Tyler Olivieri HW 4 #1

According to the Bureau of Transportation statistics, the national on-time arrival rate for flights in Nov 2018 is 79.66%. Suppose in Nov 2018, we observe 780 out of 1000 flights in JFK arrived on time.

(a) Calculate a 95% confidence interval for the true population proportion $p$ of flights arrived on time in JFK, Nov 2018.

$$\hat{p} = 780/1000 = 78\% = .78$$

It has been derived in class that a confidence interval is $\left(\hat{p} - z^* \sqrt{\frac{\sigma^2}{n}},\ \hat{p} + z^* \sqrt{\frac{\sigma^2}{n}}\right)$

where $z^* = \Phi^{-1}(1 - \alpha/2)$

here $\alpha = .05$

$$\Rightarrow z^* = \Phi^{-1}(1 - .05/2) = 1.96$$

It has also been derived that the variance of a proportion is var: $p(1-p)$, but we don't know $p$ and can formulate an estimate of the variance $s^2 = \hat{p}(1-\hat{p})$

$$\Rightarrow s^2 = .78(1-.78) = .17$$

therefore the 95% confidence interval becomes

$$\left(\hat{p} - z^*\sqrt{\frac{s^2}{n}}, \hat{p} + z^*\sqrt{\frac{s^2}{n}}\right)$$

$z^* = 1.96$

$s^2 = .17$

$n = 1000$

$\hat{p} = .78$

$$\left(.78 - 1.96\sqrt{\frac{.17}{1000}}, .78 + 1.96\sqrt{\frac{.17}{1000}}\right)$$

$\sqrt{\frac{.17}{1000}} = .013$

$1.96 * .013 = .026$

$(.78 - .026, .78 + .026)$

$(.754, .806)$ 95% confidence interval for $p$

(b) Without hypothesis testing, can you give evidence that the on-time arrival rate in JFK is different than the national statistics in Nov 2018? Explain.

A hypothesis test with $\alpha = .05$ significance would not be able to give evidence that the arrival rate in JFK is different than the national average. The acceptance region for a $\alpha = .05$ significance test is the 95% confidence interval. Thus, under a test of the form

$H_0: p = .7966$
$H_a: p \neq .7966$

the null hypothesis would be accepted.

Tyler Olivieri HW #4 2)

Let $X_1, \dots, X_n$ be a sample from a Poisson distribution.

Find the liklihood ratio for testing $H_0: \lambda = \lambda_0$ vs $H_A: \lambda = \lambda_1$, wherere $\lambda_1 > \lambda_0$

Use the fact that the sum of independent poisson R.V follow a Poisson distribution to explain how to determine a rejection region for a test at level alpha

pmf $f(X, \lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}$

$$\Lambda(x) = \frac{f(\lambda_0 \mid x)}{f(\lambda = \lambda_1 \mid x)} = \frac{\prod_{i=1}^{n} \frac{\lambda_0^{x_i} e^{-\lambda_0}}{x_i!}}{\prod_{i=1}^{n} \frac{\lambda_1^{x_i} e^{-\lambda_1}}{x_i!}} =$$

To use the fact that sum of poisson is poisson, the R.V must be independent I will assume iid since we have the whole class

Then $X_i \sim \text{poisson}(\lambda_0)$ under the null hypothesis

the liklihood of $X_1, \dots, X_n$ under null hypothesis

is $\sum_{i=1}^{n} X_i \sim \text{poisson}(n\lambda_0)$

under alternative hypothesis, the liklihood

$\sum X_i \sim \text{poisson}(\lambda_1)$

$\sum_{i=1}^{n} X_i \sim \text{poisson}(n\lambda_1)$

$$\Lambda(x) = \frac{\dfrac{\prod_{i=1}^{n} (\lambda_0)^{x_i} e^{-(\lambda_0)}}{\prod_{i=1}^{n} x_i!}}{\dfrac{\prod_{i=1}^{n} (\lambda_1)^{x_i} e^{-(\lambda_1)}}{\prod_{i=1}^{n} x_i!}} = \frac{\lambda_0^{\sum_{i=1}^{n} x_i} e^{-n\lambda_0}}{\lambda_1^{\sum_{i=1}^{n} x_i} e^{-n\lambda_1}}$$

$$= \left(\frac{\lambda_0}{\lambda_1}\right)^{\sum_{i=1}^{n} x_i} e^{-n\lambda_0 - (-n\lambda_1)}$$

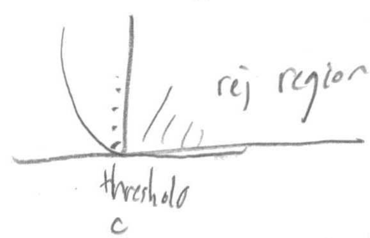
$$= \left(\frac{\lambda_0}{\lambda_1}\right)^{\sum_{i=1}^{n} x_i} e^{-n(\lambda_0 - \lambda_1)}$$

here $n, \lambda_0, \lambda_1$ are constants since $\lambda_1 > \lambda_0$

$\frac{\lambda_0}{\lambda_1} < 1$ and the liklihood function is of the form

$c^{\sum_{i=1}^{n} x_i}$ where $c < 1$

$c^x \quad c<1$




we want to reject $H_0$ when $\Lambda(x)$, the liklihood ratio is small. Thus the rejection region should be on the "right" of the threshold

when $\Lambda(x) > c$ reject $H_0$.

c can be determind by solving $\alpha = Pr\{\Lambda(x) > c \mid H_0 \text{ is true}\}$ for c given alpha

Tyler Olivieri HW4 #3

Suppose that $X_1, \dots, X_n$ form a random sample from a density function, $f(X|\theta)$, for which $T$ is a sufficient statistic for $\theta$. Show that the liklihood ratio test of $H_0: \theta = \theta_0$ versus $H_A: \theta = \theta_1$ is a function of $T$. Explain how, if the distribution of $T$ is known under $H_0$, the rej region of the test may be chosen so that the test has the level $\alpha$.

if $T$ is a sufficient statistic for $\theta$, by the factorization theorem, $f(X|\theta) = h(X)\, g(T(X), \theta)$

the liklihood ratio for the test $H_0: \theta = \theta_0$
$H_a: \theta = \theta_1$
is defined by

$$\Lambda(x) = \frac{\mathcal{L}(\theta_0|x)}{\mathcal{L}(\theta_1|x)}$$

where $\mathcal{L}(\theta|X)$ is the liklihood of the data under $\theta$.

with a fixed $\theta_0, \theta_1$ the liklihood test is equivalent to

$$\Lambda(x) = \frac{f(X|\theta_0)}{f(X|\theta_1)} = \frac{h(x)\, g(T(x), \theta_0)}{h(x)\, g(T(x), \theta_1)} = \frac{g(T(x), \theta_0)}{g(T(x), \theta_1)}$$

thus $\Lambda$ is a function of $T$

$$\Lambda(T(x)) = \frac{g(T(x), \theta_0)}{g(T(x), \theta_1)}$$

If the distribution of $T$ is known under $H_0$, then the null distribution is completely described. A test statistic could be derived that has the property that the probability that $H_0$ is rejected $= \alpha$.

Let this statistic be $\overline{T}$

then $\Pr\{\overline{T} \bowtie t \mid H_0 \text{ is true}\} = \alpha$ where $\bowtie \in \{<, \leq, >, \geq\}$ depending on the context

and $t$ is a threshold for the test.

The threshold is always chosen under the condition that $H_0$ is true, and thus since the null distribution is completley described, the test statistic and threshold can be calculated.

Tyler Olivier. HW4 #4

Let $X_1, \dots, X_n$ be a random sample from an exponential distribution with the density function $f(x|\theta) = \theta \exp(-\theta x)$

Derive a likelihood ratio test of $H_0: \theta = \theta_0$ vs. $H_A: \theta \neq \theta_0$ and show that the rejection region is of the form $\bar{X}\exp(-\theta_0 \bar{X}) \leq c$

$$\Lambda(x) = \frac{\mathcal{L}(\theta_0 | x)}{\mathcal{L}(\theta \neq \theta_0 | x)} = \frac{\prod_{i=1}^{n} \theta_0 (\exp(-\theta_0 x_i)}{\max\limits_{\theta \neq \theta_0} \left[\prod_{i=1}^{n} \theta \exp(-\theta x_i)\right]}$$

$$\arg\max_{\theta \neq \theta_0} \left( \prod_{i=1}^{n} \theta \exp(-\theta x_i) \right)$$

MLE estimate $\hat{\theta} = \frac{1}{\bar{X}}$ where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$

$$\Lambda(x) = \frac{\theta_0^n \exp\left(\sum_{i=1}^{n} (-\theta_0 x_i)\right)}{\prod_{i=1}^{n} \hat{\theta} \exp(-\hat{\theta} x_i)} = \frac{\theta_0^n \exp\left(-\frac{n\theta_0}{n} \sum_{i=1}^{n} x_i\right)}{\hat{\theta}^n \exp\left(\sum_{i=1}^{n} -\hat{\theta} x_i\right)}$$

$$= \frac{\theta_0^n \exp(-n\theta_0 \bar{X})}{\hat{\theta}^n \exp(-n\hat{\theta}\bar{X})}$$

$e^{-n\theta_0 \bar{x}} = (e^{\theta_0 \bar{x}})^{-n}$

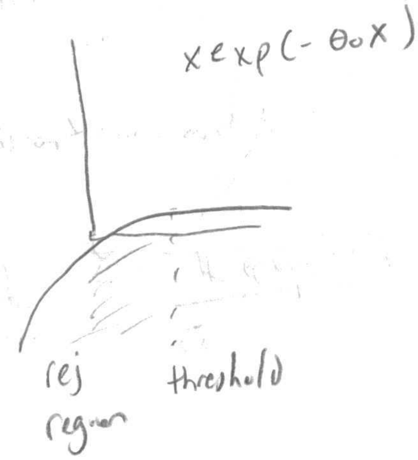
$$e^n \theta_0^n \, \bar{x}^n \exp(\theta_0 \bar{x})^n = \frac{\theta_0^n \exp(-n\theta_0 \bar{x})}{\left(\frac{1}{\bar{X}}\right)^n \exp\left(-n\left(\frac{1}{\bar{X}}\right)\bar{X}\right)} \Rightarrow \frac{\theta_0^n \exp(-n\theta_0 \bar{X})}{\left(\frac{1}{\bar{X}}\right)^n \exp(-n)}$$

$\underbrace{e^n \theta_0^n}_{\text{constant}} [\bar{x} \exp(\theta_0 \bar{x})]^n$

$$\bar{x} \exp(\theta_0 \bar{x}) \qquad = \theta_0^n e^n e^{-n\theta_0 \bar{x}} \, \bar{X}^n = \theta_0^n e^n \left[e^{-\theta_0 \bar{x}} \bar{x}\right]^n$$

$$\Lambda(x) = \theta_0^n e^n \left[\bar{x} e^{-\theta_0 \bar{x}}\right]^n$$

the liklihood function is a function of x, $\bar{x}$ is a function of x, therefor we can look at the function $x e^{-\theta_0 x}$. We want to reject $H_0$ when $\Lambda(x)$ is small.




we can see that this function is small "on the left" of the threshold.

thus the rejection region is

$$\theta_0^n e^n \left[\bar{x} e^{-\theta_0 \bar{x}}\right]^n < K$$

Since $\theta_0, n$ are constants, we can compose them with the threshold to get the desired form

$$\left[\bar{x} e^{-\theta_0 \bar{x}}\right]^n < \frac{K}{\theta_0^n e^n}$$

$$\bar{x} e^{-\theta_0 \bar{x}} < \sqrt[n]{\frac{K}{\theta_0^n e^n}} = c$$

$$\bar{x} e^{-\theta_0 \bar{x}} < c$$

Tyler Olivieri HW4 #5

Let X be a binomial r.v with n trials and probability p of success. $X \sim B(n,p)$ $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$

a) What is the generalized liklihood ratio for testing $H_0: p = .5$ versus $H_A: p \neq .5$

$$\Lambda(x) = \frac{\mathcal{L}(p=.5 \mid x)}{\mathcal{L}(p \neq .5 \mid x)} = \frac{\binom{n}{x} .5^x (1-.5)^{n-x}}{\underset{\substack{p \neq .5 \\ p \in (0,1)}}{\arg\max} \binom{n}{x} p^x (1-p)^{n-x}}$$

max liklihood

$$\underset{p}{\arg\max} \binom{n}{x} p^x (1-p)^{n-x}$$

log liklihood

$$\underset{p}{\arg\max} \log\left(\prod_{i=1}^{n} \binom{n}{x_i} p^x (1-p)^{n-x}\right)$$

$$= \sum_{i=1}^{n} \log\binom{n}{x} + x \log p + (n-x)\log(1-p)$$

$$\frac{dl}{dp} = \frac{x}{p} + \frac{(n-x)}{(1-p)} = 0$$

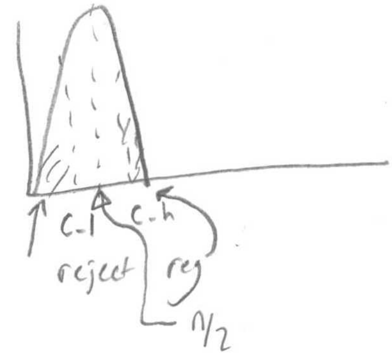
$\hat{p} = \frac{x}{n}$ is MLE estimate

$$\Lambda(x) = \frac{\binom{n}{x} .5^{x} .5^{n-x}}{\binom{n}{x} \hat{p}^{x} (1-\hat{p})^{n-x}} = \frac{.5^{x} .5^{n-x}}{\hat{p}^{x} (1-\hat{p})^{n-x}} = \left(\frac{.5}{\hat{p}}\right)^{x} \left(\frac{.5}{(1-\hat{p})}\right)^{n-x}$$

$$= \left(\frac{.5}{\hat{p}}\right)^{x} \left(\frac{.5}{(1-\hat{p})}\right)^{n} \left(\frac{(1-\hat{p})}{.5}\right)^{x} = \left(\frac{.5(1-\hat{p})}{\hat{p}\,.5}\right)^{x} \left(\frac{.5}{(1-\hat{p})}\right)^{n}$$

$$= \left(\frac{1-\hat{p}}{\hat{p}}\right)^{x} \left(\frac{.5}{(1-\hat{p})}\right)^{n} = \left(\frac{1-\frac{x}{n}}{\frac{x}{n}}\right)^{x} \left(\frac{.5}{1-\frac{x}{n}}\right)^{n}$$

$$\Lambda(x) = \left(\frac{n-x}{x}\right)^{x} \left(\frac{.5}{1-\frac{x}{n}}\right)^{n} \xrightarrow{\text{graph}}$$




reject $H_0$ when

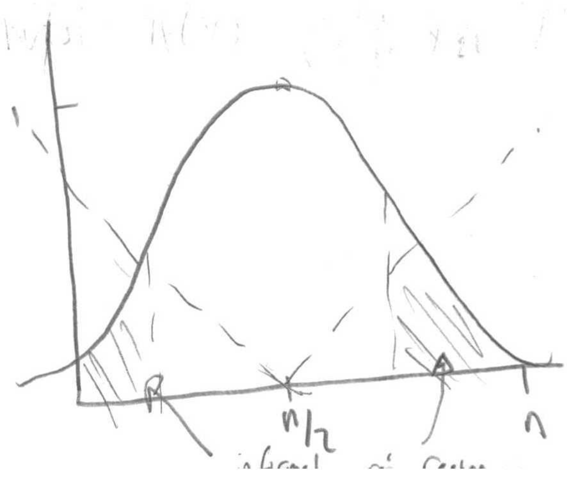
$\Lambda(x) < c\_low$ or $\Lambda(x) > c\_h$

(b) Show that the test rejects when $|X - n/2|$ is large

The test should reject when $\Lambda(x)$ is small.

informally, we can see this by looking at the graphs of $\Lambda(x)$ and $Y = |X - n/2|$

Both are centered at $n/2$. Large values of $|X - n/2|$ are small values of $\Lambda(x)$, thus we can conclude that we should reject the test when $|X - n/2|$ is large.



Tyler Olivieri HW4 #5

c) Using the null distribution of X, show how the significance level corresponding to a rejection region $|X - n/2| > K$ can be determined

with significance level $\alpha$,

$\alpha = Pr\{ |X - n/2| > k \mid H_0 \text{ is true} \}$ Solve for k

$= 1 - Pr\{ |X - n/2| < k \mid H_0 \text{ is true} \}$

d) given $n = 10$ and $K = 2$, what is the significance level of the test?

$\alpha = Pr\{ |X - 10/2| > 2 \mid H_0 \text{ is true} \}$

$= Pr\{ |X - 5| > 2 \mid H_0 \text{ is true} \}$

$= 1 - Pr\{ |X - 5| \le 2 \mid H_0 \text{ is true} \}$

$H_0$ is binomial with $p = .5$

$|X - 5| \le 2$

$-2 \le X - 5 \le 2$

$3 \le X \le 7$

$= 1 - Pr\{ 3 \le X \le 7 \mid H_0 \text{ is true} \}$

$\alpha = 1 - \underbrace{\sum_{i=3}^{7} \binom{10}{i} .5^{i} .5^{10-i}}_{\text{CDF of bernoulli - first 3 trials}}$

$$\alpha = 1 - \left[ \binom{10}{3} .5^3 .5^7 + \binom{10}{4} .5^4 .5^6 + \binom{10}{5} .5^5 .5^5 + \binom{10}{6} .5^6 .5^4 + \binom{10}{7} .5^7 .5^3 \right]$$

$$= 1 - \left[ 120(.125)(.0078) + 210(.0625)(.016) + 252(.03)(.03) + 210(.016)(.0625) + 120(.0078)(.125) \right]$$

$$= 1 - [.117 + .209 + .223 + .209 + .117]$$

$$= 1 - .875 = .125$$

significance level $= \alpha = .125$ (Wolfram solves for $\alpha = .11$) but I rounded a lot in manual calculation which can explain the difference.

e) Use the normal approximation of the binomial distribution to find the significance level if $n = 100$ and $K = 10$

normal approximation $N(np, np(1-p))$

$N(100(.5), 100(.5)(.5))$

$N(50, 25)$

$$\alpha = Pr\{ |X - \tfrac{100}{2}| > K \mid H_0 \text{ is true} \}$$

$$= Pr\{ |X - 50| > 10 \mid H_0 \text{ is true} \}$$

$=$

where $1 - \Pr\{ |X-50| \le 10 \mid H_0 \text{ is true} \}$

$$|X-50| \le 10$$

$$-10 \le X-50 \le 10$$

$$40 \le X \le 60$$

$$= 1 - \Pr\{ 40 \le X \le 60 \mid H_0 \text{ is true} \}$$

$H_0$ is $N(50,25) \rightarrow N(0,1)$

$$\frac{N(50,25) - 50}{5} = N(0,1)$$

$$= 1 - \Pr\left\{ \frac{40-50}{5} \le X \le \frac{60-50}{5} \right\}$$

$$= 1 - \Pr\left\{ -\frac{10}{5} \le X \le \frac{10}{5} \right\}$$

$$= 1 - \Pr\{ -2 \le X \le 2 \}$$

$$= 1 - [\Phi(2) - \Phi(-2)]$$

$$= 1 - [.977 - .023] = 1 - [.954] = .046$$

$\alpha$ = significance level = .046

More on 5c

$$\alpha = 1 - \Pr\{\, |X - n/2| < k \mid H_0 \text{ is true} \,\}$$

$$\alpha = 1 - \Pr\{\, -k + n/2 < X < k + n/2 \mid H_0 \text{ is true} \,\}$$

$$\alpha = 1 - \left[ F(k + n/2) - F(-k + n/2) \right]$$

where $F$ is CDF of $H_0$ (Bernoulli)

and $k + n/2$ are known, $\alpha$ can be computed