Exam Statistics and Probability

November 3, 2022

Download this Markdown (Rmd) file, and add your R-code to answer the questions. When you are finished, render the document either to pdf, word or html and upload to Brightspace. If the rendering fails, upload the Markdown (Rmd) file with your R code, or upload your R code in any format you want. If the uploading fails, email your work to E.W.van_Zwet@lumc.nl.

There are 4 problems. Some familiar R commands and formulas are available at the end of this exam.

Problem 1.

Suppose we have an i.i.d. sample $X_1, X_2, \ldots, X_n$ from the normal distribution with mean $\mu = 1$ and unknown standard deviation σ.

(a) What is the method-of-moments estimator of σ?

Run the following code to obtain a sample of size $n = 15$ from a normal distribution with mean $\mu = 1$ and some unknown standard deviation σ.

```
x=c(6.83,4.54,5.09,1.32,4.52,-0.2,7.64,2.08,-2.87,4.98,5.02,-2.12,-5.51,1.1,3.56)
```

(b) Use the bootstrap to compute the mean squared error of the method-of-moments estimator of σ. If you were unable to find the method-of-moments estimator of σ, then you may use the sample standard deviation instead.

```
# add your R code here
```

Problem 2.

Suppose we have an i.i.d. sample $X_1, X_2, \ldots, X_{13}$ from a distribution with density

$$f_\theta(x) = \theta x^{\theta - 1}.$$

The sample is

```
x=c(0.86,0.96,0.69,0.99,0.77,0.98,0.81,0.89,0.97,0.83,0.89,0.81,0.95)
```

(a) Plot the loglikelihood function for values of θ between 0 and 20.
(b) Compute the maximum likelihood estimator of θ numerically.

Answer the following questions with pen and paper.

(c) Derive a formula for the maximum likelihood estimator of θ.
(d) Compute the asymptotic variance of the maximum likelihood estimator of θ.
(e) Compute the 95% asymptotic Wald confidence interval for θ.

```
# add your R code here.
```

Problem 3

Suppose we have an i.i.d. sample $X_1, X_2, \ldots, X_n$, with $n = 50$, from a normal distribution with unknown μ and unknown σ^2. We are interested in testing the null hypothesis that the data come from the *standard* normal distribution, i.e. $H_0 : \mu = 0$ and $\sigma^2 = 1$, against $H_A : \mu \neq 0$ and/or $\sigma^2 > 1$. We will reject H_0 for large values of the test statistic

$$T = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

Use R to solve the following exercises numerically.

(a) If we reject when $T > 1.3$, what is the Type I error of the resulting test?
(b) Supposing again that we reject when $T > 1.3$, what is the power of the resulting test when $\mu = 0.5$ and $\sigma = 1.2$?
(c) Which critical value should we use to obtain a Type I error of 0.05?
(d) Suppose we observe $T = 1.5$. What is the p-value of the test?

```
# add your R code here.
```

Problem 4.

The negative binomial distribution, with parameters r and p, is closely related to the binomial distribution. The probability mass function is given by

$$P(X = x) = \binom{x + r - 1}{x} (1 - p)^r p^x.$$

The expected value of a negative binomial random variable is $rp/(1 - p)$.

Suppose that we have iid observations $X_1, \ldots, X_n$ from a negative binomial distribution with r known and p unknown.

(a) Show that the score is
$$\frac{\sum_{i=1}^n x_i}{p} - \frac{nr}{1 - p}.$$

(b) Show that the Fisher information is
$$\frac{nr}{p(1 - p)} + \frac{nr}{(1 - p)^2}.$$

Suppose that we have a sample of size $n = 100$ from a negative binomial distribution with sample average $\bar{X} = 6.2$, and that $r = 5$.

(c) Perform a one-sided asymptotic score test for $H_0 : p = 0.5$ against $p > 0.5$. What is your conclusion about the null hypothesis?

```
# add your R code here.
```

some R functions

help: help

calculator: + - * / abs x^2 sqrt log exp

vectors: c seq rep 1:10

operations on vectors: sum prod length max which.max

plot: plot points lines hist boxplot

boolean variables: which & |

sub-setting with square brackets: x=c(5,2,6,1,2); x[3]

boolean variables and sub-setting: sum(x<4); x[x<4]

input/output: cat load

control structures: for loop and if statement

probability distributions: d/p/q/rnorm *unif *binom *exp *geom *pois

descriptive statistics: summary mean median var sd

hypothesis testing: t.test, prop.test, chisq.test

some formulas

Suppose X and Y are random variables and a and b are scalars (constants, numbers).

$$\begin{aligned} E(aX+b) &= aE(X)+b \\ E(X+Y) &= E(X)+E(Y) \\ \mathrm{Var}(X) &= E(X^2)-E(X)^2 \\ \mathrm{Var}(aX+b) &= a^2\mathrm{Var}(X) \\ \mathrm{Var}(X+Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y) \end{aligned}$$

Moreover, the mean squared error of an estimator $\hat{\theta}$ of a parameter θ is

$$\mathrm{MSE}(\hat{\theta}) = E[(\hat{\theta}-\theta)^2] = \mathrm{Var}(\hat{\theta}) + E(\hat{\theta}-\theta)^2$$

In other words, the MSE is the variance plus the square of the bias.

An overview of the Wald, score and likelihood ratio tests and their asymptotic distributions:

Statistic	Definition	Distribution
Wald	$\hat{\theta}-\theta_0$	$\mathcal{N}\{0, 1/\mathcal{I}(\hat{\theta})\}$
Score	$\text{loglikelihood}'(\theta_0)$	$\mathcal{N}\{0, \mathcal{I}(\theta_0)\}$
Likelihood ratio	$\text{loglikelihood}(\hat{\theta}) -$ $\text{loglikelihood}(\theta_0)$	$\frac{1}{2}\chi^2_{\text{df}=1}$