

# Homework 5

Problems 11, 12, 13, 20, 23, 26

```
In [ ]: from functools import partial
from math import sqrt, e

from numpy import linspace
from matplotlib.pyplot import subplots, xlabel, ylabel, title, legend
from scipy.stats import norm, chisquare, chi2, beta

from resources import type2, norm_mean, norm_stddev
from resources import bounds
```

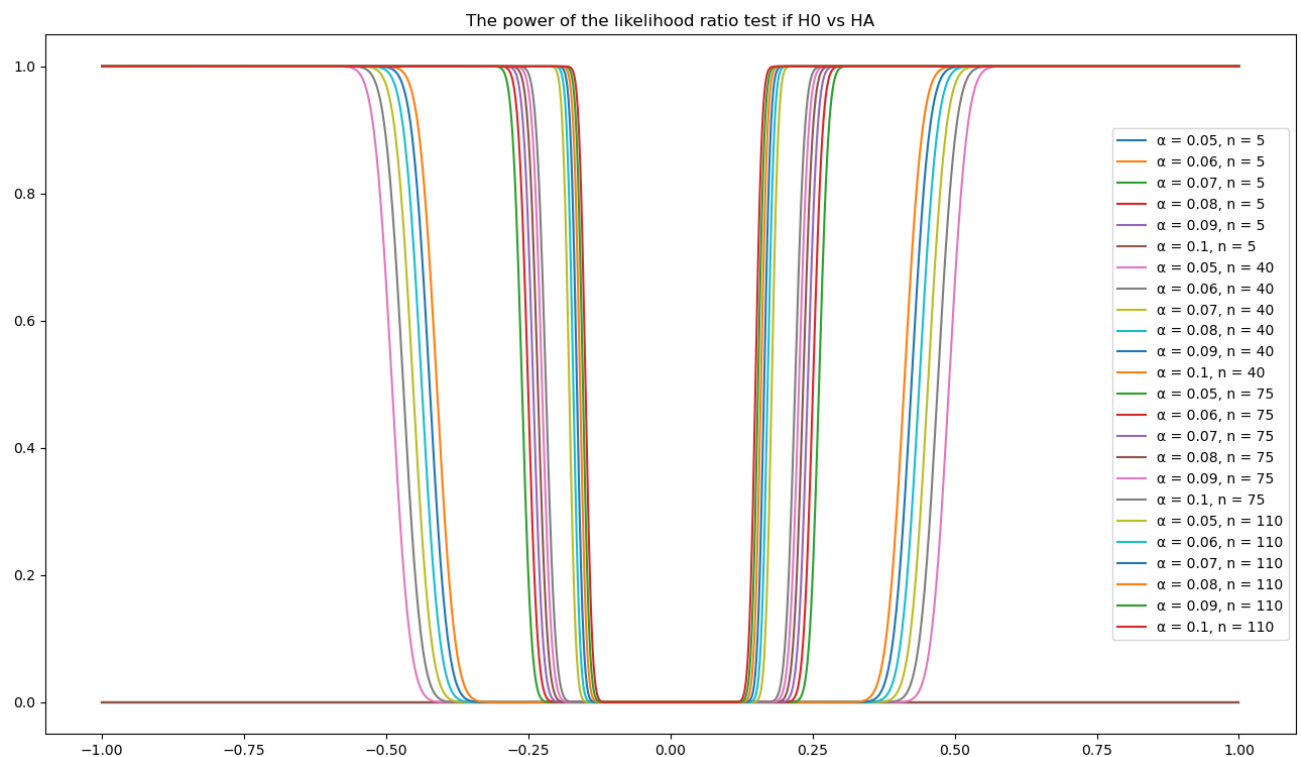
1. Suppose that  $X_1, \dots, X_{25}$  form a random sample from a normal distribution having a variance of 100. Graph the power of the likelihood ratio test of  $H_0: \mu = 0$  versus  $H_A: \mu \neq 0$  as a function of  $\mu$ , at significance levels .10 and .05. Do the same for a sample size of 100. Compare the graphs and explain what you see.

```
In [ ]: # setting vars
as = [i / 100 for i in range(5, 11)]
ns = [i for i in range(5, 111, 35)]
#as = [0.05, 0.10]
#ns = [25, 100]
mean = 0
var = 100
stddev = sqrt(var)
rv = norm

# creating calcs for axes
x = linspace(-1, 1, 1000)
fig, ax = subplots(nrows=1, ncols=1, figsize=(16, 9))
for n in ns:
    for alpha in as:
        b = [type2(p, n, alpha, rv, stddev=stddev) for p in x]
        ax.plot(x, b, label=f"alpha = {alpha}, n = {n}")

title("The power of the likelihood ratio test if H0 vs HA")
ax.legend()
```

```
Out[ ]: <matplotlib.legend.Legend at 0x7fceb33de830>
```



11 | The level of significance impacts the power of the likelihood ratio test by increasing or decreasing the width of the "fail to reject" band. An increase in the sample size appears to be proportional to a decrease the size of the "fail to reject" band, while also making the slope of the curve much sharper. This makes sense intuitively because a larger sample size could be thought of as more "evidence" behind coming to a particular conclusion, making the decision to reject the null hypothesis more stark.

## 12.

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  $H_0 : \theta = \theta_0$  versus  $H_A : \theta \neq \theta_0$ , and show that the rejection region is of the form  $\bar{X} \exp[-\theta_0 \bar{X}] \leq c$ .

Generalized likelihood ratio statistic:

$$\Lambda = \frac{\max_{\theta \in \Theta_0} \text{Lik}(\theta)}{\max_{\theta \in \Theta} \text{Lik}(\theta)}$$

The density function

$$f(x|\theta) = \theta e^{-\theta x}$$

The likelihood function

$$l(x|\theta_0) = \prod_{i=1}^n \theta_0 e^{-\theta_0 X_i}$$

The two likelihood functions

$$l(\theta_0) = \prod_{i=1}^n \theta_0 e^{-\theta_0 X_i}$$

$$l(\theta) = \prod_{i=1}^n \theta e^{-\theta X_i}$$

Expand cumulative products and make ratio

$$L\left(\frac{\theta_0}{\theta}\right) = \frac{\theta_0^{-n} e^{-\sum_{i=1}^n \theta_0 X_i}}{\theta^{-n} e^{-\sum_{i=1}^n \theta X_i}}$$

For rejection region

$$\left(\frac{\theta_0}{\theta}\right)^{-n} \frac{e^{-\theta \sum_{i=1}^n X_i}}{e^{-\theta_0 \sum_{i=1}^n X_i}} \leq c$$

Take log

$$-n \log\left(\frac{\theta_0}{\theta}\right) - \theta \sum_{i=1}^n X_i + \theta_0 \sum_{i=1}^n X_i \leq \log(c)$$

$$-n \log(\theta_0) - n \log(\theta) - \theta \sum_{i=1}^n X_i + \theta_0 \sum_{i=1}^n X_i \leq \log(c)$$

Take  $\frac{\partial}{\partial \theta}$

$$\frac{-n}{\theta} - \sum_{i=1}^n X_i = 0$$

Max likelihood when:

$$\theta = -\frac{1}{\bar{X}}$$

Sub in for rejection region & simplify

$$\frac{\theta_0^{-n} e^{-\theta_0 n \bar{X}}}{\frac{1}{\bar{X}}^{-n} e^n} \leq c$$

group by exponent

$$\left( \frac{e^{-\theta_0 \bar{X}}}{\bar{X} \theta_0 e} \right)^n \leq c$$

Aggregate constants and rearrange

$$\bar{X} \theta_0 e^{\theta_0 \bar{X}} \leq c$$

## 13.

Suppose, to be specific, that in Problem 12,  $\theta_0 = 1$ ,  $n = 10$ , and that  $\alpha = .05$ . In order to use the test, we must find the appropriate value of  $c$ .

13a. Show that the rejection region is of the form  $\bar{X} \leq x_0 \cup \bar{X} \geq x_1$ , where  $x_0$  and  $x_1$  are determined by  $c$ .

$$\alpha = P(LR_n > z) = 1 - P(LR_n \leq z) \approx 1 - F(z)$$

Where  $F(z)$  is the cdf of a Chi-square random variable with  $df$  degrees of freedom

If the null hypothesis is true, the likelihood ratio statistic converges to a Chi-square random variable with  $df$  degrees of freedom

The acceptance region will therefore be in the range  $x_0 \leq \bar{X} \leq x_1$ , which implies that the rejection region is of the form  $\bar{X} \leq x_0 \cup \bar{X} \geq x_1$

13b. Explain why  $c$  should be chosen so that  $P(\bar{X} \exp(-\bar{X}) \leq c) = .05$  when  $\theta_0 = 1$ .

$c$  should be chosen so that  $P(\bar{X} \exp(-\bar{X}) \leq c) = .05$  when  $\theta_0 = 1$  because the significance level of the hypothesis is 0.05. The cdf of the distribution that the statistic converges to would make the test appropriate for the hypothesis.

13c. Explain why  $\sum_{i=1}^{10} X_i$  and hence  $\bar{X}$  follow gamma distributions when  $\theta_0 = 1$ . How could this knowledge be used to choose  $c$ ?

The sum of exponential random variables follows a gamma distribution. This knowledge can therefore allow us to estimate the value of  $c$  by modeling a probability distribution to represent  $\bar{X}$

13d. Suppose that you hadn't thought of the preceding fact. Explain how you could determine a good approximation to  $c$  by generating random numbers on a computer (simulation).

A good approximation to  $c$  could be created by estimating the distribution of the sum of random variables. This is known as the bootstrap method.

1. Consider two probability density functions on  $[0, 1]$ :  $f_0(x) = 1$ , and  $f_1(x) = 2x$ . Among all tests of the null hypothesis  $H_0 : X \sim f_0(x)$  versus the alternative  $X \sim f_1(x)$ , with significance level  $\alpha = 0.10$ , how large can the power possibly be?

$$\alpha = \int_{\Gamma_1} f_0(x)dx = 0.1$$

$$\beta = \int_{\Gamma_1} f_1(x)dx$$

Hypothesis also can be written as

$$H_0 : \theta = 1, H_1 : \theta = 2$$

Where

$$\begin{aligned} X &\sim \beta(\theta, 1) \\ f_X(X) &= \theta x^{\theta-1} \\ x &\in [0, 1] \end{aligned}$$

$$\begin{aligned} f_0(\theta = 1) &= \theta x^{\theta-1} = 1 \\ f_1(\theta = 2) &= \theta x^{\theta-1} = 2x \\ \frac{1}{2x} &\leq c \end{aligned}$$

Is uniformly most powerful because doesn't depend on  $\theta$

Power is greatest when c is smallest, so occurs when x is maximized, at  $x = 1$

```
In [ ]: alpha = 0.10
print("max power = ", 1 - beta(0.5, 1).ppf(alpha))

max power = 0.99
```

1. Suppose that a 99% confidence interval for the mean  $\mu$  of a normal distribution is found to be  $(-2.0, 3.0)$ . Would a test of  $H_0: \mu = -3$  versus  $H_A: \mu \neq -3$  be rejected at the .01 significance level?

Yes, the test would result in a rejection of the null hypothesis.

The 99% confidence interval contains the true parameter with  $p = 0.99$ ,

## 26. True or false:

- a. The generalized likelihood ratio statistic is always less than or equal to 1.

*True.*

$$\Lambda = \frac{\max_{\theta \in \Theta_0} \text{Lik}(\theta)}{\max_{\theta \in \Theta} \text{Lik}(\theta)}$$

for  $\Theta_0 \subset \Theta$

- b. If the p-value is .03, the corresponding test will reject at the significance level .02.

*False.*

The p-value is the lowest significance level of tests that will reject the null hypothesis. &  $0.03 > 0.02$

- c. If a test rejects at significance level .06, then the p-value is less than or equal to .06.

*True.*

The p-value must be  $\leq$  the significance level for the test to reject.

- d. The p-value of a test is the probability that the null hypothesis is correct.

*False.*

The p-value is the prob that the null hypothesis is rejected while also being true.

e. In testing a simple versus simple hypothesis via the likelihood ratio, the p-value equals the likelihood ratio.

*False.* The p-value is the prob that the null hypothesis is rejected while also being true.

The likelihood ratio test uses the probability that the likelihood ratio  $\frac{H_0}{H_1} <$  the observed value, given  $H_0$

f. If a chi-square test statistic with 4 degrees of freedom has a value of 8.5, the p-value is less than .05.

```
In [ ]: f_alpha = 0.05
        f_pval = round(1 - chi2.cdf(8.5, 4), 8)
        print(f"26f. The pval of the chi-sq test stat when val=8.5, df=4 = {f_pval} | is {f_pval} < {f_alpha}? {f_pval < f_alpha}")

26f. The pval of the chi-sq test stat when val=8.5, df=4 = 0.07488723 | is 0.07488723 < 0.05? False

In [ ]: chi2.pdf(8.5, 4)

Out[ ]: 0.03031149705662341
```