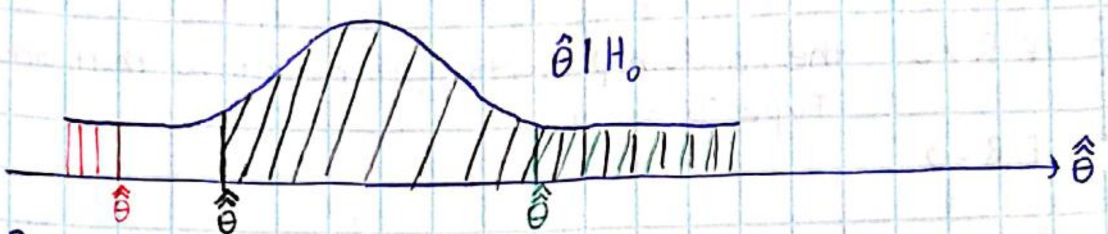
Lecture 19 11/18/2020

Here's a relevant fact about p-val that's important in our discussion about multiple comparisons. If H_0 is true, what is the distribution of the p-value?



Consider this as right-tail test

Proof for why p-vals under the H_0 are realizations from an uniform; $U(0,1)$ distribution. Assume left-sided test. The proof for right-sided and two sided is similar.

$$P_{value} := F_{\hat{\theta}|H_0}(\hat{\theta})$$

↑ R.V. model for p-val's

Let's examine the CDF of P_{val} to try and figure out it's distribution. This is a proof from 368.

$$F_{P_{val}}(p_{val}) = P(P_{val} \le p_{val}) = P(F_{\hat{\theta}|H_0}(\hat{\theta}) \le p_{val}) = P(\hat{\theta} \le F^{-1}_{\hat{\theta}|H_0}(p_{val}))$$

↑ assume $\hat{\theta}|H_0$ is constant

$$= F_{\hat{\theta}|H_0}(F^{-1}_{\hat{\theta}|H_0}(p_{val})) = p_{val} \Rightarrow P_{val} \sim U(0,1).$$

We will return to testing now. We previously proved...

$$\frac{\hat{\theta}^{MLE} - \theta}{\sqrt{\frac{I(\theta)^{-1}}{n}}} \xrightarrow{d} N(0,1) \xrightarrow{\text{Wald Test}} \text{for } H_a: \theta \neq \theta_0,\ RET \approx \left[\theta_0 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{I(\theta)^{-1}}{n}}\right]$$

⇓ Wald CI via Richardization

$$CI_{\theta, 1-\alpha} \approx \left[\hat{\hat{\theta}}^{MLE} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{I(\theta)^{-1}}{n}}\right]$$

We'll now derive a related means of testing

$$H_a: \theta \neq \theta_0$$

Recall for an i.i.d. DGP,

$$S(\theta; X_1, \ldots, X_n) = \sum_{i=1}^{n} \underbrace{\ell'(\theta; X_i)}_{W_i} \qquad \text{def 8, lec 9}$$

$$\frac{S(\theta; X_1, \ldots, X_n)}{\sqrt{n\, I(\theta)}} \xrightarrow{d} N(0,1)$$

$$\Rightarrow \frac{1}{n} S(\theta; X_1, \ldots, X_n) = \overline{W} \qquad \Updownarrow$$

$E[W_i] = 0$ Fact 1b, lec 9

$Var[W_i] = I(\theta)$ Lec 9-10

$$\Rightarrow \frac{\frac{1}{n} S(\theta; X_1, \ldots, X_n)}{\sqrt{\frac{I(\theta)}{n}}} = \frac{\overline{W} - E[\bar{X}]}{SE[\overline{W}]} \xrightarrow{d} N(0,1)$$

$$\Rightarrow \frac{S(\theta_0; X_1, \ldots, X_n)}{\sqrt{n\, I(\theta)}} \,\dot{\sim}\, N(0,1)$$

Using this as a z-test statistic was discovered by Rao in 1948 and is called the "score test" but others called the "Lagrange multiplier test"

At $\alpha = 5\%$

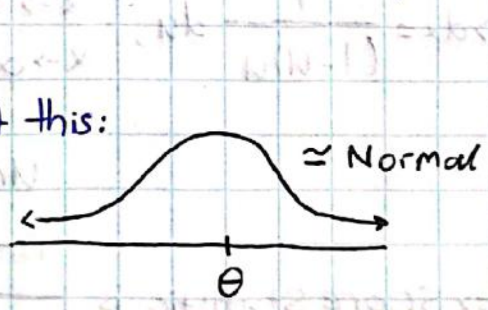
$$\Rightarrow \frac{S(\theta_0; X_1, \ldots, X_n)}{\sqrt{n\, I(\theta_0)}} \in [-1.96, 1.96] \Rightarrow \text{Retain } H_0$$

Note: this is "one-dimensional". There's only one θ being tested. You can derive the generalization with multiple θ's but we won't in this class.

This test statistic is really strange. Where is the estimator $\hat{\theta}$? You usually find an estimate that gauges the departure from H_0, and you find / approximate its distribution (the sampling distribution) and then check if $\hat{\hat{\theta}}$ looks weird. If so, reject. But we won't do that here. The estimator is not in the expression! And if you just want to test $H_a: \theta \neq \theta_0$, you don't really need an estimator or an estimate.

Many times, it is the same as the Wald test when you actually algebraically solve for the test statistic

Here's an example why you may care about this:



DGP: i.i.d. Logistic $(\theta, 1) := \dfrac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}$

$$\mathcal{L} = \prod_{i=1}^{n} \frac{e^{-x_i} e^{\theta}}{(1+e^{-x_i} e^{\theta})^2} = \frac{e^{-\sum x_i} e^{n\theta}}{\prod_{i=1}^{n} (1+e^{-x_i} e^{\theta})^2}$$

$$\ell = -\sum X_i + n\theta - 2\sum \ln(1+e^{-X_i}e^{\theta})$$

Logistic $(\theta,1)$ $= \dfrac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}$

$$S = \ell' = n - 2\sum \frac{e^{-X_i}e^{\theta}}{1+e^{-X_i}e^{\theta}}$$

To get the MLE I set the above equal to zero and solve for θ. It's not possible in closed form.

$$\ell'(\theta;x) = 1 - 2\,\frac{e^{-x}e^{\theta}}{1+e^{-x}e^{\theta}}, \quad -\ell''(\theta;x) = 2\,\frac{(1+e^{-x}e^{\theta})e^{-x}e^{\theta} - (e^{-x}e^{\theta})^2}{(1+e^{-x}e^{\theta})^2}$$

$$= 2\,\frac{e^{-x}e^{\theta}}{(1+e^{-x}e^{\theta})^2}$$

$$I(\theta) = E[-\ell''(\theta;X)] = E\left[2\,\frac{e^{-X}e^{\theta}}{(1+e^{-X}e^{\theta})^2}\right] = 2\int_{\mathbb{R}} \frac{e^{-x}e^{\theta}}{(1+e^{-x}e^{\theta})^2}\,\overset{\text{Logistic}(\theta,1)}{f_X(x)}\,dx$$

$$= 2\int_{\mathbb{R}} \frac{e^{-x}e^{\theta}}{(1+e^{-x}e^{\theta})^2}\,\frac{e^{-x}e^{\theta}}{(1+e^{-x}e^{\theta})^2}\,dx$$

$$= 2\int_{\mathbb{R}} \frac{(e^{-x}e^{\theta})^2}{(1+e^{-x}e^{\theta})^4}\,dx = 2\cdot\frac{1}{6} = \boxed{\frac{1}{3}}$$

Integration calculus:

$$\int_{\mathbb{R}} \left(\frac{1}{1+e^{-x}e^{\theta}}\right)^2 \left(\frac{e^{-x}e^{\theta}}{1+e^{-x}e^{\theta}}\right)^2 dx = \int_0^1 u^{\cancel{2}}(1-u)^{\cancel{2}}\,\frac{1}{\cancel{(1-u)}\cancel{u}}\,du = \int_0^1 (u-u^2)\,du$$

$$= \left[\frac{u^2}{2} - \frac{u^3}{3}\right]_0^1 = \frac{1}{6}$$

Let $u = \dfrac{1}{1+e^{-x}e^{\theta}} \Rightarrow 1-u = \dfrac{e^{-x}e^{\theta}}{1+e^{-x}e^{\theta}} \Rightarrow \dfrac{du}{dx} = \cancel{-}(1+e^{-x}e^{\theta})^{-2}(+e^{-x}e^{\theta})$

$$= \frac{e^{-x}e^{\theta}}{1+e^{-x}e^{\theta}}\;\frac{1}{1+e^{-x}e^{\theta}}$$

$$= (1-u)u$$

$\Rightarrow dx = \dfrac{1}{(1-u)u}\,du$, $\quad x \to -\infty,\ u=0$; $\quad x\to\infty,\ u=1$

Under $H_0: \theta = \theta_0$

$\Rightarrow$ Score statistic is $\dfrac{n - 2\sum \dfrac{e^{-X_i}e^{\theta_0}}{1+e^{-X_i}e^{\theta_0}}}{\sqrt{n/3}} \;\dot\sim\; N(0,1)$

R-studio

In our data example, we get $\dfrac{10 - 2(0.646)}{\sqrt{\frac{10}{3}}} = 4.77 \notin [-1.96, 1.96]$

$\Rightarrow$ Reject H_0

Here's another also related testing procedure to the Wald and Score. Here too we wish to test against $H_0 : \theta = \theta_0$. Remember, we want an estimate that gauges departure from this. How about...

$$\hat{LR} := \frac{\mathcal{L}(\hat{\theta}^{MLE}; x_1, \dots, x_n)}{\mathcal{L}(\theta_0; x_1, \dots, x_n)} \overset{\text{i.i.d. DGP}}{=} \frac{\prod_{i=1}^{n} \mathcal{L}(\hat{\theta}^{MLE}; x_i)}{\prod_{i=1}^{n} \mathcal{L}(\theta_0; x_i)} = \prod_{i=1}^{n} \frac{\mathcal{L}(\hat{\theta}^{MLE}; x_i)}{\mathcal{L}(\theta_0; x_i)}$$

Likelihood Ratio. If it's significantly > 1, then we reject H_0.
Now we just need $\hat{LR}$, the sampling distr. You can prove that:

capitalize greek letter lambda $\rightarrow \hat{\Lambda} : 2 \ln(\hat{LR}) \xrightarrow{d} \chi^2_1$ Recall $F_{\chi^2_1}(3.84) = 95\%$

E.g. i.i.d. Bern(θ), $H_a : \theta \neq \theta_0$.

$$\hat{LR} = \prod_{i=1}^{n} \frac{\mathcal{L}(\bar{x}; x_i)}{\mathcal{L}(\theta_0; x_i)} = \prod_{i=1}^{n} \frac{\bar{x}^{x_i}(1-\bar{x})^{1-x_i}}{\theta_0^{x_i}(1-\theta_0)^{1-x_i}} = \left(\frac{\bar{x}}{\theta_0}\right)^{\sum x_i} \left(\frac{1-\bar{x}}{1-\theta_0}\right)^{n - \sum x_i}$$

$$\hat{\Lambda} = 2\left(\sum x_i \ln\left(\frac{\bar{x}}{\theta_0}\right) + \left(n - \sum x_i\right) \ln\left(\frac{1-\bar{x}}{1-\theta_0}\right)\right) = 2\underbrace{\left(O_1 \ln\left(\frac{O_1}{E_1}\right) + O_2 \ln\left(\frac{O_2}{E_2}\right)\right)}_{\text{Discrete KL-divergence (Back in Lec 2)}}$$

Let O_1 := # ones
O_2 := # zeroes
E_1 := # expected ones $(n\theta_0)$
E_2 := # expected zeroes $n(1-\theta_0)$