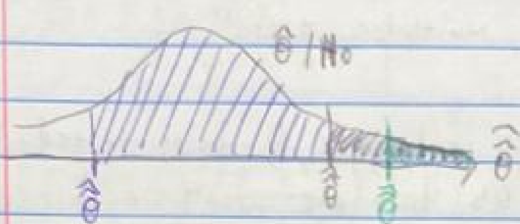
Lecture 19 11/18/20

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



Proof for why p-values under the null hypothesis are realizations from a $U(0,1)$ distribution. Assume left-sided test. The proof for right-sides and two-sided is similar.

$P_{val} := F_{\hat{\theta}/H_0}(\hat{\theta})$

r.v. model for Pval's

Let's examine the CDF of P-val to try and figure out its distribution. This is a proof from 365.

assume $\hat{\theta}/H_0$ is continuous

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

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

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 Test via Continuous Mapping & Slutsky's (Rxhandize)

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

We'll now derive a related means of testing $H_a: \theta \neq \theta_0$

Recall for an iid DGP,

$$S(\theta; x_1, \ldots, x_n) = \sum_{i=1}^{n} \underbrace{l'(\theta; x_i)}_{w_i} \quad \text{def 8, lec 9.}$$

$$\Rightarrow \frac{1}{n} S(\theta; x_1, \ldots, x_n) = \bar{w}$$

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

$Var[w_i] = I(\theta)$ lec 9-10

$$\Rightarrow \frac{\frac{1}{n} S(\theta; x_1, \ldots, x_n)}{\sqrt{\frac{I(\theta)}{n}}} = \frac{\bar{w} - E[\bar{X}]}{SE[\bar{w}]} \xrightarrow{d} N(0,1)$$

$$= \frac{S(\theta; x_1, \ldots, x_n)}{\sqrt{n I(\theta)}} \xrightarrow{d} N(0,1)$$

$$\Rightarrow \frac{S(\theta_0; x_1, \dots, x_n)}{\sqrt{n\, I(\theta_0)}} \sim N(0,1)$$

At $\alpha = 5\%$

$$\Rightarrow \frac{S(\theta_0; x_1, \dots, x_n)}{\sqrt{n\, I(\theta)}} \in [-1.96, 1.96] \Rightarrow \text{RET } H_0$$

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

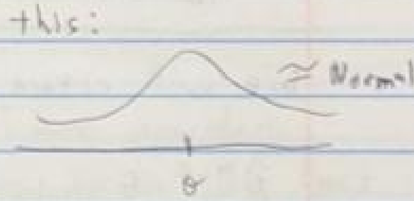
Note! this is "one-dimensional." There's only one θ being tested. You can derive the generalization with multiple θ's but we won't in this class.

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 $\hat{\theta}$ looks weird. If so, reject. But we don'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 (HW you'll do it for Bern).

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

DGP: iid Logistic $(\theta, 1) := \frac{e^{-(x-\theta)}}{\left(1 + e^{-(x-\theta)}\right)^2}$



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

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

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

To get the MLE I set the above equal to zero and solve for θ. Good luck! It's not possible in closed form. You can use a computer to do a numerical solve if you wish.

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

$$\Downarrow$$

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

$$I(\theta) = E\left[2\frac{e^{-x}e^{\theta}}{(1+e^{-x}e^{\theta})^2}\right] = 2\int_{\mathbb{R}} \frac{e^{-x}e^{\theta}}{(1+e^{-x}e^{\theta})^2} f_x(x)\,dx = 2\int_{\mathbb{R}} \frac{(e^{-x}e^{\theta})^2}{(1+e^{-x}e^{\theta})^4}\,dx = 2\cdot\frac{1}{6} = \frac{1}{3}$$

$$\int_{\mathbb{R}} \left(\frac{1}{1+e^{-x}e^{\theta}}\right)^2\left(\frac{e^{-x}e^{\theta}}{1+e^{-x}e^{\theta}}\right)^2 dx = \int_0^1 u^{\cancel{2}}(1-u)^{\cancel{2}}\frac{1}{\cancel{(1-u)}\cancel{u}}\,du = \int_0^1 (u-u^2)\,du = \left[\frac{u^2}{2} - \frac{u^3}{3}\right]_0^1 = \frac{1}{6}$$

let $u = \frac{1}{1+e^{-x}e^{\theta}} \Rightarrow 1-u = \frac{e^{-x}e^{\theta}}{1+e^{-x}e^{\theta}} \Rightarrow \frac{du}{dx} = \cancel{-}(1+e^{-x}e^{\theta})^{-2}(\cancel{-}e^{-x}e^{\theta}) = \underbrace{\frac{e^{-x}e^{\theta}}{1+e^{-x}e^{\theta}}}_{(1-u)}\underbrace{\frac{1}{1+e^{-x}e^{\theta}}}_{u}$

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

$\Rightarrow$ Score statistic is $\dfrac{n - 2\sum \frac{e^{-x_i}e^{\theta_0}}{1+e^{-x_i}e^{\theta_0}}}{\sqrt{n/3}} \overset{\text{under } H_0:\theta=\theta_0}{\sim} N(0,1)$

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

$\Rightarrow$ REJ 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 ...

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

Likelihood Ratio. If it's significantly greater than one, then we reject H_0. Now we just need ~~LR~~ $\widehat{LR}$, the sampling distribution.

You can prove that:

$\hat{\Lambda} := 2\ln(\widehat{LR}) \xrightarrow{d} \chi^2_1$ Recall $F_{\chi^2_1}(3.84) = 95\%$

E.g. iid Bern(θ)

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

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

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