The LRT is a *lot* more general than the LRT we saw in the previous lecture. E.g. assume DGP iid $f(x; \theta_1, \theta_2, \dots, \theta_K)$ and we want to test

$$H_a: \theta_1 \neq \theta_{1_0} \text{ and/or } \theta_2 \neq \theta_{2_0} \text{ and/or} \dots \text{ and/or } \theta_K \neq \theta_{K_0} \quad \text{vs.}$$

$$H_0: \theta_1 = \theta_{1_0} \text{ and } \theta_2 = \theta_{2_0} \text{ and} \dots \text{ and } \theta_K = \theta_{K_0} \quad \text{then}$$

$$\hat{\Lambda} := 2\ln\left(\frac{\mathcal{L}(\hat{\theta}_1^{MLE}, \dots, \hat{\theta}_K^{MLE}; X_1, \dots X_n)}{\mathcal{L}(\theta_{1_0}, \dots, \theta_{K_0}; X_1, \dots, X_n)}\right) \xrightarrow{d} \chi^2_K$$

For example, let H_0: DGP is iid N(0, 1) vs H_a: DGP iid $N(\theta_1, \theta_2)$ where $\theta_1 \neq 0$ and $\theta_2 \neq 1$.

$$\hat{\Lambda} = 2\ln\left(\frac{\mathcal{L}(\hat{\theta}_1^{MLE}, \hat{\theta}_2^{MLE}; X_1, \dots, X_n)}{\mathcal{L}(0, 1; X_1, \dots X_n)}\right) \xrightarrow{d} \chi^2_2$$

$$\hat{\theta}_1^{MLE} = \bar{X}, \quad \hat{\theta}_2^{MLE} = \frac{1}{n}\sum(X_i - \bar{X})^2$$

$$\text{numerator} = \prod_{i=1}^n \frac{1}{\sqrt{\cancel{2\pi}\hat{\theta}_2^{MLE}}} e^{-\frac{1}{2\hat{\theta}_2^{MLE}}(X_i - \hat{\theta}_1^{MLE})^2}$$

$$\text{denominator} = \prod_{i=1}^n \frac{1}{\cancel{\sqrt{2\pi}}} e^{-\frac{1}{2}X_i^2}$$

For example. Let's do the die roll. Let's test if a die is fair using n rolls and counts of the number of 1's, 2's, ..., 6's. You can show the MLE's for the probability of a certain face is

$$\hat{\theta}_j^{MLE} = \frac{n_j}{n}$$ where n_j is the number of rolls of face j.

$$H_0: \theta_1 = \dots = \theta_6 = \frac{1}{6}$$

$$\hat{\Lambda} = 2\ln\left(\frac{\mathcal{L}(\hat{\theta}_1^{MLE}, \dots, \hat{\theta}_6^{MLE}; X_1, \dots X_n)}{\mathcal{L}(\frac{1}{6}, \dots, \frac{1}{6}; X_1, \dots, X_n)}\right) \to \chi^2_5$$

here's really 5 parameters here (not 6) since if you know 5, the 6th is fixed

$$= 2\ln\left(\frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2} \cdots \left(\frac{n_6}{n}\right)^{n_6}}{\underbrace{\left(\frac{1}{6}\right)^{n_1}\left(\frac{1}{6}\right)^{n_2} \cdots \left(\frac{1}{6}\right)^{n_6}}_{\frac{1}{6^n}}}\right)$$

$$F_{\chi^2_5}(.95) = 11.07$$

$$= 2\ln\left(6^n\left(\frac{n_1}{n}\right)^{n_1} \cdots \left(\frac{n_6}{n}\right)^{n_6}\right) = 2\left(n\ln(6) + n_1\ln\left(\frac{n_1}{n}\right) + \dots + n_6\ln\left(\frac{n_6}{n}\right)\right)$$

using the data from the class where we introduced the chi-sq goodness of fit test, we get

$$= 2\left(15\ln(6) + 4\ln\left(\tfrac{4}{15}\right) + 1\ln\left(\tfrac{1}{15}\right) + 3\ln\left(\tfrac{3}{15}\right) + 2\ln\left(\tfrac{2}{15}\right) + 1\ln\left(\tfrac{1}{15}\right) + 4\ln\left(\tfrac{4}{15}\right)\right)$$

$= 4.056 \approx 3.8$ from the chi-squared computation

$\not> 11.07 \Rightarrow$ Retain H_0

They're not equal because Pearsons's chi-squared test is a different test the the generalized likelihood ratio test!

The generalized likelihood ratio test

In most general terms, if the DGP is iid $f(x; \theta_1, \theta_2, \dots, \theta_K)$ and you wish to test a subset of all K parameters, where the subset has $K_0 \leq K$ parameters e.g. $K = 17$ and $K_0 = 3$ and the null could be something like:

$$H_0: \theta_1 = \theta_{1_0} \text{ and } \theta_7 = \theta_{7_0} \text{ and } \theta_{11} = \theta_{11_0} \Rightarrow K_0 = 3$$

then...

$$\hat{\Lambda} = 2\ln\left(\frac{\mathcal{L}(\hat{\theta}_1^{MLE}, \dots, \hat{\theta}_K^{MLE}; X_1, \dots, X_n)}{\mathcal{L}(\theta_{1_0}, \hat{\tilde{\theta}}_2^{MLE}, \dots, \theta_{7_0}, \dots, \theta_{11_0}, \dots, \hat{\tilde{\theta}}_K^{MLE}; X_1, \dots X_n)}\right) \xrightarrow{d} \chi^2_{K_0}$$

where in the numerator there are K degrees of freedom (K free parameters) and the denominator has K - K_0 degrees of freedom since K_0 parameters are fixed by the null hypothesis. Then the LR statistic is asymptotically distributed as a chi-squared with numerator - denominator = K - (K - K_0) = K_0 degrees of freedom.

$\hat{\tilde{\theta}}_j^{MLE} :=$ the MLE for parameter j assuming the null hypothesis (i.e. assuming the K_0 values of K_0 parameters).

Sometimes the top model with all K parameters free is called the "full model" and the bottom model with < K parameters free is called the "reduced model". The generalized LRT is a test of the "full model" vs "reduced model" answering the question: is the extra complexity of the full model really required when compared to a simpler reduced model? Are the extra K - K_0 parameters justified? The reduced model is said to be "nested" within the full model as a subspace is nested within the entire space.

For example, let's assume the DGP is iid $N(\theta_1, \theta_2)$ and test against the null of $\theta_1 = 0 = \theta_{1_0}$. This is an alternative test to the 1-sample t-test.

$$\frac{\mathcal{L}(\hat{\theta}_1^{MLE}, \hat{\theta}_2^{MLE}; X_1, \dots X_n)}{\mathcal{L}(0, \hat{\tilde{\theta}}_2^{MLE}; X_1, \dots X_n)}$$

$$\begin{cases} \hat{\theta}_1^{MLE} = \bar{x} \\ \hat{\theta}_2^{MLE} = \frac{1}{n}\sum(x_i - \bar{x})^2 \\ \hat{\tilde{\theta}}_2^{MLE} = \frac{1}{n}\sum x_i^2 = \overline{q} \end{cases}$$

$\frac{1}{n}\sum x_i^2 - \bar{x}^2 = \overline{q} - \bar{x}^2$ // ← prev lec

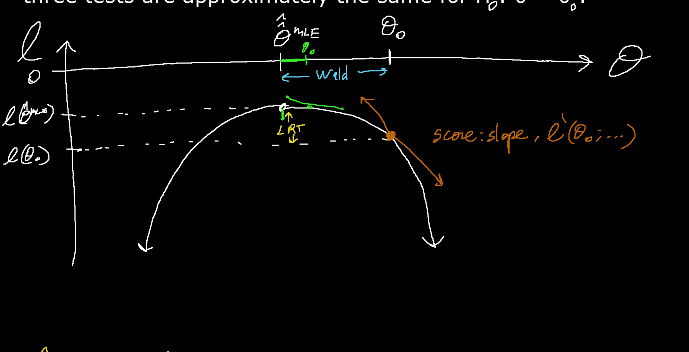
$$= \prod_{i=1}^n \frac{\frac{1}{\sqrt{\cancel{2\pi}(\overline{q} - \bar{x}^2)}} e^{-\frac{1}{2(\overline{q}-\bar{x}^2)}(x_i - \bar{x})^2}}{\frac{1}{\sqrt{\cancel{2\pi}\,\overline{q}}} e^{-\frac{1}{2\overline{q}}x_i^2}}$$

$$= \left(\frac{\overline{q}}{\overline{q} - \bar{x}^2}\right)^{n/2} \cancel{e^{-\frac{1}{2}\left(\frac{1}{\overline{q}-\bar{x}^2}\underbrace{\sum(x_i - \bar{x})^2}_{n(\overline{q}-\bar{x}^2)} - \frac{1}{\overline{q}}\underbrace{\sum x_i^2}_{n\overline{q}}\right)}}$$

$$\Rightarrow \hat{\Lambda} = \cancel{2}\frac{n}{\cancel{2}}\ln\left(\frac{\overline{q}}{\overline{q} - \bar{x}^2}\right) \quad \text{compare to} \quad F_{\chi^2_1}(.95) = 3.84$$

Thm: Wald, Score and LRT are all asymptotically equivalent i.e. as n gets large, they will all give the same results i.e. the same decision in a hypothesis test and the same CI.

Here's an illustration for one parameter that shows how these three tests are approximately the same for H_0: $\theta = \theta_0$:



$$\hat{\Lambda} = 2\ln\left(\frac{\mathcal{L}(\hat{\theta}_{MLE})}{\mathcal{L}(\theta_0)}\right) = 2\left(\ell(\hat{\theta}_{MLE}) - \ell(\theta_0)\right)$$