

Practice Exam Statistics and Probability

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

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

Problem 1.

Suppose we have a sample of size 10 from the exponential distribution with unknown rate λ . Generate a small dataset by running the following R code:

```
set.seed(123)
x = rexp(10,rate=2)
```

As you know, the mean of the exponential distribution is $1/\lambda$ and the variance is $1/\lambda^2$. We decide to estimate the variance with the square of the sample average. Use the bootstrap to answer the following questions.

- (a) What is the bias of the estimator?
- (b) What is the variance of the estimator?
- (c) What is the mean squared error of the estimator?

(a)

```
set.seed(123)
lambda_hat_v = replicate(1e5, {
  x = rexp(10,rate=2)
  lambda_hat = (1/(var(x)*(10-1)/10))**0.5
  return(lambda_hat)
})
lambda_hat_v[1:10]
```

```
## [1] 2.497630 1.836606 4.147442 2.812384 1.025384 2.760223 4.044387 2.550313
## [9] 1.795285 3.214233
```

```
bias = mean(lambda_hat_v) - 2
bias
```

```
## [1] 0.7066743
```

(b)

```
var(lambda_hat_v)
```

```
## [1] 1.47759
```

(c)

```
mse = mean((lambda_hat_v - 2) ** 2)
mse
```

```
## [1] 1.976963
```

Problem 2

We assume that $X = (1.02, 3.35, 6.83, 4.60, 1.80)$ are distributed according to a Rayleigh distribution with parameter θ^2 . The Rayleigh distribution only takes positive values and its density for $x > 0$ is

$$f(x) = \frac{x}{\theta^2} e^{-x^2/(2\theta^2)}.$$

The expectation of the Rayleigh distribution is $\theta\sqrt{\pi/2}$.

- (a) Calculate a method-of-moments estimate of θ^2 .
- (b) Find the Maximum likelihood estimator of θ^2 analytically (by pen and paper).
- (c) Find the maximum likelihood estimator of θ^2 numerically using optimize.

(a)

```
# add your R code
X = c(1.02, 3.35, 6.83, 4.60, 1.80)
theta_squre_hat = (mean(X)**2) / (pi/2)
theta_squre_hat
```

```
## [1] 7.887974
```

(c)

```
n = length(X)
mle = function (theta_sq) {
  result = -1*n*log(theta_sq)+sum(log(X) - (X**2)/(2*theta_sq))
  return(-1*result)
}
opt_result = optimize(mle, c(0, 100))
theta_squre_hat1 = opt_result$minimum
theta_squre_hat1
```

```
## [1] 8.331179
```

Problem 3

Suppose we have a *single* observation X from the Poisson distribution with unknown λ . We want to perform a one-sided test

$$H_0 : \lambda = 10 \quad \text{versus} \quad A : \lambda > 10$$

We reject the null hypothesis if $X > 13$. Use the R function `ppois` to answer the following questions.

- (a) Determine the probability of a type I error.
- (b) Determine the probability of a type II error when $\lambda = 12$.
- (c) Draw the power curve for λ between 1 and 20.
- (d) What is the smallest critical value such that the test has level of significance $\alpha \leq 0.05$.

(a)

```
# add your R code
type1 = 1-ppois(13, lambda = 10)
type1
```

```
## [1] 0.1355356
```

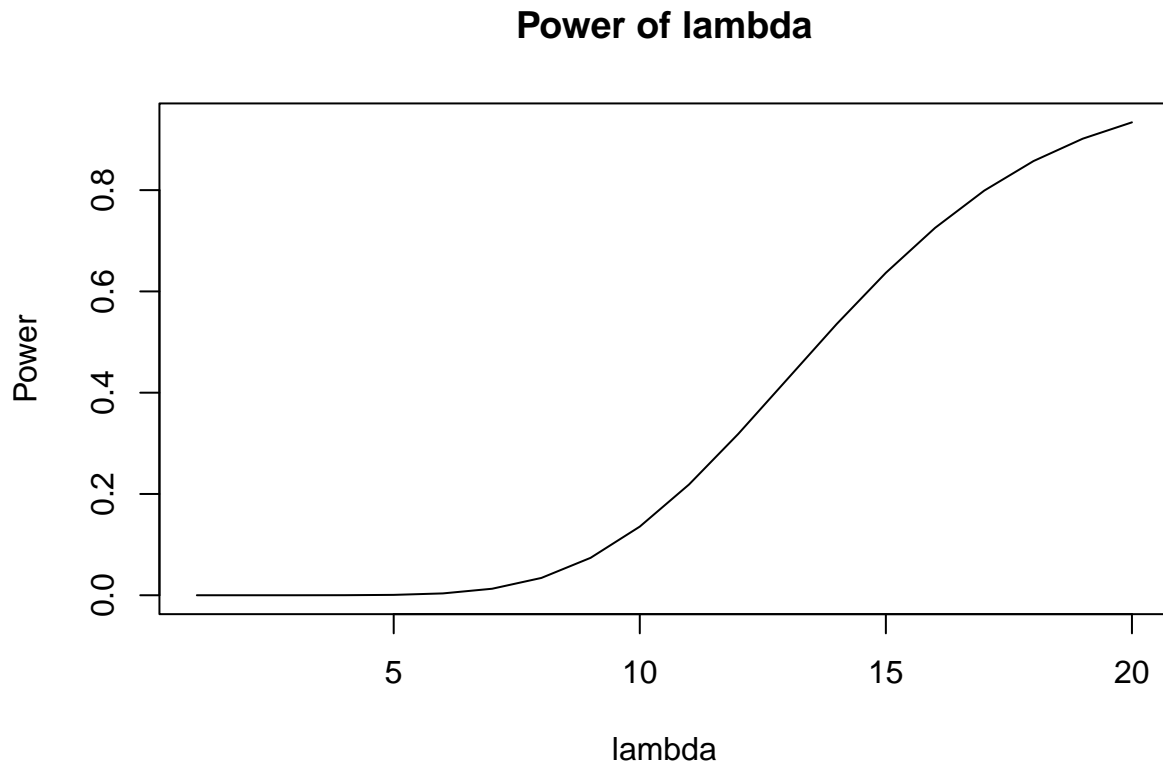
(b)

```
type2 = ppois(13, lambda = 12)
type2
```

```
## [1] 0.6815356
```

(c)

```
lambda_v = 1:20
power_v = 1 - ppois(13, lambda = lambda_v)
plot(lambda_v, power_v, type = 'l', main='Power of lambda', xlab = 'lambda', ylab = 'Power')
```



(d)

```
min_cv = qpois(1-0.05, lambda = 10)
min_cv
```

```
## [1] 15
```

Problem 4.

The hypergeometric distribution is used for random draws *without replacement*. It is implemented in R in the `d/r/p/qhyper` function. Have a look at the help file of this function in R. Suppose we know that an urn contains $m + n = 100$ balls, but we do not know how many white balls m or black balls n it contains. We formulate the null hypothesis $H_0 : m = n = 50$. Drawing $k = 10$ balls without replacement, we observe $x = 2$ white balls.

- Show that the likelihood estimate for m is $\hat{m} = 20$ numerically. Hint: use `which.max` instead of `optimize`.
- Calculate the test statistic and the critical value of the likelihood ratio test statistic for H_0 at $\alpha = 0.05$ numerically. Use the asymptotic distribution of the likelihood ratio test. What is your conclusion about H_0 ?
- Calculate the p-value corresponding to the test of exercise b.

(a)

```
# add your R code
mle = function (m) {
  dhyper(x = 2, m = m, n = 100 - m, k = 10)
}
m_v = 0:100
mle_v = mle(m_v)
names(mle_v) = m_v
names(which.max(mle_v))
```

```
## [1] "20"
```

(b)

```
m_hat = 20
m_0 = 50
lrt = mle(m_hat) - mle(m_0)
lrt
```

```
## [1] 0.2801773
```

```
cv1 = -1*qchisq(0.975, df = 1)/2
cv2 = qchisq(0.975, df = 1)/2
c(cv1, cv2)
```

```
## [1] -2.511943 2.511943
```

Because $cv1 < lrt < cv2$, the conclusion is not reject the null hypothesis.

(c)

```
p_value = 1 - pchisq(2*lrt, df = 1)
p_value
```

```
## [1] 0.4541174
```

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).

$$E(aX + b) = aE(X) + b$$

$$E(X + Y) = E(X) + E(Y)$$

$$\text{Var}(X) = E(X^2) - E(X)^2$$

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

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

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

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	$\text{loglikelihood}(\hat{\theta}) - \text{loglikelihood}(\theta_0)$	$\frac{1}{2}\chi_{\text{df}=1}^2$