Math 362, Problem set 4

Due 2/23/10

Note: If you turn HW in on 2/21, I will return it for you by 2/23 for the purpose of studying for the exam.

1. (5.5.13) Let $p$ denote the probability that, for a particular tennis player, the first serve is good. Since $p = .4$, this player decided to take lessons in order to increase $p$. When the lessons are completed, the hypothesis $H_0 = p = .4$ will be tested against $H_1 : p > .4$ based on $n = 25$ trials. Let $y$ equal the number of first serves that are good, and let the critical region be defined by $C = \{ y : y \geq 13 \}$

(a) Determine $\alpha = P_{p=.4}(Y \geq 13)$.

(b) Find $\beta = P_{p=.6}(Y < 13)$. That is $1 - \beta$ is the power at $p = 0.6$.

Answer: For (a),

$$\alpha = P_{p=.4}(Y \geq 13) = \sum_{k=13}^{25} \binom{25}{k} (.4)^k (.6)^{25-k} \approx 0.154.$$ 

For (b),

$$\beta = P_{p=.6}(Y < 13) = \sum_{k=1}^{12} \binom{25}{k} (.6)^{25-k} (.4)^k \approx 0.154.$$ 

Note that these are actually the same sums!

2. (5.6.2) Consider the power function $\gamma(\mu)$ and it’s derivative given by (5.6.5). Show that $\gamma'(\mu)$ is strictly negative for $\mu < \mu_0$, and strictly positive for $\mu > \mu_0$. What does this mean about $\gamma(\mu)$?

Answer:

Let

$$a = \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma}.$$
Then $a < 0$ if $\mu > \mu_0$ and $a > 0$ if $\mu < \mu_0$. Using (5.6.6)

$$
\gamma'(\mu) = \frac{\sigma}{\sqrt{n}} \left( e^{-\left( a - z_{\alpha/2}\right)^2/2} - e^{-\left( a + z_{\alpha/2}\right)^2/2} \right).
$$

If $\mu < \mu_0$ then $(a + z_{\alpha/2})^2 > (a - z_{\alpha/2})^2$, so $e^{-\left( a + z_{\alpha/2}\right)^2/2} < e^{-\left( a - z_{\alpha/2}\right)^2/2}$, so $\gamma'(\mu) < 0$.

If $\mu > \mu_0$, then $(a + z_{\alpha/2})^2 < (a - z_{\alpha/2})^2$, so $e^{-\left( a + z_{\alpha/2}\right)^2/2} > e^{-\left( a - z_{\alpha/2}\right)^2/2}$, so $\gamma'(\mu) > 0$.

Therefore $\gamma(\mu)$ has a minimum at $\mu_0$: That is to say that the probability that I accept $H_1$ is smallest when $H_0$ is true; a very good thing!

3. (5.6.9) In exercise 5.4.14, (on HW 2) we found a confidence interval for the variance $\sigma^2$ using the variance $S^2$ of a random sample of size $n$ arising from $N(\mu, \sigma^2)$, where the mean $\mu$ is unknown. In testing $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$, use the critical region defined by $(n - 1)S^2/\sigma_0^2 \geq c$.

That is, reject $H_0$ and accept $H_1$ if $S^2 \geq c\sigma_0^2/(n - 1)$. If $n = 13$ and the significance level $\alpha = 0.025$, determine $c$.

**Answer**

Recall, $(n - 1)S^2/\sigma^2$ has $\chi^2(n - 1)$ distribution. Thus if $H_0 : \sigma^2 = \sigma_0^2$ is true, $(n - 1)S^2/\sigma_0^2$ has $\chi^2(n - 1)$ distribution, and

$$
\alpha = \mathbb{P}(\chi^2(n - 1) \geq c) = \mathbb{P}(\chi^2(n - 1) > c)
$$

Since $n = 13$, and we want $\alpha = 0.025$, we consult the table and find that $c = 23.337$.

**Note:** Since our test is 'one sided' (in the sense that we only reject $H_0$ if $(n - 1)S^2/\sigma_0^2$ is large enough, we consult the table to find the $c$ so that $\mathbb{P}(\chi^2(n - 1) \leq c) = .975$.

4. (5.7.3) A die was cast $n = 120$ time independent times, and the following data resulted:

<table>
<thead>
<tr>
<th>Spots Up</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>$b$</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>40</td>
</tr>
</tbody>
</table>

If we use a chi-square test, for what values of $b$ would the hypothesis that the die is unbiased be rejected at the 0.025 significance level.

**Answer:**

If the die is unbiased, $p_i = \frac{1}{6}$ for $i = 1, \ldots, 6$. Thus

$$
\sum \frac{(X_i - pn)^2}{pn} = \frac{(b - 20)^2}{20} + \frac{(40 - b - 20)^2}{20} = \frac{(b - 20)^2}{10}.
$$
Since our test (at the 0.025 significance level) is to reject this if \( \frac{(b-20)^2}{10} > 12.833 \) (this number is obtained by comparing to a \( \chi^2(5) \) distribution), we solve
\[
\frac{(b-20)^2}{10} > 12.833
\]
if
\[
b > \sqrt{128.33} + 20
\]
or
\[
b < 20 - \sqrt{128.33}
\]
There are two sides, coming from the positive and negative square root of \((b-20)^2\); both need to be accounted for.

5. (5.7.7) A certain genetic model suggest that the probabilities of a particular trinomial distribution are, respectively,
\[
p_1 = p^2, \quad p_2 = 2p(1-p) \quad \text{and} \quad p_3 = (1-p)^2,
\]
where \(0 < p < 1\). If \(X_1, X_2, X_3\) represent the respective frequencies in \(n\) independent trials, explain how we could check on the adequacy of the genetic model.

**Answer:** If \(p = p_0\) is given, then we simply compute
\[
\frac{(X_1 - p_0^2n)^2}{p_0^2n} + \frac{(X_2 - 2p_0(1-p_0)n)^2}{2p_0(1-p_0)n} + \frac{(X_3 - (1-p_0)^2n)^2}{(1-p_0)^2n}
\]
and for our desired significance level, we compare the outcome with the \(c\) we find for a \(\chi^2(2)\) random variable.

Generally, we are not so lucky as to be given \(p_0\), and our goal is to test whether there exists a \(p\) so that the data fits the distribution. Hence we need to find an estimate \(\hat{p} = \hat{p}(X_1, X_2, X_3)\) and use that estimate to compute
\[
\frac{(X_1 - \hat{p}^2n)^2}{\hat{p}^2n} + \frac{(X_2 - 2\hat{p}(1-\hat{p})n)^2}{2\hat{p}(1-\hat{p})n} + \frac{(X_3 - (1-\hat{p})^2n)^2}{(1-\hat{p})^2n}.
\]
Then we would compare to the value of \(c\) coming from our desired significance level, comparing now with a \(\chi^2(1)\) random variable. **Important point:** The number of degrees of freedom is reduced by one since we estimate a parameter.

As far as finding \(\hat{p}\), there are a couple of reasonable ways to estimate the parameter. I think what the book had in mind is the following:

Note that (if \(H_0\) is correct)
\[
X_1 \approx p^2n \quad \text{and} \quad X_2 \approx 2(p - p^2)n
\]
Therefore
\[
2X_1 + X_2 \approx 2pn
\]
so
\[ p \approx \frac{2X_1 + X_2}{2n} = \frac{2X_1 + X_2}{2(X_1 + X_2 + X_3)}. \]

Therefore we can take
\[ \hat{p} = \frac{2X_1 + X_2}{2(X_1 + X_2 + X_3)}. \]

*Note:* I am flexible as to how you found \( \hat{p} \).

6. (5.8.3). Suppose \( X \) is a random variable with the pdf \( f_X(x) = b^{-1}f((x - a)/b) \), where \( b > 0 \). Suppose we can generate observations from \( f(z) \). Explain how we can generate observations from \( f_X(x) \).

*Answer:*
Recall: If \( X = g(Y) \) (where \( g(y) \) is increasing or decreasing) and \( Y \) has pdf \( f(z) \), then
\[ f_X(x) = f(g^{-1}(x))|g^{-1}'(x)|. \]

We see \( f_X(x) = b^{-1}f((x - a)/b) \) is of this form, where \( g^{-1}(x) = \frac{x-a}{b} \).
That is \( g(y) = by + a \). In other words, if we can generate \( Y \) with pdf \( f(z) \), we can find \( X \) with pdf \( f_X(x) \) by taking \( X = bY + a \).

*Remark:* I promised some test-like questions, but the book questions are pretty good in this regard this time - especially 1, 3, 4. Even problems like 5 written a bit less open endedly is okay. 6, I would try and make more concrete. As I hit sections where the book problems are less suitable for exam problems, though, I will try and write some of my own problems. Please give me feedback as to whether how the homework problems are going for you as we head towards this exam.