$

where $\hat{\psi}_n$ is the profiled MLE of ψ .

Theorem 27. [Part of Wilks' Theorem (Appendix...)] Assume a statistical model $\{X_i \stackrel{IID}{\sim} f(\cdot|\theta)\}$ with unknown parameter $\theta = (\psi, \phi)$, where $\theta \in \Theta \subset \mathbb{R}^d$, $\psi \in \Psi \subset \mathbb{R}^{d_\psi}$, and $\phi \in \Phi \subset \mathbb{R}^{d_\phi}$. Assume the conditions from the Cramer Theorem 19 (Handout 4) are satisfied. Let $\theta_0 = (\psi_0, \phi_0)$ be the real value of $\theta = (\psi, \phi)$. Then

$$W_{LR,p}(\psi_0) = -2(\ell_{n,p}(\psi_0) - \ell_{n,p}(\hat{\psi}_n)) \xrightarrow{D} \chi_{d_\psi}^2 \quad (3.9)$$

where $\hat{\psi}_n$ is the profiled MLE of θ .

Proof. The proof is available in [3, (Ch. 4)], [1, (Ch. 22)]. □

Proposition 28. Given a statistical model $\{X_i \stackrel{IID}{\sim} f(\cdot|\theta)\}$ where $\theta = (\psi, \phi)$, the profiled LR hypothesis test for

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

has rejection area, at sig. level α , is

$$RA(X_{1:n}) = \{X_{1:n} : W_{LR}(\theta_*) \geq \chi_{d_\psi, 1-\alpha}^2\} \quad (3.10)$$

Proposition 29. Given a statistical model $\{X_i \stackrel{IID}{\sim} f(\cdot|\theta)\}$ where $\theta = (\psi, \phi)$, the $(1 - \alpha)$ profiled LR confidence region for ψ is

$$CI(\psi) = \{\psi \in \Psi : W_{LR}(\psi) \leq \chi_{d_\psi, 1-\alpha}^2\} \quad (3.11)$$

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

4 Examples

Example 30. Let random sample $x_1, \dots, x_n \stackrel{\text{IID}}{\sim} N(\mu, \sigma^2)$, where μ and σ^2 are unknown. We are interested in inference on μ .

1. Calculate the profile likelihood for μ
2. Find the likelihood ratio rejection area (at sig. level α) for the hypothesis test

$$H_0 : \mu = \mu_* \text{ vs. } H_1 : \mu \neq \mu_*$$

with respect to the t statistic $t = \sqrt{n} \frac{(\bar{x} - \mu_*)}{s}$, $s = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

Solution. Ok, I need to perform inference about the parameter of interest μ under the presence of a nuisance parameter σ^2 .

1. The profile likelihood is

$$L_{n,p}(\mu) = \sup_{\forall \sigma^2} L_n(\mu, \sigma^2) = L_n(\mu, \hat{\sigma}_\mu^2)$$

where $\hat{\sigma}_\mu^2$ is the n MLE of σ^2 for a given μ .

So first I need to find $\hat{\sigma}_\mu^2$. Okay, then, ...

The joint likelihood is

$$L_n(\mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

The joint log likelihood is

$$\ell_n(\mu, \sigma^2) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 + \text{const.}$$

So, to find $\hat{\sigma}_\mu^2$

$$0 = \frac{d}{d\sigma^2} \ell_n(\mu, \sigma^2) \big|_{\sigma^2 = \hat{\sigma}_\mu^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2} \frac{1}{(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 \big|_{\sigma^2 = \hat{\sigma}_\mu^2}$$

then

$$\hat{\sigma}_\mu^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

Hence the profile likelihood for μ is

$$\begin{aligned} L_{n,p}(\mu) &= L_n(\mu, \hat{\sigma}_\mu^2) = \left(\frac{1}{2\pi\hat{\sigma}_\mu^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\hat{\sigma}_\mu^2} \sum_{i=1}^n (x_i - \mu)^2\right) \\ &= \left(\frac{1}{2\pi} \frac{n}{\sum_{i=1}^n (x_i - \mu)^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2}n\right) \end{aligned}$$

2. To test

$$H_0 : \mu = \mu_* \text{ vs. } H_1 : \mu \neq \mu_*$$

I need to find the log likelihood ratio

$$W_{LR,p}(\mu_*) = -2 \log \left(\frac{\sup_{H_0} L_{n,p}(\mu)}{\sup_{H_1} L_{n,p}(\mu)} \right)$$

Under the null hypothesis H_0 it is

$$L_{n,p}(\mu_*) = \left(\frac{1}{2\pi} \frac{n}{\sum_{i=1}^n (x_i - \mu_*)^2} \right)^{\frac{n}{2}} \exp\left(-\frac{1}{2}n\right)$$

Under the alternative hypothesis H_1 it is

$$\begin{aligned} \sup_{H_1} L_{n,p}(\mu) &= \sup_{\forall \mu} \left(\frac{1}{2\pi} \frac{n}{\sum_{i=1}^n (x_i - \hat{\mu})^2} \right)^{\frac{n}{2}} \exp\left(-\frac{1}{2}n\right) \\ &= \left(\frac{1}{2\pi} \frac{n}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^{\frac{n}{2}} \exp\left(-\frac{1}{2}n\right) \end{aligned} \quad (4.1)$$

because the MLE of μ under the H_1 is $\hat{\mu} = \bar{x}$: In fact, under H_1 it is

$$0 = \frac{d}{d\sigma^2} \ell_n(\mu, \sigma^2) |_{\sigma^2 = \hat{\sigma}^2, \mu = \hat{\mu}} \implies \begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{cases}$$

or otherwise you can see that $L_{n,p}(\mu)$ maximizes by minimizing the sum-of-squares term. So

$$\begin{aligned} W_{LR,p}(\mu_*) &= -2 \log \left(\frac{\sup_{H_0} L_{n,p}(\mu)}{\sup_{H_1} L_{n,p}(\mu)} \right) = -2 \log \left(\frac{\left(\frac{1}{2\pi} \frac{n}{\sum_{i=1}^n (x_i - \mu_*)^2} \right)^{\frac{n}{2}} \exp\left(-\frac{1}{2}n\right)}{\left(\frac{1}{2\pi} \frac{n}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^{\frac{n}{2}} \exp\left(-\frac{1}{2}n\right)} \right) \\ &= n \log \left(\frac{\sum_{i=1}^n (x_i - \mu_*)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) = n \log \left(\frac{\sum_{i=1}^n (x_i \pm \bar{x} - \mu_*)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \\ &= n \log \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_*)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \\ &= n \log \left(1 + \frac{1}{n-1} \underbrace{n \frac{(\bar{x} - \mu_*)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}_{=t^2} \right) \\ &= n \log \left(1 + \frac{1}{n-1} t^2 \right) \xrightarrow{D} \underbrace{\chi_{2-1}^2}_{=1} \end{aligned}$$

where $t = \sqrt{n} \frac{(\bar{x} - \mu_*)}{s}$ with $s = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

Therefore the rejection area at sig. level α is

$$RA(x_{1:n}) = \{x_{1:n} : n \log \left(1 + \frac{1}{n-1} t^2 \right) \geq \chi_{1,1-\alpha}^2\}$$

References

- [1] Thomas S Ferguson. *A course in large sample theory*. Routledge, 2017.
- [2] Y. Pawitan. *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press, 2001.
- [3] Robert J Serfling. *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons, 2009.
- [4] T. A. Severini. *Likelihood methods in statistics*. Oxford University Press, 2000.
- [5] G. A. Young and R. L. Smith. *Essentials of statistical inference*, volume 16. Cambridge University Press, 2005.