Probabilities and statistics

Lecturer: Prof. Dr. Sara van de Geer
Prof. Dr. Martin Larsson

Serie 12

May 25th, 2015

Q1. We toss 100 times a coin and we get 60 head. We want to do a test to know whether the coin is fair.

(a) Test the hypothesis with a 0.01 level of significance. Should this test be one or two-tailed?.

(b) What is the biggest amount of head should we have in 100 tossings so we cannot discard H_0 := "The coin biased towards tail".

(c) Calculate all p_0 so that the null hypothesis

$$H_0(p_0) := \text{"Probability of head is } p_0 \text{ "},$$

would not be rejected in a test with 0.05 level of significance.
Hint: It will be useful to use the central limit theorem in all of this question.

Solution

Let $(X_i)_{i=1}^n$ an i.i.d sequence of Bernoulli(p). Let $X = \sum_{i=1}^{100} X_i$, then $X \sim Bin(100, p)$.

(a) We want to know whether the coin is fair or not, so our hypothesis are

$$\begin{aligned} H_0 &: \text{The coin is fair,} \\ H_1 &: \text{The coin is not fair.} \end{aligned}$$

That is to say

$$\begin{aligned} H_0 &: p = p_0 = \frac{1}{2}, \\ H_1 &: p \neq p_0. \end{aligned}$$

Then we should o a two-tailed test.
Under H_0 we have that $\mathbb{E}_0(X) = 50$ and $Var_0(X) = np_0(1-p_0) = 25$. Now, take c_1 and c_2 such that

$$\begin{aligned} 0.01 &\geq \mathbb{P}_0(X \notin (c_1, c_2)) \\ &= 1 - \mathbb{P}_0\left(\frac{c_1 - 50}{5} \leq \frac{X-50}{5} \leq \frac{c_1 - 50}{5}\right) \\ &\sim 1 - \phi\left(\frac{c_2 - 50}{5}\right) + \phi\left(\frac{c_1 - 50}{5}\right). \end{aligned}$$

We would like to make the rejection zone as biggest as we can, given that ϕ is symmetric, we just have to make

$$\begin{aligned} \frac{c_1 - 50}{5} &= -\frac{c_2 - 50}{5}, \\ \Rightarrow c_1 &= 100 - c_2. \end{aligned}$$

Finally

$$\begin{aligned}
0.01 \geq 1-\phi\left(\frac{c_2-50}{5}\right)+1-\phi\left(\frac{c_2-50}{5}\right), \\
\Leftarrow \phi\left(\frac{c_2-50}{5}\right) \geq 0.995 \\
\Leftarrow c_2 \geq 62.9,
\end{aligned}$$

then

$$K_{1\%}=[0,37]\cup[63,100].$$

Given that $60 \notin K_{1\%}$ we cannot reject H_0 with 0.01 level of significance.

(b) We have to take our hypothesis

$$\begin{aligned}
H_0 &: p \leq p_0 = \frac{1}{2} \\
H_1 &: p > p_0,
\end{aligned}$$

now we have to find c such that

$$0.01 \geq \mathbb{P}_0(X \geq c) \sim 1-\phi\left(\frac{c-50}{5}\right).$$

Then we have to choose $c \geq 61.5$, from what we have that $K_{1\%}=[62,100]$. So we reject H_0 if we have 62 or more heads.

(c) Take $p_0 \in (0,1)$ and

$$\begin{aligned}
H_0 &: p = p_0, \\
H_1 &: p \neq p_0.
\end{aligned}$$

Note that $\mathbb{E}_0(X)=100p_0$ and $Var_0(X)=100p_0(1-p_0)$. Then, take c_1 and c_2 such that

$$\begin{aligned}
0.01 &\geq \mathbb{P}_0(X \notin (c_1,c_2)) \\
&= 1-\mathbb{P}_0\left(\frac{c_1-100p_0}{10\sqrt{p_0(1-p_0)}} \leq \frac{X-100p_0}{10\sqrt{p_0(1-p_0)}} \leq \frac{c_1-100p_0}{10\sqrt{p_0(1-p_0)}}\right) \\
&\sim 1-\phi\left(\frac{c_2-100p_0}{10\sqrt{p_0(1-p_0)}}\right)+\phi\left(\frac{c_1-100p_0}{10\sqrt{p_0(1-p_0)}}\right).
\end{aligned}$$

Doing the same that in question a we have that $c_1(p_0)=100p_0-19.6\sqrt{p_0(1-p_0)}$ and $c_2(p_0)=100p_0+19.6\sqrt{p_0(1-p_0)}$. Then we are interested in the set

$$\begin{aligned}
&\left\{p : 100p-19.6\sqrt{p(1-p)} \leq 60 \leq 100p+19.6\sqrt{p(1-p)}\right\} \\
&=\left\{p : 19.6^2 p(1-p) \leq (60-100p)^2 \leq 19.6^2 p(1-p)\right\} = [0.502, 0.691].
\end{aligned}$$

Q2. Consider the null hypothesis $X \sim f(x)dx$ and the alternative $X \sim f(x-1)dx$ for the following cases:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$
$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Compute the form of the rejection of the likelihood area ratio test (Neyman-Pearson Lemma). Comment the difference

Solution:

Using the Neyman-Pearson test with the hypothesis

$$H_0 : X \sim f(x)dx,$$
$$H_1 : X \sim f(x-1)dx.$$

We have that the likelihood ratio is given by

$$L = \frac{f(x-1)}{f(x)}.$$

then in the case of the normal variable $L = e^{x-\frac{1}{2}}$, and we have rejection when $L > c$, i.e.$x > \ln c + \frac{1}{2}$. Then the rejection set is of the form (a, ∞).

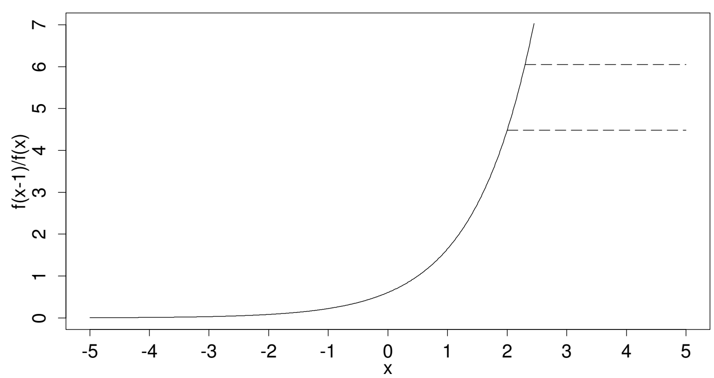

Figure 1: Rejections sets of the normal case.

In the case of the Cauchy random variable the likelihood ration is given by $L = \frac{x^2+1}{x^2-2x+2}$, then we have an interesting behavior as you can see in figure 2. If you put $c = 1$ then you will have a non bounded interval, but if you put $c > 1$ you will have a bounded interval.

This happens because the Cauchy distribution is heavy tailed.

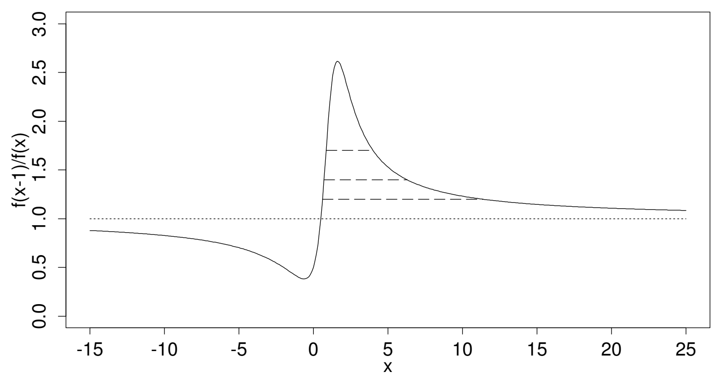

Figure 2: Rejections sets of the cauchy case.

Q3. Let $(X_i)_{i=1}^n$ be an i.i.d F-distributed sequence. Let F be absolutely continuous. The Sign test is a test where the null hypothesis is that the median of X is m, i.e.

$$F^{-1}(m) = \frac{1}{2}.$$

Use the Theorem 6.4 of the Skript to construct the test with significance level $\alpha = 0.05$.

Solution:

We have to construct the test between the hypothesis:

$$H_0 : F^{-1}(0.5) = m,$$
$$H_1 : F^{-1}(0.5) \neq m.$$

We will use the statistic $T_{n,m} = \sum_{i=1}^m \mathbf{1}_{\{X_n \leq m\}}$ and the test is going to be given by

$$\phi(x) = 1 \Leftrightarrow \left|T_{n,m} - \frac{n}{2}\right| > c(n, \alpha),$$

where $x = (x_i)_{i=1}^n$, n is the size of the experiment and α is the level of the test. We have that under H_0 $T_{n,m}$ follows a law of $Bin(n, \frac{1}{2})$. Then if we define $k = \frac{n}{2} - c(n, \alpha)$, we should have, thanks to the symmetry of the Binomial:

$$\sum_{j=0}^{k-1} \binom{n}{j} 0.5^n \leq \frac{\alpha}{2} < \sum_{j=0}^{k} \binom{n}{j} 0.5^n,$$

and $\frac{n}{2} + c(n, \alpha) = n - k$. Take $C \subseteq \mathbb{R}^{n+1}$

$$C := \left\{ (x_1, ..., x_n, m) : k \leq \sum_{i=1}^n \mathbf{1}_{\{x_i \geq m\}} < n - k \right\},$$

then we would like $B((x_1, .., x_n)) = \{m : (x_1, ..., x_n, m) \in C\}$ to be a confidence interval of level α of $F(\frac{1}{2})$. Note that

$$m \in [x_{(j)}, x_{(j+1)}) \Leftrightarrow \sum_{i=1}^n \mathbf{1}_{\{x_i > m\}} = j$$

then $B((x_1,...,x_n)) = [x_{(k)}, x_{(n-k)}]$. Thanks to the central limit theorem

$$\begin{aligned}\mathbb{P}_m\left(k \leq \sum_{i=1}^n \mathbf{1}_{\{x_i>m\}} \leq n-k\right) &= \mathbb{P}_m\left(\frac{2}{\sqrt{n}}\left(k-\frac{n}{2}\right) \leq \frac{2}{\sqrt{n}}\left(\sum_{i=1}^n \mathbf{1}_{\{x_i>m\}} - \frac{n}{2}\right) \leq \frac{2}{\sqrt{n}}\left(\frac{n}{2}-k\right)\right)\\ &\approx \phi\left(\frac{2}{\sqrt{n}}\left(\frac{n}{2}-k\right)\right) - \phi\left(\frac{2}{\sqrt{n}}\left(k-\frac{n}{2}\right)\right)\\ &= 2\phi\left(\frac{2}{\sqrt{n}}\left(\frac{n}{2}-k\right)\right) - 1.\end{aligned}$$

We want it to be bigger than 0.95, so $k \approx \frac{n}{2} - \frac{1.96}{2}\sqrt{n} \approx \lfloor \frac{n}{2} \rfloor - \sqrt{n}$. Then

$$B((x_1,...,x_n)) = \left[x_{(\lfloor \frac{n}{2}\rfloor - \sqrt{n})}, x_{(\lfloor \frac{n}{2}\rfloor + \sqrt{n})}\right],$$

is a confidence interval with 95% of confidence level. Then using the same notation as in Theorem 6.4 we have that

$$\begin{aligned}C &:= \{(x, B(x))\}\\ &= \left\{(x,m) : m \in \left[x_{(\lfloor \frac{n}{2}\rfloor - \sqrt{n})}, x_{(\lfloor \frac{n}{2}\rfloor + \sqrt{n})}\right]\right\}.\end{aligned}$$

Then $A(m) = \{x \in \mathbb{R}^n : \theta \in \left[x_{(\lfloor \frac{n}{2}\rfloor - \sqrt{n})}, x_{(\lfloor \frac{n}{2}\rfloor + \sqrt{n})}\right]\}$, so thanks to the Theorem 6.4 a test to the level 0.95 is given by $\varphi_m(x) = \mathbf{1}_{\left\{x_{(\lfloor \frac{n}{2}\rfloor - \sqrt{n})} \leq m \leq x_{(\lfloor \frac{n}{2}\rfloor + \sqrt{n})}\right\}}$.