Stats 100B: Homework #8

Professor Nicolas Christou
Assignment: 1-5

Eric Chuu
UID: 604406828

March 15, 2017

Exercise 1.
(a) The lifetime of certain batteries are supposed to have variance of 150^2 hours. Using $\alpha = 0.05$, test the following hypothesis.

$$H_0 : \sigma^2 = 150$$
$$H_a : \sigma^2 > 150$$

if the lifetime of 15 of these batteries (which constitutes a random sample from a normal population) have:

$$\sum_{i=1}^{15} x_i = 250, \quad \sum_{i=1}^{15} x_i^2 = 8000$$

where X denotes the lifetime of a battery.

(b) Consider the confidence interval for σ^2. Show that the expected value of the midpoint of this confidence interval is not equal to σ^2.

Solution
(a) Since $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$, we first calculate s^2.

$$s^2 = \frac{1}{15-1}\left(\sum_{i=1}^{15} x_i^2 - n\bar{X}^2\right) = \frac{1}{14}\left(\sum_{i=1}^{15} x_i^2 - \frac{\left(\sum_{i=1}^{15} x_i\right)^2}{n}\right) = \frac{1}{14}\left(8000 - \frac{250^2}{15}\right) = 273.8$$

We reject H_0 if $\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{1-\alpha;n-1}$, and since

$$\frac{(n-1)s^2}{\sigma_0^2} = \frac{14 \cdot 273.8}{150} = 25.55 > \chi^2_{0.95;14} = 23.68$$

we reject H_0.

(b) We first consider the confidence interval given by

$$1 - \alpha = \mathbf{Pr}\left(\chi^2_{\frac{\alpha}{2};n-1} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{1-\frac{\alpha}{2};n-1}\right)$$

Then the the midpoint, M can be written as

$$M = \frac{1}{2}\left(\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2};14}} + \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2};14}}\right)$$
$$\mathrm{E}[M] = \frac{1}{2}\left(\frac{(n-1)\sigma^2}{26.12} + \frac{(n-1)\sigma^2}{5.63}\right) \neq \sigma^2$$

We conclude that the confidence interval is biased. □

Exercise 2.
Let $X \sim \text{Unif}(0, \theta)$. You have exactly one observation from this distribution, and you want to test the null hypothesis $H_0 : \theta = 10$ against $H_a : \theta > 10$, and you want to use the significance level $\alpha = 0.10$. Two testing procedures are being considered:

Procedure G reject H_0 if and only if $X \geq 9$.
Procedure K rejects H_0 if and only if $X \geq 0.5$ or if $X \leq 0.5$.

(a) Confirm that Procedure G has a Type I error probability of 0.10.
(b) Confirm that Procedure K has a Type I error probability of 0.10.
(c) Find the power of Procedure G when $\theta = 12$
(d) Find the power of Procedure K when $\theta = 12$

Solution
(a) We calculate α for Procedure G, the probability of Type I error

$$\alpha = \mathbf{Pr}\left(\text{reject } H_0 \mid H_0 \text{ is true}\right) = \mathbf{Pr}\left(X \geq 9 \mid \theta = 10\right) = \int_9^{10} \frac{1}{10} dx = 0.10$$

(b) We calculate α for Procedure K, the probability of Type I error

$$\alpha = \mathbf{Pr}\left(\text{reject } H_0 \mid H_0 \text{ is true}\right) = \mathbf{Pr}\left(X \geq 9.5 \mid \theta = 10\right) + \mathbf{Pr}\left(X \leq 0.5 \mid \theta = 10\right)$$
$$= \int_{9.5}^{10} \frac{1}{10} dx + \int_0^{0.5} \frac{1}{10} dx = 0.05 + 0.05 = 0.10$$

(c) We calculate $1 - \beta$, the power of Procedure G when $\theta = 12$

$$1 - \beta = \mathbf{Pr}\left(\text{reject } H_0 \mid H_0 \text{ is false}\right) = \mathbf{Pr}\left(X \geq 9 \mid \theta = 12\right) = \int_9^{12} \frac{1}{12} dx = 0.25$$

(d) We calculate $1 - \beta$, the power of Procedure K when $\theta = 12$

$$1 - \beta = \mathbf{Pr}\left(\text{reject } H_0 \mid H_0 \text{ is false}\right) = \mathbf{Pr}\left(X \geq 9.5 \mid \theta = 12\right) + \mathbf{Pr}\left(X \leq 0.5 \mid \theta = 12\right)$$
$$= \int_{9.5}^{12} \frac{1}{12} dx + \int_0^{0.5} \frac{1}{12} dx = \frac{2.5}{12} + \frac{0.5}{12} = 0.25$$

□

Exercise 3.
Consider the simple regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_1, i = 1, \ldots, n$. Find a confidence interval for σ^2. Use $1 - \alpha$ confidence level.

Solution
Consider the standard error,

$$S_e^2 = \frac{s^2}{n}, \qquad \frac{(n-2)^2 S_e^2}{\sigma^2} \sim \chi^2_{n-2}$$

Then we can use S_e^2 as the pivotal quantity to construct the confidence interval.

$$\mathbf{Pr}\left(\chi^2_{\frac{\alpha}{2};n-2} \leq \frac{(n-2)S_e^2}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2};n-2}\right) = 1 - \alpha$$
$$\mathbf{Pr}\left(\frac{1}{\chi^2_{1-\frac{\alpha}{2};n-2}} \leq \frac{\sigma^2}{(n-2)S_e^2} \leq \frac{1}{\chi^2_{\frac{\alpha}{2};n-2}}\right) = 1 - \alpha$$
$$\mathbf{Pr}\left(\frac{(n-2)S_e^2}{\chi^2_{1-\frac{\alpha}{2};n-2}} \leq \sigma^2 \leq \frac{(n-2)S_e^2}{\chi^2_{\frac{\alpha}{2};n-2}}\right) = 1 - \alpha$$

$\square$

Exercise 4.
(a) Let $X_1, X_2, \ldots, X_n$ denote the random sample from a Poisson distribution with parameter λ. Use the Neyman-Pearson Lemma to find the best critical region for testing

$$H_0 : \lambda = 2$$
$$H_a : \lambda = 5$$

(b) Let $Y_1, Y_2, \ldots, Y_n$ be the outcomes of n independent Bernoulli trials. Use the Neyman-Pearson Lemma to find the best critical region for testing

$$H_0 : p = p_0$$
$$H_a : p > p_0$$

Solution
(a) We consider the ratio of the likelihoods,

$$\frac{L(\lambda_0)}{L(\lambda_a)} = \frac{L(2)}{L(5)} = \frac{2^{X_1 + \cdots X_n}}{5^{X_1 + \cdots X_n}} \cdot \frac{e^{-2n}}{e^{-5n}} = \left(\frac{2}{5}\right)^{\sum_{i=1}^n X_i} e^{3n} < k$$

We reject H_0 when $\left(\frac{2}{5}\right)^{\sum_{i=1}^n X_i} e^{3n} < k$

$$\ln\left(\frac{2}{5}\right) \sum_{i=1}^n X_i + 3n < \ln k$$
$$\sum_{i=1}^n X_i > \frac{\ln k - 3n}{\ln\left(\frac{2}{5}\right)}$$

The rejection region is then given by $\sum_{i=1}^n X_i > k'$.

(b) We consider the ratio of the likelihoods,

$$\begin{aligned}
\frac{L(p_0)}{L(p_1)} = \frac{\prod_{i=1}^n f(y_i;p_0)}{\prod_{i=1}^n f(y_i;p_1)} &= \frac{p_0^{Y_1+\cdots+Y_n}(1-p_0)^{n-\sum_{i=1}^n Y_i}}{p_1^{Y_1+\cdots+Y_n}(1-p_1)^{n-\sum_{i=1}^n Y_i}} \\
&= \left(\frac{p_0}{p_1}\right)^{\sum_{i=1}^n Y_i}\left(\frac{1-p_0}{1-p_1}\right)^{n-\sum_{i=1}^n Y_i} \\
&= \left(\frac{p_0(1-p_1)}{p_1(1-p_0)}\right)^{\sum_{i=1}^n Y_i}\left(\frac{1-p_0}{1-p_1}\right)^{n} < k
\end{aligned}$$

Taking log on both sides and find a quantity on the left hand side can be used with a known distribution, we can obtain the following rejection region:

$$\begin{aligned}
&\sum_{i=1}^n Y_i \cdot \ln\left(\frac{p_0(1-p_1)}{p_1(1-p_0)}\right) + n\ln\left(\frac{1-p_0}{1-p_1}\right) < \ln k \\
&\frac{\sum_{i=1}^n Y_i}{n}\cdot \ln\left(\frac{p_0(1-p_1)}{p_1(1-p_0)}\right) + \ln\left(\frac{1-p_0}{1-p_1}\right) < \frac{\ln k}{n} \\
&\Rightarrow \frac{\sum_{i=1}^n Y_i}{n} > \frac{\frac{\ln k}{n} - \ln\left(\frac{1-p_0}{1-p_1}\right)}{\ln\left(\frac{p_0(1-p_1)}{p_1(1-p_0)}\right)} \\
&\Rightarrow \frac{\sum_{i=1}^n Y_i}{n} > k'
\end{aligned}$$

□

Exercise 5.
Suppose that the length in millimeters of metal fibers produced by a certain process will follow the normal distribution with mean μ and standard deviation σ, both unknown. We will test

$$H_0 : \mu = 5.2$$
$$H_a : \mu \neq 5.2$$

A sample size of $n = 15$ metal fibers was selected and was found that $\bar{x} = 5.4$ and $s = 0.4266$.

(a) Approximate the $p-$value using only your t table and use it to test this hypothesis. Assume $\alpha = 0.05$.

(b) Assume now that the population standard deviation is known, $\sigma = 0.4266$. Compute the power of the test when the actual mean is $\mu_a = 5.35$, and you can accept 0.05.

(c) Draw the two distributions (under H_0 and under H_a) and show that the Type I error and Type II error on them.

(d) Assume now that the hypothesis we are testing is

$$H_0 : \mu = 5.2 \quad (1)$$
$$H_a : \mu > 5.2 \quad (2)$$

Determine the sample size needed in order to detect with probability 95% a shift from $\mu_0 = 5.2$ to $\mu_a = 5.3$ if you are willing to accept a Type I error $\alpha = 0.05$. Assume $\sigma = 0.4266$.

Solution

(a) We first calculate the test statistic

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{5.4 - 5.2}{0.4266/\sqrt{15}} = 1.8157$$
$$p = 2 \cdot \mathbf{Pr}\left(t > 1.8157\right) = 2 \cdot \left[1 - \mathbf{Pr}\left(t \leq 1.8157\right)\right]$$

Using the t table, we see that $0.05 \leq p \leq 0.10$, so we do not reject H_0.

(b) Suppose that $\sigma = 0.4266$. We recall that we reject H_0 when the test statistic $Z \leq -Z_{\frac{\alpha}{2}}$ or $Z \geq Z_{\frac{\alpha}{2}}$. In particular, we reject if either inequality holds

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq -Z_{\frac{\alpha}{2}} \Rightarrow \bar{X} \leq -1.96 \cdot \frac{0.4266}{\sqrt{15}} + 5.2 = 4.9841$$
$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq Z_{\frac{\alpha}{2}} \Rightarrow \bar{X} \geq 1.96 \cdot \frac{0.4266}{\sqrt{15}} + 5.2 = 5.4159$$

Then we can compute the power of the test

$$\begin{aligned} 1 - \beta = \mathbf{Pr}\left(\text{reject } H_0 \mid \mu = 5.35\right) &= \mathbf{Pr}\left(Z \leq \frac{4.9841 - 5.35}{0.4266/\sqrt{15}}\right) + \mathbf{Pr}\left(Z \geq \frac{5.4159 - 5.35}{0.4266/\sqrt{15}}\right) \\ &= \mathbf{Pr}\left(Z \leq -0.322)\right) + \mathbf{Pr}\left(Z \geq 0.5982\right) \\ &= 0.2753 \end{aligned}$$

(c) The two distributions under H_0 and under H with Type I and Type II error shown is given below.

(d) Suppose the hypothesis is the one given in lines (1), (2) above. In order to determine the sample size needed in order to detect with probability 0.95 a shift from $\mu_0 = 5.2$ to $\mu_a = 5.3$, we need only determine the sample size to obtain a power of 0.95. Since we reject H_0 when

$$\bar{X} \geq z_\alpha \cdot \frac{\sigma}{\sqrt{n}} + \mu_0 = 1.645 \cdot \frac{0.4266}{\sqrt{n}} + 5.2$$

we can calculate the power as follows and find the sample size required to give us 0.95 power,

$$\begin{aligned} 1 - \beta = 0.95 = \mathbf{Pr}\left(Z \geq \frac{1.645 \cdot \frac{0.4266}{\sqrt{n}} + 5.2 - 5.3}{0.4266/\sqrt{n}}\right) \\ \Rightarrow Z_{0.05} = -1.645 = 1.645 - \frac{0.1}{0.4266/\sqrt{n}} \\ \Rightarrow 2 \cdot 1.645 = \frac{0.1\sqrt{n}}{0.4266} \\ \Rightarrow n = 197 \end{aligned}$$

□