Review Session 8 – Asymptotic Analysis of Tests and Estimators

References/suggested reading

(i) Casella & Berger sections 5.5.4, 10.1, 10.3.

1 Introduction

Building on previous review sessions, we'll study the asymptotic properties of the point estimation and hypothesis testing procedures we've developed thus far. In particular, we will focus on the maximum likelihood estimator (MLE), and this will in turn, not too surprisingly, give us information about the asymptotics of the likelihood ratio test statistic for the hypothesis problem.

2 Consistency

Definition 2.1 (consistency)**.** A sequence of estimators $W_n = W_n(X_1, \ldots, X_n)$ is a *consistent* sequence of estimators of the parameter θ if, $W_n \xrightarrow{P} \theta$ or for every $\epsilon > 0$ and $\theta \in \Theta$:

$$\lim_{n\to\infty} \mathbb{P}_\theta(|W_n - \theta| < \epsilon) = 1.$$

Example 2.2 (consistency of $\overline{X}_n$ and S^2)
Recall from Review Session 5 that the sample mean $\overline{X}_n$ is a consistent estimator of $\mathbb{E}[X]$ (by the law of large numbers) and the sample variance S_n^2 is a consistent estimator of $\mathrm{Var}(X)$, provided that $\mathrm{Var}(S_n^2) \to 0$. The proof of this last fact used Chebyshev's inequality or that

$$\mathbb{P}_\theta(|S_n^2 - \mathrm{Var}(X)| \geq \epsilon) \leq \frac{\mathbb{E}(S_n^2 - \mathrm{Var}(X))^2}{\epsilon^2} = \frac{\mathrm{Var}(S_n^2)}{\epsilon^2} \overset{n\to\infty}{\to} 0.$$

We can apply this technique more broadly. Chebyshev gives us for a generic sequence of estimators $\{W_n\}_{n=1}^\infty$,

$$\mathbb{P}_\theta(|W_n - \theta| \geq \epsilon) \leq \frac{\mathbb{E}_\theta[(W_n - \theta)^2]}{\epsilon^2}.$$

Thus, if the MSE $\mathbb{E}_\theta[(W_n - \theta)^2] \to 0$, then we know this sequence of estimators is consistent. Furthermore, the bias-variance decomposition gives us a more refined way of guaranteeing the consistency of W_n:

$$\mathbb{E}_\theta[(W_n - \theta)^2] = \mathrm{Var}_\theta(W_n) + (\mathbb{E}_\theta(W_n) - \theta)^2.$$

Theorem 2.3
If $\{W_n\}_{n=1}^\infty$ is a sequence of estimators of a parameter θ satisfying

1. $\lim_n \mathrm{Var}_\theta(W_n) = 0.$
2. $\lim_n(\mathbb{E}_\theta(W_n) - \theta) = 0,$

for every $\theta \in \Theta$, then $\{W_n\}_{n=1}^\infty$ is a consistent sequence of estimators of θ. In particular, if $\{W_n\}_{n=1}^\infty$ is a sequence of unbiased estimators, then it suffices to have $\lim_n \mathrm{Var}_\theta(W_n) = 0$ for all $\theta \in \Theta$.

Sometimes, we are interested in a linear transformation of an estimator. For example, we may consider the sample variance to take the unbiased form $S_n^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ or the form of the MLE $S_n^2 := \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Recall that Slutsky's theorem tells us that if $A_n \xrightarrow{\mathsf{P}} A$ and $b_n \xrightarrow{\mathsf{P}} b$ where $\{b_n\}_{n=1}^\infty$ is a real sequence, then $A_n \cdot b_n \xrightarrow{\mathsf{P}} A \cdot b$. Thus, the consistency of one of these estimators immediately implies the consistency of the other. This applies much more broadly as we see in the following result.

Theorem 2.4

Let $\{W_n\}_{n=1}^\infty$ be a consistent sequence of estimators of a parameter θ. Let $\{a_n\}_{n=1}^\infty, \{b_n\}_{n=1}^\infty$ be sequences of constants satisfying $\lim_n a_n = 1$ and $\lim_n b_n = 0$. Then, the sequence $U_n = a_n \cdot W_n + b_n$ is a consistent sequence of estimators of θ.

3 The Delta Method

Beyond the fact that a sequence of estimators $\{W_n\}_{n=1}^\infty$ is consistent, we often want to understand the rate of its convergence to a parameter θ, similar to how the Central Limit Theorem gives us rates of convergence on the estimator $\overline{X}_n$ for $\mathbb{E}[X]$. In particular, we'd like to understand the limiting behavior of the deviations $W_n - \theta$ and, in particular, determine how fast it goes to 0. Here, "how fast" should make you think of big-oh $O(\cdot)$ and little-oh $o(\cdot)$ notation. Asymptotic notation describes the speed of convergence as a sequence of real numbers $\{e_n\}_{n=1}^\infty$ such that $\frac{W_n-\theta}{e_n}$ goes to a constant. However, here, $W_n - \theta$ is a random variable. So, we'll rephrase this as seeking a sequence of real numbers $\{e_n\}_n$ such that $\frac{W_n-\theta}{e_n}$ converges in distribution to some non-degenerate distribution.

We've already seen one example of such a characterization: the central limit theorem tells us that for the sample mean estimator $\overline{X}_n$ of the population mean $\mu = \mathbb{E}[X]$, we have

$$\frac{\sqrt{n} \cdot (\overline{X}_n - \mu)}{\sqrt{\mathrm{Var}(X)}} \xrightarrow{\mathsf{d}} \mathcal{N}(0,1).$$

Thus, here the speed of convergence $\{e_n\}$ can be taken as $e_n = \frac{1}{\sqrt{n}}$. This central limit theorem will in fact be the basis for understanding the convergence rate of many other estimators. In particular, if our estimator $W(X_1, \ldots, X_n)$ is a function of $\overline{X}_n$, then we will be able to directly deduce a CLT-like statement of convergence for $\{W_n\}$.

Theorem 3.1 (delta method)

Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \sigma^2)$. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \sigma^2(g'(\theta))^2)$$

Proof. (idea) Taylor expanding $g(Y_n)$ about $Y_n = \theta$ gives us

$$g(Y_n) = g(\theta) + g'(\theta) \cdot (Y_n - \theta) + \text{remainder}.$$

where the remainder goes to 0 quickly as $Y_n \to \theta$. In other words,

$$\sqrt{n} \cdot (g(Y_n) - g(\theta)) \approx g'(\theta) \cdot \sqrt{n}(Y_n - \theta),$$

from which the result follows from the hypothesis $\sqrt{n}(Y_n - \theta) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \sigma^2)$ and Slutsky's theorem. ■

Example 3.2

We can estimate the inverse-mean $1/\mathbb{E}[X] =: 1/\mu$ by $1/\overline{X}_n$. Then, by the delta method, for $\mu \neq 0$, we have

$$\sqrt{n}\left(\frac{1}{\overline{X}_n} - \frac{1}{\mu}\right) \xrightarrow{\mathsf{d}} \mathcal{N}\left(0, \left(\frac{1}{\mu}\right)^4 \operatorname{Var}(X)\right)$$

This may not be such a useful limiting distribution if $\operatorname{Var}(X)$ is unknown. However, provided the sample variance S_n^2 is consistent for $\operatorname{Var}(X)$, Slutsky's theorem tells us that we can simply substitute S_n^2 for $\operatorname{Var}(X)$:

$$\frac{\sqrt{n}\left(\frac{1}{\overline{X}_n} - \frac{1}{\mu}\right)}{S_n} \xrightarrow{\mathsf{d}} \mathcal{N}(0, (1/\mu)^4).$$

We might also want to determine a limiting distribution which is free of μ since μ is an unknown parameter here. This is the case, for example, when forming an asymptotic confidence interval for our parameter. However, we know by consistency of $\overline{X}_n \xrightarrow{\mathsf{d}} \mu$ and the continuous mapping theorem (cf. review session 5), that $1/\overline{X}_n^4 \xrightarrow{\mathsf{d}} 1/\mu^4$. Thus,

$$\frac{\sqrt{n}\left(\frac{1}{\overline{X}_n} - \frac{1}{\mu}\right)}{\left(\frac{1}{\overline{X}_n}\right)^2 \cdot S_n} \xrightarrow{\mathsf{d}} \mathcal{N}(0,1)$$

The above example only worked provided $\mu \neq 0$. What if $\mu = 0$, or in general $g'(\theta) = 0$ in the delta method statement? Then, we can use a higher order Taylor expansion of $1/\overline{X}_n$ and follow the same principle. This is the motivation for the second-order delta method.

Theorem 3.3 (second-order delta method)

Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \xrightarrow{\mathsf{d}} \mathcal{N}(0, \sigma^2)$. For a given function g and a specific value of θ, suppose that $g'(\theta) = 0$ and $g''(\theta)$ exists and is not 0. Then

$$n(g(Y_n) - g(\theta)) \xrightarrow{\mathsf{d}} \sigma^2 \frac{g''(\theta)}{2} \chi_1^2$$

Proof. (idea) A second-order Taylor expansion of $g(Y_n)$ gives

$$g(Y_n) = g(\theta) + g'(\theta) \cdot (Y_n - \theta) + \frac{g''(\theta)}{2} \cdot (Y_n - \theta)^2 + \text{remainder}.$$

Acknowledging that $g'(\theta) = 0$ and rearranging this gives us

$$g(Y_n) - g(\theta) \approx \frac{g''(\theta)}{2} \cdot (Y_n - \theta)^2.$$

From this, we have that since the square of a standard-normal $\mathcal{N}(0,1)$ is a χ_1^2 random variable:

$$\frac{n(Y_n - \theta)^2}{\sigma^2} \xrightarrow{\mathsf{d}} \chi_1^2.$$

This gives us the desired convergence. ■

Often, we are interested in several parameters or a vector parameter $\theta \in \mathbb{R}^p$. Multivariate CLT and the multivariate Taylor expansion then give us the multivariate analogue of the delta method.

Theorem 3.4 (multivariate delta method)

If $g : \mathbb{R}^p \to \mathbb{R}^\ell$ has Jacobian $\nabla g(\mathbf{a})$ for $\mathbf{a} \in \mathbb{R}^p$ and the sequence of random vectors $\{\mathbf{X}_i\}_{i=1}^\infty \subseteq \mathbb{R}^k$ satisfies for some $b > 0$:

$$n^b(\mathbf{X}_n - \mathbf{a}) \xrightarrow{\text{d}} \mathbf{Y},$$

for some random vector $\mathbf{Y} \in \mathbb{R}^k$, then

$$n^b(g(\mathbf{X}_n) - g(\mathbf{a})) \xrightarrow{\text{d}} (\nabla g(\mathbf{a}))^T\mathbf{Y}.$$

This version of the multivariate delta method is stated in a very general form, but (using multivariate CLT for our theorem hypothesis) $\mathbf{Y}$ would typically be a multivariate Gaussian, b would be $1/2$, and $\mathbf{X}_n$ would be $(\overline{X}_{1,n}, \dots, \overline{X}_{p,n})$, i.e. the vector of sample means of the components of a multivariate random sample.

Corollary 3.5 (multivariate delta method for univariate function of mean)

Let $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}^p$ be a random sample with $\mathbb{E}[\mathbf{X}_{ij}] = \mu_i$ and $\mathrm{Cov}(\mathbf{X}_{ik}, \mathbf{X}_{jk}) = \sigma_{ij}$. For a given function $g : \mathbb{R}^p \to \mathbb{R}$ with continuous first partial derivatives and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)$, letting

$$\tau^2 = \sum_i \sum_j \sigma_{ij} \cdot \frac{\partial g(\boldsymbol{\mu})}{\partial \mu_i} \cdot \frac{\partial g(\boldsymbol{\mu})}{\partial \mu_j} > 0,$$

we have

$$\sqrt{n}(g(\overline{X}_{1,n}, \dots, \overline{X}_{p,n}) - g(\mu_1, \dots, \mu_p)) \xrightarrow{\text{d}} \mathcal{N}(0, \tau^2).$$

This just follows from the previous theorem since, by Cramér-Wold, the univariate projection of a multivariate Gaussian (i.e., that given by multivariate CLT for the limit of $\sqrt{n}((\overline{X}_{1,n}, \dots, \overline{X}_{p,n}) - \boldsymbol{\mu})$) is a univariate Gaussian.

4 Consistency and Asymptotic Normality of the MLE

We've seen that the MLE for a population mean $\mu = \mathbb{E}[X]$ is often the sample mean $\overline{X}_n$. The central limit theorem gives us an "asymptotic normality" statement for this estimator. In the previous section, we've seen that the MLE for various functions of the mean $g(\mu)$ also obey asymptotic normality by the delta method and by the functional invariance of the MLE.

It turns out the consistency and asymptotic normality of the MLE hold under much broader conditions. The exact conditions required are not so important to know, as pretty much all exam problems will involve an MLE situation which abides by these conditions. However, we'll give a brief sketch of why certain conditions are required.

Theorem 4.1 (consistency of the MLE)

Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} f(x|\theta)$. Suppose the following regularity conditions hold:

(i) Θ, the parameter space, is compact.

(ii) $\log f(x|\theta)$ is continuous.

(iii) There exists $k(X)$, $\mathbb{E}_{X\sim f(x|\theta)}[k(X)] < \infty$ such that $|\log f(x|\theta)| \le k(x)$ for almost every x and all $\theta \in \Theta$.

(iv) $\forall \epsilon > 0$, $\sup_{\|\theta-\theta_0\|>\epsilon} \mathbb{E}_{X\sim f(x|\theta_0)}[\log f(X|\theta)] < \mathbb{E}_{X\sim f(x|\theta_0)}[\log f(X|\theta_0)]$.

Then, $\hat{\theta}_{\text{MLE}} \xrightarrow{\text{d}} \theta_0$.

Proof. (idea) Recall that $L(\theta|\mathbf{x}_n) := \prod_{i=1}^n f(x_i|\theta)$ is the likelihood function. The basic idea here is to then show that $\frac{1}{n}\log L(\hat{\theta}_{\text{MLE}}|\mathbf{x})$ converges to $\mathbb{E}_{X\sim f(x|\theta_0)}[\log(f(X|\theta)]$. This looks very similar to the law of large numbers, except the value of $\hat{\theta}_{\text{MLE}}$ is random and changes with n. We resolve this by asserting a uniform law of large numbers

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\log(f(X|\theta)) - \mathbb{E}_{X\sim f(x|\theta_0)}[\log(f(X|\theta))]\right| \xrightarrow{\mathsf{d}} 0.$$

This requires conditions (i)–(iii). The intuition here is that if Θ is finite $|\Theta| < \infty$, then this uniform law of large numbers naturally follows from a union bound plus the usual law of large numbers. Θ being compact means we can approximate it with a finite covering by Euclidean balls, so we should hopefully be able to make a similar argument. For this, we need the value of $\log f(x|\theta)$ to not vary too much with θ inside of each ball, which is ensured by continuity (ii). To further ensure that $\mathbb{E}_{X\sim f(x|\theta_0)}[\log(f(X|\theta))]$ does not vary too much with θ inside of each ball, we need to be able to exchange limits and integrals, for which (iii) is a standard assumption.

Then, this gives us $\frac{1}{n}\log L(\theta|\mathbf{x}) \xrightarrow{\mathsf{d}} \mathbb{E}_{X\sim f(x|\theta_0)}[\log(f(X|\theta))]$ for all $\theta \in \Theta$. Intuitively, this should imply that the maximizer of $\frac{1}{n}\log L(\theta|\mathbf{x})$ over θ should converge to the maximizer of $\mathbb{E}_{X\sim f(x|\theta_0)}[\log(f(X|\theta))]$, which will correspond to our desired statement $\hat{\theta}_{\mathsf{MLE}} \xrightarrow{\mathsf{d}} \theta_0$. However, this is only assuming that $\theta = \theta_0$ maximizes the function $G(\theta) := \mathbb{E}_{X\sim f(x|\theta_0)}[\log f(X|\theta)]$. Condition (iv) essentially ensures this is the case by mandating that $G(\theta) < G(\theta_0)$ for θ far away from θ_0. ■

Like the estimators seen in the previous sectoin, the MLE also has asymptotic normality, with a limiting variance given by the Cramér-Rao lower bound (cf. review doc 6).

Theorem 4.2 (asymptotic normality of the MLE, also called Cramer's theorem)

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f(x|\theta)$. Under several regularity conditions in addition to (i)–(iv) from Theorem 4.1, we have

$$\sqrt{n}(\hat{\theta}_{\mathsf{MLE}} - \theta) \xrightarrow{\mathsf{d}} \mathcal{N}(0, I(\theta)^{-1}),$$

where $I(\theta) := \mathbb{E}_{X\sim f(x|\theta)}[-\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)]$ is the Fisher information for θ, so that the variance above $I(\theta)^{-1}$ is the Cramér-Rao lower bound.

Proof. (idea) Consider the log-likelihood function $\ell(\theta|\mathbf{x}_n) := \sum_{i=1}^{n}\log(f(x_i|\theta))$. Next, similar to the delta method, we will consider a Taylor expansion of the first-derivative $\ell'(\theta|\mathbf{x}_n) = \frac{\partial}{\partial\theta}\ell(\theta|\mathbf{x}_n)$, about the true parameter value $\theta = \theta_0$:

$$\ell'(\theta|\mathbf{x}_n) = \ell'(\theta_0|\mathbf{x}) + (\theta - \theta_0)\cdot\ell''(\theta_0|\mathbf{x}_n) + \text{remainder}.$$

Letting $\theta = \hat{\theta}_{\mathsf{MLE}}$ in the above, we see that the LHS is zero by the definition of the MLE. Ignoring the remainder, we then have

$$\sqrt{n}(\hat{\theta}_{\mathsf{MLE}} - \theta_0) \approx \sqrt{n}\cdot\frac{-\ell'(\theta_0|\mathbf{x}_n)}{\ell''(\theta_0|\mathbf{x}_n)} = \frac{-\frac{1}{\sqrt{n}}\ell'(\theta_0|\mathbf{x}_n)}{\frac{1}{n}\ell''(\theta_0|\mathbf{x}_n)}.$$

In the above RHS, the denominator $\frac{1}{n}\ell''(\theta_0|\mathbf{x}_n)$ will go to its mean $\mathbb{E}_{X\sim f(x|\theta_0)}[\ell''(\theta_0|X)]$ by the law of large numbers. Meanwhile, the numerator can be re-written as

$$-\frac{1}{\sqrt{n}}\ell'(\theta_0|\mathbf{x}_n) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log(f(x_i|\theta_0))\right).$$

The random variable $W_i := \frac{\partial}{\partial\theta}\log(f(X_i|\theta_0))$ can be shown to have mean zero and variance $I(\theta_0)$. Thus, by CLT

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log(f(X_i|\theta_0))\right) \xrightarrow{\mathsf{d}} \mathcal{N}(0, I(\theta_0)).$$

Then, putting everything together we get $\sqrt{n}(\hat{\theta}_{\mathsf{MLE}} - \theta) \xrightarrow{\mathsf{d}} \mathcal{N}(0, I(\theta)^{-1})$. The "extra regularity" conditions alluded to in the theorem statement here are just to make sure that the remainder terms in our Taylor expansion do indeed vanish as $n \to \infty$. ■

We see from the above that the MLE has a nice and somewhat optimal form for its asymptotic variance. We call this *asymptotic efficiency*, which we formalize as follows:

Definition 4.3 (asymptotic variance and asymptotic efficiency)**.** For a sequence of estimators $\{W_n\}_{n=1}^\infty$, suppose that $\frac{W_n - \theta}{e_n} \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2)$ for some real sequence $\{e_n\}_{n=1}^\infty$. The parameter σ^2 is called the *asymptotic variance*, or limiting variance, of $\{W_n\}_{n=1}^\infty$.

We say a sequence of estimators $\{W_n\}_{n=1}^\infty$ is *asymptotically efficient* for a parameter θ if $\sqrt{n} \cdot (W_n - \theta) \xrightarrow{\text{d}} \mathcal{N}(0, I(\theta)^{-1})$. That is, if the asymptotic variance of W_n achieves the Cramér-Rao lower bound.

Note here that the scaling $\{e_n\}_{n=1}^\infty$ or $1/\sqrt{n}$ is important. The asymptotic variance is *not* the limit of the variances of W_n, which will in general go to 0 even just to ensure consistency, as we saw earlier.

5 Asymptotics of the Likelihood Ratio Test

The asymptotics of the MLE also reveal how the likelihood ratio test statistic (LRT) behaves for large samples. Recall the LRT $\lambda(\mathbf{x})$ is

$$\lambda(\mathbf{x}) := \frac{\sup_{\theta \in \Theta_0} L(\theta|\mathbf{x})}{\sup_{\theta \in \Theta} L(\theta|\mathbf{x})},$$

where Θ_0 is the parameter set corresponding to our null hypothesis H_0.

Theorem 5.1 (asymptotics of the LRT)

For testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$, suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f(x|\theta)$, and that $f(x|\theta)$ satisfies all necessary regularity conditions for the asymptotic normality of the MLE. Then, under H_0

$$-2 \log \lambda(X_1, \ldots, X_n) \xrightarrow{\text{d}} \chi_1^2,$$

where χ_1^2 is a χ^2 random variable with 1 degree of freedom.

Proof. (idea) Again, let $\ell(\theta|\mathbf{x}) = \log L(\theta|\mathbf{x})$ be the log-likelihood and consider its Taylor expansion about $\hat{\theta}_{\text{MLE}}$:

$$\ell(\theta|\mathbf{x}) = \ell(\hat{\theta}_{\text{MLE}}|\mathbf{x}) + \ell'(\hat{\theta}_{\text{MLE}}|\mathbf{x}) \cdot (\theta - \hat{\theta}_{\text{MLE}}) + \ell''(\hat{\theta}_{\text{MLE}}|\mathbf{x}) \cdot \frac{(\theta - \hat{\theta}_{\text{MLE}})^2}{2} + \cdots .$$

For the simple null hypothesis $H_0 : \theta = \theta_0$, the LRT will be

$$\lambda(\mathbf{x}) = \frac{L(\theta_0|\mathbf{x})}{L(\hat{\theta}_{\text{MLE}}|\mathbf{x})}.$$

Thus,

$$-2 \log \lambda(\mathbf{x}) = -2\ell(\theta_0|\mathbf{x}) + 2\ell(\hat{\theta}_{\text{MLE}}|\mathbf{x}) \approx \ell''(\hat{\theta}_{\text{MLE}}|\mathbf{x}) \cdot (\theta_0 - \hat{\theta}_{\text{MLE}})^2 = -\ell''(\hat{\theta}_{\text{MLE}}|\mathbf{x})/n \cdot (\sqrt{n}(\theta_0 - \hat{\theta}_{\text{MLE}}))^2.$$

$(\sqrt{n}(\theta_0 - \hat{\theta}_{\text{MLE}})^2)$ goes to $I(\theta_0)^{-1}$ times a χ_1^2 random variable by the asymptotic normality of the MLE. Meanwhile, $-\ell''(\hat{\theta}_{\text{MLE}}|\mathbf{x})/n$ goes to $I(\theta_0)$ by the consistency of the MLE and a law of large numbers-like argument. Thus, by Slutsky, the above RHS goes to χ_1^2 in distribution. ■

Recall that the likelihood ratio test rejects $H_0 : \theta \in \Theta_0$ for small values of $\lambda(\mathbf{x})$. To ensure that the test is level α, we need to choose a rejection threshold c such that

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\lambda(\mathbf{X}) \leq c) \leq \alpha.$$

This can be difficult to do in general if the distribution of $\lambda(\mathbf{X})$ is complicated. A decent proxy is to find the critical values c of the asymptotic distribution of $\lambda(\mathbf{X})$. Our theorem gives us a nice and simple, parameter-free distribution, the χ_1^2, whose critical values we understand well.

Example 5.2 (Poisson LRT)

Suppose we wish to test $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$ based on observing $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim}$ Poisson(λ). The MLE in this case is $\hat{\lambda} := \overline{X}_n$. Then,

$$-2\log\lambda(X_1,\ldots,X_n) = -2\log\left(\frac{e^{-n\lambda_0}\lambda_0^{\sum_i X_i}}{e^{-n\hat{\lambda}}\hat{\lambda}^{\sum_i X_i}}\right) = 2n\left((\lambda_0 - \hat{\lambda}) - \hat{\lambda}\log(\lambda_0/\hat{\lambda})\right).$$

Now, since $-2\log\lambda(X_1,\ldots,X_n) \approx \chi_1^2$. We can determine an approximate critical value (that of the χ^2) by rejecting H_0 at level α if $-2\log\lambda(X_1,\ldots,X_n) > \chi_{1,\alpha}^2$.

6 Approximate Confidence Intervals

We've seen several examples thus far of two sequences of estimators $\{W_n\}_{n=1}^{\infty}, \{V_n\}_{n=1}^{\infty}$ satisfying

$$\frac{W_n - \theta}{V_n} \xrightarrow{\text{d}} \mathcal{N}(0,1).$$

Here, W_n could be an estimator for θ and V_n could be the sample-analogue of $\mathrm{Var}(W_n)$, with the right scaling. Similar to the previous section, this also gives us a way of forming approximate confidence intervals for θ. In particular, if we conflate $\frac{W_n-\theta}{V_n}$ and its limit (which is often justifiable for large samples), then we can say

$$W_n - z_{\alpha/2}\cdot V_n \leq \theta \leq W_n + z_{\alpha/2}\cdot V_n.$$

Thus, we obtain an approximate confidence interval with lower and upper bounds in terms of our sample and free of the parameter θ.

Example 6.1 (approximate C.I. from CLT)

If $X_1,\ldots,X_n$ are i.i.d. with mean μ and variance σ^2, then CLT gives us

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\text{d}} \mathcal{N}(0,1).$$

Moreover, thanks to Slutsky's theorem, we can substitute σ^2 for the sample variance S_n^2 giving

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \xrightarrow{\text{d}} \mathcal{N}(0,1).$$

This gives us the approximate $1-\alpha$ confidence interval

$$\overline{X}_n - z_{\alpha/2}\cdot S_n/\sqrt{n} \leq \mu \leq \overline{X}_n + z_{\alpha/2}\cdot S_n/\sqrt{n}.$$

Approximate C.I.'s for the MLE can be derived in the same manner using the asymptotic normality result from earlier. In general, if $X_1,\ldots,X_n \overset{\text{i.i.d.}}{\sim} f(x|\theta)$, then we have that the variance of $\hat{\theta}_{\text{MLE}}$ is approximated by

$$\widehat{\mathrm{Var}}(\hat{\theta}_{\text{MLE}}) := \frac{1}{-\frac{\partial^2}{\partial\theta^2}\log L(\theta|X_1,\ldots,X_n)|_{\theta=\hat{\theta}_{\text{MLE}}}}.$$

Note that the above is the sample-based analogue of the variance term from Cramer's theorem $\frac{1}{I(\theta)}$. Thus, by a law of large numbers-like argument, we should have $\widehat{\mathrm{Var}}(\hat{\theta}_{\text{MLE}})\cdot n \to I(\theta)^{-1}$ as $n\to\infty$. Then, Cramer's theorem gives us that

$$\frac{\hat{\theta}_{\text{MLE}} - \theta}{\sqrt{\widehat{\mathrm{Var}}(\hat{\theta}_{\text{MLE}})}} \xrightarrow{\text{d}} \mathcal{N}(0,1).$$

This gives us approximate $1-\alpha$ C.I.

$$\hat{\theta}_{\text{MLE}} - z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\hat{\theta}_{\text{MLE}})} \leq \theta \leq \hat{\theta}_{\text{MLE}} + z_{\alpha/2}\cdot\sqrt{\widehat{\mathrm{Var}}(\hat{\theta}_{\text{MLE}})}.$$

Note that it was necessary for us to approximate the asymptotic variance $I(\theta)^{-1}$ with a sample-based analogue free of θ since we want to easily isolate θ in forming the confidence interval bounds.

7 Problems

7.1 Previous Core Competency Problems

Problem 1 (2018 Summer Practice, # 2)**.** Suppose that $X_1, \dots, X_n$ are i.i.d. $\exp(1/\mu)$, where $\mathbb{E}(X_1) = \mu > 0$.

(i) Find the mean and variance of $\overline{X}_n = \sum_{i=1}^n X_i/n$. Hence, find the asymptotic distribution of $\overline{X}_n$ (properly standardized).

(ii) Let $T = \log \overline{X}_n$. Find the corresponding asymptotic distribution of T (properly standardized).

(iii) How can the asymptotic distribution of T be used to construct an approximate $(1-\alpha)$ confidence interval (CI) for μ? Explain your answer and give the desired CI.

Problem 2 (2018 Summer Practice, # 5)**.** Suppose that $Y_1, \dots, Y_n$ are i.i.d Poisson(λ), $\lambda > 0$ unknown. Assume that n is even, i.e., $n = 2k$ for some integer k. Consider

$$\hat{\lambda} = \frac{1}{2k} \sum_{i=1}^{k} (Y_{2i} - Y_{2i-1})^2.$$

(a) Is $\hat{\lambda}$ an unbiased estimator of λ (show your steps)?

(b) Is $\hat{\lambda}$ a consistent estimator of λ, as $k \to \infty$ (show your steps)?

Problem 3 (2018 September, # 6)**.** Suppose that $X_1, X_2, \dots, X_n$ are i.i.d. $N(\theta, 1)$, where $\theta \in \mathbb{R}$ is unknown. Let $\psi = \mathbb{P}_\theta(X_1 > 0)$.

(a) Find the maximum likelihood estimator $\hat{\psi}$ of ψ.

(b) Find an approximate 95% confidence interval for ψ.

(c) Let $Y_i = \mathbf{1}\{X_i > 0\}$, for $i = 1, \dots, n$. Define $\tilde{\psi} = (1/n) \sum_{i=1}^n Y_i$. Show that $\tilde{\psi}$ is a consistent estimator of ψ.

(d) Find the asymptotic distribution of both the estimators. Which estimator of ψ, $\hat{\psi}$ or $\tilde{\psi}$, is more preferable in this model and why?

Problem 4 (2019 May, # 3)**.** n_1 people are given treatment 1 and n_2 people are given treatment 2. Let X_1 be the number of people on treatment 1 who respond favorably to the treatment and let X_2 be the number of people on treatment 2 who respond favorably. Assume that $X_1 \sim$ Binomial(n_1, p_1), and $X_2 \sim$ Binomial(n_2, p_2). Let $\psi = p_1 - p_2$.

(i) Find the maximum likelihood estimator $\hat{\psi}$ of ψ.

(ii) Find the Fisher information matrix $I(p_1, p_2)$.

(iii) Use the delta method to find the asymptotic standard error of $\hat{\psi}$.

Problem 5 (2019 May, # 6)**.** Denote by $\hat{\zeta}_n$ the MLE of $\zeta = p(1-p)$ based on n i.i.d. samples from Binomial$(1, p)$. Denote by p_0 the true value of p.

(a) If $p_0 \neq 1/2$, find the limiting (non-degenerate) distribution of $\hat{\zeta}_n$, with proper normalization.

(b) Derive the asymptotic distribution of $\hat{\zeta}_n$, when $p_0 = 1/2$.

Problem 6 (2019 September, # 3)**.** Suppose that X_n and Y_n are independent random variables, where X_n is asymptotically normal with mean 4 and standard deviation $1/\sqrt{n}$ (i.e., $\sqrt{n}(X_n - 4) \xrightarrow{\text{d}} N(0,1)$) and Y_n is asymptotically normal with mean 3 and standard deviation $2/\sqrt{n}$. Use the delta method to get an approximate mean and standard deviation of Y_n/X_n.

Problem 7 (2019 September, # 5)**.** Let $X_1, \dots, X_n$ be the number of accidents at an important intersection in the past n years. We are interested in estimating the probability of zero accidents next year. We model the X_i's as independent random variables distributed according to a Poisson distribution with mean λ.

(i) Let $q(\lambda)$ be the probability that there will be no accidents next year. Find an unbiased and consistent estimator of $q(\lambda)$.

(ii) Compute the maximum likelihood estimator of $q(\lambda)$ and derive its asymptotic distribution. Compare this estimator with the one obtained in (i).

Problem 8 (2020 May, # 8)**.** Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p) random sample, i.e. $P(X_i = 1) = 1 - P(X_i = 0) = p$, $p \in (0, 1)$. Further, let $\theta = \mathrm{Var}(X_i)$.

(i) Find $\hat{\theta}$, the maximum likelihood estimator of θ.

(ii) Show that $\hat{\theta}$ is asymptotically normal when $p \neq 1/2$ and give the asymptotic variance.

(iii) When $p = 1/2$, derive a non-degenerated asymptotic distribution of $\hat{\theta}$ with an appropriate normalization. *Hint: try relating $\hat{\theta}$ to the statistic* $(\overline{X}_n - 1/2)^2$.

Problem 9 (2020 May, # 9)**.** Let $X_1, \ldots, X_{2n}$ be an i.i.d. random sample with common pdf $f(x) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}x}$ for $x > 0$. Consider the three estimators $\hat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^n X_i$, $\hat{\lambda}_2 = \frac{1}{n}\sum_{i=n+1}^{2n} X_i$, and $\hat{\lambda} = \frac{1}{2n}\sum_{i=1}^{2n} X_i$.

(i) Show that $T_1 = \hat{\lambda}_1\hat{\lambda}_2$ is an unbiased and consistent estimator of λ^2.

(ii) Show that $T_2 = \hat{\lambda}^2$ is consistent for λ^2, but not unbiased.

(iii) Derive the asymptotic distribution of the estimators T_1 and T_2. Which one is more efficient asymptotically?

Problem 10 (2020 September, # 6)**.** Suppose $(X_1, \ldots, X_n)$ are i.i.d. from a Normal distribution with $\mathbb{E}X_i = \mathrm{Var}(X_i) = \theta$, where $\theta > 0$ is unknown.

(a) Find the MLE for θ explicitly.

(b) Find the asymptotic distribution of your MLE.

Problem 11 (2021 May, # 3)**.** A random sample $X_1, \ldots, X_n$ is drawn from a population with p.d.f.

$$f_\theta(x) = \frac{1}{2}(1 + \theta x), x \in [-1, 1],$$

and $f_\theta(x) = 0$ if $x \notin [-1, 1]$, where $\theta \in [-1, 1]$ is the unknown parameter.

1. Find an unbiased estimator of θ.
2. Is the estimator in (i) consistent? Provide a justification for your answer.

Problem 12 (2021 May, # 5)**.** Let X and Y be a pair of random variables with the following distributional specification. $P(Y = 1) = 1 - P(Y = 0) = \alpha$ where $\alpha \in (0, 1)$ and $X|Y = 0 \sim N(0, \sigma^2)$ and $X|Y = 1 \sim N(\mu, \sigma^2)$.

1. Find the conditional distribution of Y given X, i.e. $P(Y = 1|X = x)$.
2. Suppose that we have an i.i.d. random sample from this population, i.e. we observe i.i.d. copies (X_i, Y_i), $i = 1, \ldots, n$. Write down the likelihood function and find maximum likelihood estimators $\hat{\alpha}_n, \hat{\mu}_n$ and $\hat{\sigma}_n^2$ of α, μ, and σ^2.
3. What are the asymptotic distributions of $\hat{\alpha}_n, \hat{\mu}_n$, and $\hat{\sigma}_n^2$ (properly standardized)?

Problem 13 (2021 May, # 6)**.** Suppose $X_1, \ldots, X_n$ are independent, with $X_i \sim N\left(\frac{\theta}{i}, 1\right)$. Here, $\theta \in \mathbb{R}$ is an unknown parameter.

(i) Find an unbiased estimator $\hat{\theta}_n$ for θ which depends on the entire data.

(ii) Find asymptotic non-degenerate distribution of your estimator, i.e. $d_n(\hat{\theta}_n - \theta)$ converges to a non-degenerate distribution.

(iii) Suppose that we impose a normal prior $\theta \sim N(0, \tau)$, where $\tau > 0$ is an known constant. Find the posterior distribution of θ given data $X_1, \ldots, X_n$.

Problem 14 (2021 September, # 4)**.** Suppose $\{U_i\}_{i\geq 1} \overset{\text{i.i.d.}}{\sim} U(0, \theta)$, for some $\theta > 0$.

1. Show that $T_n := \left(\prod_{i=1}^n U_i\right)^{1/n}$ converges in probability to a constant, and find this constant.
2. Find a function of T_n that is a consistent estimator for θ.
3. Find constants a_n and b_n such that $a_n(T_n - b_n)$ converges to a non-degenerate distribution.