Maximum Likelihood Tests

A. Stepanov
D. Simpson
S. Culpepper

Recall that the likelihood and log-likelihood functions are,

$$L(\theta;\mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta), \qquad \ell(\theta;\mathbf{x}) = \sum_{i=1}^{n} \ln[f(x_i;\theta)]$$

$$\hat{\theta} = Argmax\, L(\theta;\mathbf{x})$$

We are often interested in two-sided hypotheses,

$$H_0\colon \theta = \theta_0 \quad vs. \quad H_1\colon \theta \neq \theta_0$$

Likelihood Ratio Test (LRT): $L(\theta;\mathbf{x})$ is maximized at $\hat{\theta}$, so it intuitively suggests that we should Reject H_0 in favor of H_1 if,

$$\Lambda = \frac{L(\theta_0;\mathbf{x})}{L(\hat{\theta};\mathbf{x})} \leq c$$

where c is such that $\alpha = P_{\theta_0}(\Lambda \leq c)$.

Example 1. Let $Y_1 < \cdots < Y_n$ be the order statistics of a random sample of size n from a $U(0,\theta)$ distribution for $\theta > 0$. $H_0\colon \theta = \theta_0 \quad vs. \quad H_1\colon \theta \neq \theta_0$

Recall $\hat{\theta} = Y_n$ and $F_{Y_n}(x) = (x/\theta)^n, 0 < x < \theta$.

$$\Lambda = \frac{\left(\frac{1}{\theta_0}\right)^n I(Y_n < \theta_0)}{\left(\frac{1}{Y_n}\right)^n} = \left(\frac{Y_n}{\theta_0}\right)^n I(Y_n < \theta_0)$$

$$\Lambda \leq k \quad \Leftrightarrow \quad Y_n \leq c \;\; or \;\; Y_n \geq \theta_0$$

Reject H_0 if $Y_n \leq c \;\; or \;\; Y_n \geq \theta_0$. If $Y_n \leq c$, $\alpha = P_{\theta_0}(Y_n \leq c)$,

Suppose we use an α Type I error rate.

$$\alpha = P_{\theta_0}(Y_n \leq c) = \left(\frac{c}{\theta_0}\right)^n$$
$$\Rightarrow c = \theta_0 \alpha^{\frac{1}{n}}$$

Example 2. Let $X_1, \dots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$ distribution (σ^2 known).

$$H_0: \mu = \mu_0 \quad vs. \quad H_1: \mu \neq \mu_0$$

Recall,

$$\begin{aligned} L(\mu; \mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu_0)^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2\right]\right\} \end{aligned}$$

$$\hat{\mu} = \bar{X} \sim N(\mu, \sigma^2/n)$$

$$\begin{aligned} \Lambda = \frac{L(\mu_0; \mathbf{x})}{L(\hat{\mu}; \mathbf{x})} &= \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2\right]\right\}}{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}} \\ &= \exp\left\{-\frac{n(\bar{x} - \mu_0)^2}{2\sigma^2}\right\} \end{aligned}$$

$$\Lambda \leq k \quad \Leftrightarrow \quad -2\ln\Lambda = \frac{n(\bar{x} - \mu_0)^2}{\sigma^2} \geq c$$

$$\frac{n(\bar{x} - \mu_0)^2}{\sigma^2} \sim \chi^2(1)$$

Reject H_0 if $\frac{n(\bar{x} - \mu_0)^2}{\sigma^2} \geq c = \chi^2_\alpha(1)$.

Maximum Likelihood Tests

A. Stepanov
D. Simpson
S. Culpepper

Theorem 6.3.1. Assume regularity conditions (R0) to (R5). Under the null hypothesis, $H_0\colon \theta = \theta_0$,

$$\chi_L^2 = -2 \ln \Lambda \xrightarrow{D} \chi^2(1)$$

Proof. A Taylor series of $\ell(\hat{\theta}_n; \mathbf{x})$ about the true value under the null, θ_0, yields,

$$\ell(\hat{\theta}_n; \mathbf{x}) = \ell(\theta_0; \mathbf{x}) + (\hat{\theta}_n - \theta_0)\ell'(\theta_0; \mathbf{x}) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2 \ell''(\theta_n^*; \mathbf{x})$$

for θ_n^* between $\hat{\theta}_n$ and θ_0 such that $\hat{\theta}_n \xrightarrow{P} \theta_0 \Rightarrow \theta_n^* \xrightarrow{P} \theta_0$.

$$-\frac{1}{n}\ell''(\theta_0; \mathbf{x}) \xrightarrow{P} I(\theta_0)$$

$$\frac{1}{\sqrt{n}}\ell'(\theta_0; \mathbf{x}) = \sqrt{n}(\hat{\theta}_n - \theta_0)I(\theta_0) + R_n$$

where $R_n \xrightarrow{P} 0$. Rearranging the expansion for $\ell(\hat{\theta}_n; \mathbf{x})$ yields,

$$\ell(\hat{\theta}_n; \mathbf{x}) - \ell(\theta_0; \mathbf{x}) = (\hat{\theta}_n - \theta_0)^2 nI(\theta_0) - \frac{1}{2}(\hat{\theta}_n - \theta_0)^2 nI(\theta_0) + (\hat{\theta}_n - \theta_0)R_n$$

$$\chi_L^2 = -2 \ln \Lambda = (\hat{\theta}_n - \theta_0)^2 nI(\theta_0) + R_n^*$$

where $R_n^* \xrightarrow{P} 0$. ■

Theorem 6.3.1. implies that, for large n, we can Reject H_0 if $\chi_L^2 > \chi_\alpha^2(1)$.

Wald Test

Theorem 6.2.2 implies that,

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{D} N\left(0, \frac{1}{I(\theta_0)}\right)$$

So,

$$\sqrt{nI(\theta_0)}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{D} N(0,1)$$

We can test

$$H_0: \theta = \theta_0 \quad vs. \quad H_1: \theta \neq \theta_0$$

with

$$\chi_W^2 = nI(\theta_0)\left(\hat{\theta}_n - \theta_0\right)^2 \xrightarrow{D} \chi^2(1)$$

Reject H_0 if $\chi_W^2 > \chi_\alpha^2(1)$.

Rao's Score Test

Recall that **Theorem 6.1.1** implies under assumptions (R0) and (R1),

$$\lim_{n\to\infty} P[L(\theta_0; \mathbf{x}) > L(\theta; \mathbf{x})] = 1 \;\forall\; \theta_0 \neq \theta.$$

Asymptotically the likelihood function is maximized at the true value θ_0.

$$\frac{1}{\sqrt{n}} \ell'(\theta_0; \mathbf{x}) \xrightarrow{D} N\left(0, I(\theta_0)\right)$$

Define the statistic,

$$\chi_R^2 = \frac{[\ell'(\theta_0; \mathbf{x})]^2}{nI(\theta_0)}$$

Reject H_0 if $\chi_R^2 > \chi_\alpha^2(1)$.

Example 3. Let $X_1, \dots, X_n$ be a random sample of size n from the distribution with probability density function,

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} x^{\frac{1-\theta}{\theta}}, & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$$

$$H_0: \theta = \theta_0 \quad vs. \quad H_1: \theta \neq \theta_0$$

a) Find Λ.

$$L(\theta;\mathbf{x}) = \frac{1}{\theta^n}\left[\prod_{i=1}^{n} x_i\right]^{\frac{1-\theta}{\theta}} \quad \Rightarrow \quad \hat{\theta} = -\frac{1}{n}\sum_{i=1}^{n} \ln x_i = \bar{y}$$

$$\Lambda = \frac{L(\theta_0;\mathbf{x})}{L(\hat{\theta};\mathbf{x})} = \frac{\frac{1}{\theta_0^n}\left[\prod_{i=1}^{n} x_i\right]^{\frac{1-\theta_0}{\theta_0}}}{\frac{1}{\bar{y}^n}\left[\prod_{i=1}^{n} x_i\right]^{\frac{1-\bar{y}}{\bar{y}}}} = \left(\frac{\bar{y}}{\theta_0}\right)^n \left[\prod_{i=1}^{n} x_i\right]^{\frac{1-\theta_0}{\theta_0} - \frac{1-\bar{y}}{\bar{y}}}$$

$$= \left(\frac{\bar{y}}{\theta_0}\right)^n \left(e^{-n\bar{y}}\right)^{\frac{1}{\theta_0} - \frac{1}{\bar{y}}} = e^n \left(\frac{\bar{y}}{\theta_0}\right)^n \exp\left[-n\frac{\bar{y}}{\theta_0}\right]$$

Note that because Λ maximized at $\bar{y} = \frac{2}{\lambda_0}$ that rejecting H_0 if $\Lambda \le c$ is equivalent to rejecting H_0 if $\bar{y} \le c_1$ and $\bar{y} \ge c_2$.

Recall that,

$$\bar{y} \sim Gamma\left(\alpha = n, "\theta" = \frac{\theta}{n}\right) \Rightarrow \frac{2n\bar{y}}{\theta} \sim \chi^2(2n)$$

So, reject H_0 if,

$$\frac{2n\bar{y}}{\theta_0} \le \chi^2_{1-\frac{\alpha}{2}}(2n) \;\; or \;\; \frac{2n\bar{y}}{\theta_0} \ge \chi^2_{\alpha/2}(2n)$$

b) Suppose $H_0: \theta = \frac{1}{2}$ $vs.$ $H_1: \theta \neq \frac{1}{2}$, $\bar{y} = 1$, $n = 15$. Use the large sample LRT to test the hypothesis.

$$-2\ln\Lambda = -2\left[n + n\ln\left(\frac{\bar{y}}{\theta_0}\right) - n\frac{\bar{y}}{\theta_0}\right] = 2n\left[\frac{\bar{y}}{\theta_0} - \ln\left(\frac{\bar{y}}{\theta_0}\right) - 1\right]$$
$$= 30[2 - \ln(2) - 1] \approx 9.21 > \chi^2_{.05}(1) = 3.84$$

Reject H_0.

c) Suppose $H_0: \theta = \frac{1}{2}$ $vs.$ $H_1: \theta \neq \frac{1}{2}$, $\bar{y} = 1$, $n = 15$. Use the large sample Wald statistic to test the hypothesis.

Recall that $I(\theta) = 1/\theta^2$.

$$\chi^2_W = nI(\hat{\theta}_n)(\hat{\theta}_n - \theta_0)^2 = 15\left(1 - \frac{1}{2}\right)^2 = \frac{15}{4} > \chi^2_{.05}(1) = 3.84$$

Do not reject H_0.

d) Suppose $H_0: \theta = \frac{1}{2}$ $vs.$ $H_1: \theta \neq \frac{1}{2}$, $\bar{y} = 1$, $n = 15$. Use the large sample Rao score statistic to test the hypothesis.

$$\ell'(\theta_0; \mathbf{x}) = -\frac{n}{\theta_0} - \frac{1}{\theta_0^2}\sum_{i=1}^{n} \ln x_i = -\frac{n}{\theta_0} + n\frac{\bar{y}}{\theta_0^2} = -30 + 60 = 30$$

$$\chi^2_R = \frac{[\ell'(\theta_0; \mathbf{x})]^2}{nI(\theta_0)} = \frac{[30]^2}{60} = 15 > \chi^2_{.05}(1) = 3.84$$

Reject H_0.

Example 4. Let $\lambda > 0$ and let $X_1, \dots, X_n$ be a random sample from the distribution with the probability density function,

$$f(x;\theta) = 2\lambda^2 x^3 e^{-\lambda x^2}, \qquad x > 0$$

a) Find Λ.

$$L(\lambda;\mathbf{x}) = 2^n \lambda^{2n} \exp\left(-\lambda \sum_{i=1}^n x_i^2\right)\left(\prod_{i=1}^n x_i\right)^3 \quad \Rightarrow \quad \hat{\lambda} = \frac{2n}{\sum_{i=1}^n x_i^2} = \frac{2}{\bar{y}}.$$

$$\Lambda = \frac{L(\lambda_0;\mathbf{x})}{L(\hat{\lambda};\mathbf{x})} = \frac{2^n \lambda_0^{2n} \exp\left(-\lambda_0 \sum_{i=1}^n x_i^2\right)\left(\prod_{i=1}^n x_i\right)^3}{2^n \left(\frac{2}{\bar{y}}\right)^{2n} \exp(-2n)\left(\prod_{i=1}^n x_i\right)^3}$$

$$= e^{2n}\left(\frac{\bar{y}\lambda_0}{2}\right)^{2n} \exp(-n\lambda_0\bar{y}) \le c$$

Note that because Λ maximized at $\bar{y} = \frac{2}{\lambda_0}$ that rejecting H_0 if $\Lambda \le c$ is equivalent to rejecting H_0 if $\bar{y} \le c_1$ and $\bar{y} \ge c_2$. Recall that,

$$\bar{y} \sim Gamma(\alpha = 2n, \theta = 1/n\lambda) \Rightarrow 2n\lambda\bar{y} \sim \chi^2(4n)$$

So, reject H_0 if $2n\lambda_0\bar{y} \le \chi^2_{1-\alpha/2}(4n)$ or $2n\lambda_0\bar{y} \ge \chi^2_{\alpha/2}(4n)$.

Suppose $H_0: \lambda = 1 \ \ vs. \ H_1: \lambda \ne 1, \bar{y} = 1, n = 10$.

$$2n\lambda_0\bar{y} = 20, \qquad \chi^2_{.975}(40) \approx 24.3, \qquad \chi^2_{.025}(40) = 59.34$$

Reject H_0.

b) Suppose $H_0: \lambda = 1 \;\; vs. \;\; H_1: \lambda \neq 1, \bar{y} = 1, n = 10$. Use the large sample LRT to test the hypothesis.

$$-2\ln\Lambda = -2\left[2n + 2n\ln\left(\frac{\bar{y}\lambda_0}{2}\right) - n\lambda_0\bar{y}\right] = 2n\left[\lambda_0\bar{y} - 2 - 2\ln\left(\frac{\bar{y}\lambda_0}{2}\right)\right]$$
$$= 20\left[1 - 2 - \ln\left(\frac{1}{4}\right)\right] \approx 7.73 > \chi^2_{.05}(1) = 3.84$$

Reject H_0.

c) Suppose $H_0: \lambda = 1 \;\; vs. \;\; H_1: \lambda \neq 1, \bar{y} = 1, n = 10$. Use the large sample Wald statistic to test the hypothesis.

Recall that

$$I(\lambda) = \frac{2}{\lambda^2}.$$

$$\chi^2_W = nI(\hat{\theta}_n)(\hat{\theta}_n - \theta_0)^2 = 10 \cdot \frac{2}{2^2}\left(\frac{2}{1} - 1\right)^2 = \frac{10}{4} < \chi^2_{.05}(1) = 3.84$$

Do not reject H_0.

d) Suppose $H_0: \lambda = 1 \;\; vs. \;\; H_1: \lambda \neq 1, \bar{y} = 1, n = 10$. Use the large sample Rao score statistic to test the hypothesis.

$$\ell(\lambda; \mathbf{x}) = n\ln(2) + 2n\ln(\lambda) - \lambda\sum_{i=1}^{n} x_i^2 + 3\sum_{i=1}^{n}\ln(x_i).$$
$$\ell'(\lambda; \mathbf{x}) = \frac{2n}{\lambda} - \sum_{i=1}^{n} x_i^2 \Rightarrow \ell'(1; \bar{y} = 1) = 10$$

$$\chi^2_R = \frac{[\ell'(\theta_0; \mathbf{x})]^2}{nI(\theta_0)} = \frac{[10]^2}{20} = 5 > \chi^2_{.05}(1) = 3.84$$

Reject H_0.

A. Stepanov
D. Simpson
S. Culpepper

Testing between two simple hypotheses using the likelihood ratio

Example 5. Suppose $X{\sim}Exp(\theta)$ and consider the somewhat artificial situation where there are only two possible values for θ: $\theta = 1$ or $\theta = A \gg 1$. Based on one observation we wish to test the null hypothesis

$$H_0\!: X{\sim}Exp(1)$$

versus the alternative

$$H_1\!: X{\sim}Exp(A)$$

It turns out the most powerful test is based on the likelihood ratio:

$$\Lambda(\mathrm{x}) = \frac{f_0(x)}{f_1(x)} = \frac{e^{-x}}{\frac{1}{A}e^{-\frac{x}{A}}} = Ae^{-\frac{A-1}{A}x}$$

For a level-α test we set a cutoff value c for the statistic such that the probability of falsely rejecting $H_0 = \alpha$. To do this solve for c in:

$$\alpha = P_0(\Lambda \le c) = P_0\left(Ae^{-\frac{A-1}{A}X} \le c\right)$$
$$= P_0\left(X \ge \frac{A}{A-1}\ln\left(\frac{A}{c}\right)\right)$$

$$= \exp\left(-\frac{A}{A-1}\ln\left(\frac{A}{c}\right)\right) = \left(\frac{c}{A}\right)^{\frac{A}{A-1}}$$

Solving for *c* gives

$$c = A\alpha^{\frac{A-1}{A}}$$

For example, *A*=10 and $\alpha = 0.05$ gives the cutoff value *c* = 0.6746.

If we were to observe, say, *X* = 3.2, then $\Lambda(\mathrm{x}) = 0.56$ and we would reject H_0. If instead we observed *X* = 2.7, then $\Lambda(\mathrm{x}) = 0.88$ and we would fail to reject H_0 at level 0.05.

What is the power of this test? Compute

$$P_1(\Lambda \leq c) = P_1\left(X \geq \frac{A}{A-1}\ln\left(\frac{A}{c}\right)\right)$$

$$= \exp\left(-\frac{1}{A}\,\frac{A}{A-1}\ln\left(\frac{A}{c}\right)\right)$$

$$= \exp\left(\frac{1}{A-1}\ln\left(\frac{c}{A}\right)\right) = \left(\frac{c}{A}\right)^{\frac{1}{A-1}}$$

Now insert the value of *c* found above for an α-level test to get:

$$P_1(\Lambda \leq c) = \left(\frac{A\alpha^{\frac{A-1}{A}}}{A}\right)^{\frac{1}{A-1}} = \alpha^{\frac{1}{A}}$$

Note that we can see directly the tradeoff between the false rejection rate α and the power (probability of correctly rejecting) $\alpha^{\frac{1}{A}}$. Make α too small and we lose power, demand too much power and we increase the false rejection rate too much, so we try to find a happy intermediate.

In the example, if the alternative $A = 10$ and α=0.05, then the power of the test is

$$0.05^{0.10} = 0.741.$$