Statistics 210A Lecture 24 Notes

Daniel Raban

November 18, 2021

1 Generalized Likelihood Ratio Tests, Asymptotic Relative Efficiency, and Pearson's χ^2 Test

1.1 Recap: Likelihood-ratio based hypothesis tests

We have been assuming a parametric model $X_1, \dots, X_n \overset{\text{iid}}{\sim} p_{\theta_0}(x)$, where $\theta_0 \in \Theta^o \subseteq \mathbb{R}^d$. $p_\theta(x)$ sufficiently regular in θ. We have the MLE

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \ell_n(\theta; X),$$

which we assume converges in probability to θ_0. The central limit theorem tells us that

$$\frac{1}{\sqrt{n}} \nabla \ell_n(\theta_0; X) \implies n_d(0, J_1(\theta_0)),$$

where we can think of $\nabla \ell_n$ as a complete sufficient statistic for all the likelihood ratios. We had the Taylor expansion

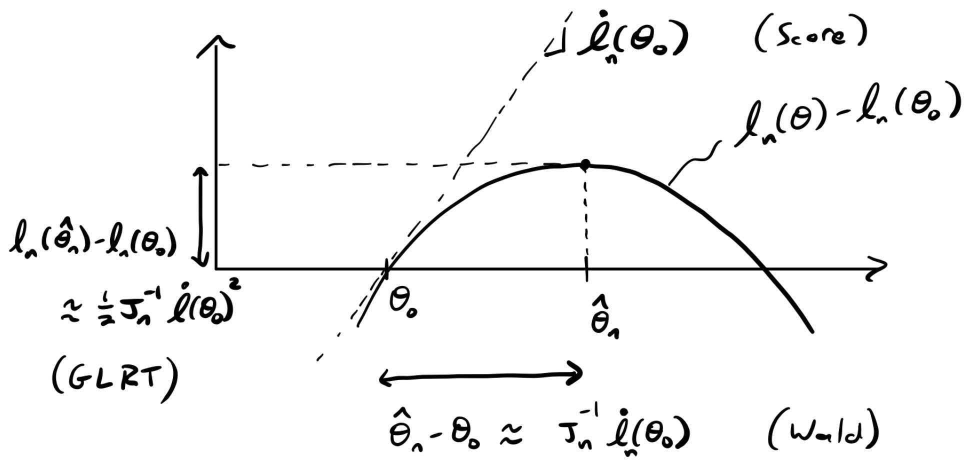
$$0 = \nabla \ell_n(\widehat{\theta}_n) = \nabla \ell_n(\theta_0) + \nabla^2 \ell_n(\widetilde{\theta}_n)(\widehat{\theta}_n - \theta_0),$$

which told us that

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_n - \theta_0) &= \underbrace{\left(-\frac{1}{n} \nabla^2 \ell_n(\widetilde{\theta}_n) \right)^{-1}}_{\xrightarrow{p} J_1(\theta_0)^{-1}} \underbrace{\frac{1}{\sqrt{n}} \nabla \ell_n(\theta_0)}_{\implies N(0, J_1(\theta_0))} \\ &\implies N_d(0, J_1(\theta_0)^{-1}).\end{aligned}$$

We have following picture of the second order Taylor approximation of the log-likelihood

$$\ell_n(\theta) - \ell_n(\theta_0) \approx \dot{\ell}_n(\theta_0)(\theta - \theta_0) - \frac{1}{2} J_n(\theta_0)(\theta - \theta_0)^2.$$



Different parts of this picture give us different likelihood-based test statistics for hypothesis testing.

For large n,

$$2(\ell_n(\widehat{\theta}_n) - \ell_n(\theta_0)) \approx \|J_n^{1/2}(\widehat{\theta}_n - \theta_0)\|^2,$$

which gives us the Wald test. Looking at

$$2(\ell_n(\widehat{\theta}_n) - \ell_n(\theta_0)) \approx \|J_n(\theta_0)^{-1/2}\nabla \ell_n(\theta_0)\|^2,$$

gives us the score test, and

$$2(\ell_n(\widehat{\theta}_n) - \ell_n(\theta_0)) \approx \|J_n(\theta)^{1/2}(\widehat{\theta}_n - \theta_0)\|^2,$$

gives us the generalized likelihood ratio test. This is looking at the vertical distance in the above picture.

1.2 Generalized likelihood ratio tests

1.2.1 GLRT with a simple null

Suppose we want to test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. We have

$$\begin{aligned}
\ell_n(\theta_0) - \ell_n(\widehat{\theta}_n) &= \cancelto{0}{\nabla \ell_n(\widehat{\theta}_n)} + \frac{1}{2}(\theta_0 - \widehat{\theta}_n)^{-1}\nabla^2 \ell_n(\widetilde{\theta}_n)(\theta_0 - \widehat{\theta}_n) \\
&= -\frac{1}{2}\| \underbrace{(-\tfrac{1}{n}\nabla^2 \ell_n(\widetilde{\theta}_n))^{1/2}}_{\xrightarrow{p} J_1(\theta_0)^{1/2}} \quad \underbrace{\sqrt{n}(\theta_0 - \widehat{\theta}_n)}_{\implies N_d(0, J_1(\theta_0)^{-1})} \quad \|^2 \\
&\implies -\frac{1}{2}\chi^2_d.
\end{aligned}$$

This means that

$$2(\ell_n(\widehat{\theta}_n) - \ell_n(\theta_0)) \implies \chi^2_d.$$

We should reject θ_0 if and only if

$$\ell_n(\theta_0) \leq \ell_n(\widehat{\theta}_n) - \frac{1}{2}\chi^2_d(\alpha).$$

This has some of the advantages of the Wald test, such as invariance under parameterization, but without requiring the confidence set to always be an ellipsoid.

1.2.2 GLRT with a composite null or with nuisance parameters

Theorem 1.1. *Suppose we are testing $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \notin \Theta_0$. Assume that*

- $\Theta \subseteq \mathbb{R}^d$*, where* $\Theta_0 \subseteq \Theta$ *is a* d_0*-dimensional manifold contained in* Θ^o.
- θ_0 *is in the relative interior of* Θ_0.
- $\widehat{\theta}_n \xrightarrow{p} \theta_0$ *with smooth likelihood.*
- $J_1(\theta) \succ 0$.

Then

$$2(\ell_n(\widehat{\theta}_n) - \ell_n(\widehat{\theta}_0)) \implies \chi^2_{d-d_0},$$

where $\widehat{\theta}_0 = \arg\max_{\theta \in \Theta_0} \ell_n(\theta; X)$.

Here is an informal derivation.

Proof. Assume without loss of generality that $\theta_0 = 0$ and $J_1(0) = I_d$. Then $\widehat{\theta}_n \approx N_d(0, \frac{1}{n} I_d)$, and locally ($\theta \approx 0$), $\nabla^2 \ell_n(\theta) \approx -n I_d$. Then

$$\ell_n(\theta) - \ell_n(\widehat{\theta}_n) \approx \frac{n}{2} \|\theta - \widehat{\theta}_n\|^2.$$

Then

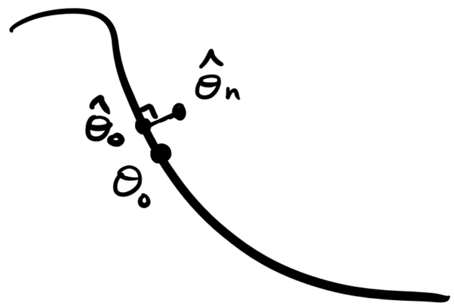
$$\widehat{\theta}_0 \approx \operatorname*{arg\,min}_{\theta \in \Theta_0} \|\theta - \widehat{\theta}_n\|^2 = \mathrm{Proj}_{\Theta_0}(\widehat{\theta}_n).$$

This means that -1 times the test statistic looks like

$$\begin{aligned} 2(\ell_n(\widehat{\theta}_0) - \ell(\widehat{\theta}_n)) &\approx -n \|\widehat{\theta}_n - \mathrm{Proj}_{\Theta_0}(\widehat{\theta}_n)\|^2 \\ &= -\| \mathrm{Proj}^{\perp}_{\Theta_0}(\underbrace{\sqrt{n}\widehat{\theta}_n}_{\approx N(0, I_d)})\|^2 \\ &\implies -\chi^2_{d-d_0}. \end{aligned}$$

□

Here is a picture when $d = 2$ and $d_0 = 1$.



The segment looks like χ_1^2.

1.3 Asymptotic relative efficiency

Suppose $\widehat{\theta}_n^{(i)}$ with $i = 1, 2$ are two estimators with $d = 1$ and

$$\sqrt{n}(\widehat{\theta}_n^{(i)} - \theta_0) \implies N(0, \sigma_i^2.$$

Definition 1.1. The **asymptotic relative efficiency (ARE)** of $\widehat{\theta}^{(2)}$ with respect to $\widehat{\theta}^{(1)}$ is σ_1^2/σ_2^2.

This has a nice interpretation of telling us that using an inefficient estimator is really like throwing away a fraction of our data set. Suppose $\sigma_1^2/\sigma_2^2 = \gamma \in (0, 1)$. Then

$$\begin{aligned} \widehat{\theta}_{\lfloor \gamma n \rfloor}^{(1)}(X_1, \ldots, X_{\lfloor \gamma n \rfloor}) &\approx N(\theta_0, \sigma_2^2/n) \\ &\overset{D}{\approx} \widehat{\theta}_n^{(2)}(X_1, \ldots, X_n). \end{aligned}$$

1.4 Pearson's χ^2 test for goodness of fit

Let $N = (N_1, \ldots, N_d) \sim \text{Multinom}(n, \pi)$, where $\pi = (\pi_1, \ldots, \pi_d)$ with $\sum_j \pi_j = 1$ and all $\pi_j > 0$. The multinomial density is

$$p_\theta(N) = \frac{n! \pi_1^{N_1} \cdots \pi_d^{N_d}}{N_1! \cdots N_d!} \mathbb{1}_{\{\sum_j N_j = n\}}.$$

We can parameterize this as a $d-1$-parameter exponential family by

$$\pi_j = \begin{cases} \frac{1}{1+\sum_{k>1} e^{\eta_k}} & j = 1, \\ \frac{e^{\eta_j}}{1+\sum_{k>1} e^{\eta_k}} & j > 1 \end{cases}$$

so that

$$\eta_j = \log(\pi_j + \pi_1).$$

We can calculate the score

$$\nabla \ell_n(\eta, N) = (N_2, \ldots, N_d) - (n\pi_2, \ldots, n\pi_d).$$

The variance of the score is

$$\begin{aligned}
\mathrm{Var}_\eta(\nabla \ell_n(\eta; N)) &= \begin{bmatrix} n\pi_2(1-\pi_2) & \cdots & -\pi_i\pi_j & & \cdots \\ & \ddots & & & \\ & & & & \\ & & & & n\pi_d(1-\pi_d) \end{bmatrix} \\
&= n(\mathrm{diag}(\pi_{2-d}) - \pi_{2-d}\pi_{2-d}^\top)
\end{aligned}$$

If we use the formula

$$(A + uv^\top)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u},$$

we get

$$J_n(\eta)^{-1} = \frac{1}{2}(\mathrm{diag}(\pi_{2,\ldots,d}^{-1}) + \pi_1^{-1}\mathbf{1}_{d_1}\mathbf{1}_{d_2}^\top).$$

After some algebra, it follows that the score test fo $H_0 : \pi = \pi_0$ vs $H_1 : \pi \neq \pi_0$ is

$$\begin{aligned}
\nabla \ell_n(\eta_0) J_n^{-1}(\eta_0) \nabla \ell_n(\eta_0) &= (N_{2,\ldots,d} - n\pi_{2,\ldots,d})^\top (\tfrac{1}{n}(\mathrm{diag}(\pi_{2,\ldots,d}^{-1}) + \pi_0\mathbf{1}))(N_{2,\ldots,d} - n\pi_0) \\
&= \sum_{j>1} \frac{(N_j - n\pi_j)^2}{n\pi_j} - \frac{1}{n\pi_n}\mathbf{1}^\top (N_{2,\ldots,d} + n\pi_{2,\ldots,d})^2 \\
&= \sum_j \frac{(N_j - n\pi_j)^2}{n\pi_j}.
\end{aligned}$$

This is the test statistic for Pearson's χ^2 test.