PC8 – Hypothesis testing

Octobre 16 2018

In PC7, we have given six steps to build hypothesis tests and study properties of these tests (as the significance level, and the power function). The steps are

(1) write the statistical model,

(2) choose and write the hypotheses,

(3) choose an estimator/a test statistic,

(4) choose the shape of the rejection region,

(5) compute the boundaries of the rejection region using the chosen significance level,

(6) summarize the previous steps by precising the rule of rejection for H_0 and then conclude considering the observed values.

In the following, we are going to define a p-value associated to a test (it could be a 7-th step) and how to conclude from the p-value. Then we will consider a particular test, namely the likelihood ratio test which enables to automatically do steps (3) and (4). We will discuss uniformly most powerful test at the end.

P-values

Let $(Z, \mathcal{Z}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ be a statistical model and z_{obs} the observed value, a realization of Z. We want to test $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$. Consider that the estimator $S(Z)$ is the test statistic we have chosen and $\mathcal{S}_R$ the rejection region for this statistic. The p-value is the probability to observe a $S(Z)$ as extreme or more extreme than the observed one $S(z_{obs})$ under H_0. For instance if the rejection region is one sided and has this shape: $\mathcal{S}_R = [a, +\infty)$ then the p-value is

$$\mathbb{P}_{H_0}\left(S(Z) \geqslant S(z_{obs})\right).$$

Equivalently, the p-value is the smallest significance level at which the null hypothesis would be rejected. Then the chosen test is equivalent to reject H_0 iff the p-value is smaller than the chosen significance level.

Exercise 1: P-values and doping controls

We consider the same doping control as in Exercise 1 of PC6. Namely, during a sport meeting, J.C. and S.R. are subject to an unannounced doping control. Doctors measure their hematocrit levels in their blood. Normally this rate is equal to $\tau_0 = 45\%$ but it can be increased by taking some drug. The measure of this rate is assumed to be Gaussian with a standard deviation of 2. The observed value of J.C. is 48 and the one of S.R. is 50. We want to know if these values are abnormal, i.e. that they have taken a drug, or if they just are the result of the imprecision of the measurements. In PC6, we chose to reject the H_0 hypothesis if the measurement x of the hematocrit level is larger than $\tau_c = 45 + 2q_{1-\alpha} \simeq 48.29$ and we don't reject H_0 otherwise. So we didn't reject H_0^{JC} for J.C. and we have rejected H_0^{SR} for S.R..

1. How do you compute the p-value in the doping experiment?

The rejection region associated to the significance level α is $\mathcal{R}_\alpha = \{x : \ x \geqslant 45 + 2q_{1-\alpha}\}$. Since $\alpha \mapsto q_{1-\alpha}$ is decreasing, then for any $0 < \alpha < \alpha' < 1$, then $\mathcal{R}_\alpha \subset \mathcal{R}_{\alpha'}$ and the p-value of this family of tests, associated to the observed value x_{obs}, is formally defined as the smallest significance level at which the null hypothesis would be rejected:

$$\hat{\alpha}(x_{obs}) = \inf\{\alpha \in [0, 1] : \ x_{obs} \in \mathcal{R}_\alpha\}.$$

In our case

$$\begin{aligned}\inf\{\alpha\in[0,1]:\ x_{obs}\in\mathcal{R}_\alpha\} = \inf\{\alpha\in[0,1]:\ x_{obs}\geqslant 45+2q_{1-\alpha}\} &= \inf\left\{\alpha\in[0,1]:\ \frac{x_{obs}-45}{2}\geqslant q_{1-\alpha}\right\}\\ &= \inf\left\{\alpha\in[0,1]:\ F_{\mathcal{N}(0,1)}\left(\frac{x_{obs}-45}{2}\right)\geqslant F_{\mathcal{N}(0,1)}\left(q_{1-\alpha}\right)\right\}\\ &= \inf\left\{\alpha\in[0,1]:\ F_{\mathcal{N}(0,1)}\left(\frac{x_{obs}-45}{2}\right)\geqslant 1-\alpha\right\}\\ &= 1-F_{\mathcal{N}(0,1)}\left(\frac{x_{obs}-45}{2}\right)\end{aligned}.$$

The p-value is also the probability under H_0 to observe an observation X even more extreme than the one observed x_{obs}, i.e. the probability under H_0 to observe an observation X such that $X\geqslant x_{obs}$, i.e. $\mathbb{P}_{H_0}(X\geqslant x_{obs})$.

$$\mathbb{P}_{H_0}(X\geqslant x_{obs}) = \mathbb{P}_{Z\sim\mathcal{N}(0,1)}\left(Z\geqslant\frac{x_{obs}-45}{2}\right) = 1-F_{\mathcal{N}(0,1)}\left(\frac{x_{obs}-45}{2}\right).$$

2. What are the p-values for J.C. and S.R.?

For J.C.:

$$\hat{\alpha}(48) = 1-F_{\mathcal{N}(0,1)}\left(\frac{48-45}{2}\right) = 1-F_{\mathcal{N}(0,1)}\left(\frac{3}{2}\right)\simeq 0.067.$$

```
1-pnorm(3/2)
```

```
## [1] 0.0668072
```

For S.R.:

$$\hat{\alpha}(50) = 1-F_{\mathcal{N}(0,1)}\left(\frac{50-45}{2}\right) = 1-F_{\mathcal{N}(0,1)}\left(\frac{5}{2}\right)\simeq 0.006.$$

```
1-pnorm(5/2)
```

```
## [1] 0.006209665
```

3. What are your conclusions using the p-values?

The p-value of S.R. $\hat{\alpha}(48)\simeq 0.067$ is larger than the chosen signficance level $\alpha=0.05$, so we don't reject $H_0^{S.R.}$.

The p-value of J.C. $\hat{\alpha}(50)\simeq 0.006$ is smaller than the chosen signficance level $\alpha=0.05$, so we reject $H_0^{J.C.}$.

As expected, we obtain exactly the same results.

Exercise 2: p-values and Student test

Let

$$X_1,\ldots,X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu,\sigma^2).$$

1. We assume that σ is not known and we want to test H_0: $\mu=\mu^*$ versus H_1: $\mu<\mu^*$ for $\mu^*\in\mathbb{R}$.

a. Propose a test at level 0.05 (steps 1 to 6 and as step 7 the computation of the p-value) and create an R function **rej_reg** that takes as input a vector of observations $x=(x_1,\ldots,x_n)$, the parameter μ^* and a level of signficance $\alpha\in(0,1)$, and outputs a vector of size four containing the upper bound of the acceptation region, the statistics of the test, the result of the test: 0 if H_0 is not rejected and 1 if H_0 is rejected and the value of the p-value.

Step 1: The statistical model is

$$\left(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n),\{\mathcal{N}^{\otimes n}(\mu,\sigma^2);\mu\in\mathbb{R},\sigma\in\mathbb{R}_+\}\right).$$

Step 2: The two hypotheses we want to test are $H_0 : \mu = \mu^*$ against $H_1 : \mu < \mu^*$.

Step 3: The MLE of μ is $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Under H_0: $\bar{X}_n \sim \mathcal{N}(\mu^*, \sigma^2/n)$, this distribution depends on σ which is not known. This statistics cannot let us compute the significance level. Thus we use Student/Gosset's Theorem which says that

$$T_n = \sqrt{n}\frac{\bar{X}_n - \mu^*}{S_n}$$

follows a Student distribution with $n-1$ degrees of freedom under H_0, where $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

Step 4: We want to reject the null hypothesis when the estimator of the mean is too small with respect to μ^*. So we choose a rejection region of the following shape: $(-\infty, \delta]$.

Step 5: We choose δ to obtain a test with significance level α, i.e. such that

$$\alpha = \mathbb{P}_{H_0}(T_n \leqslant \delta) = \mathbb{P}_{T_n \sim \mathcal{T}(n-1)}\left(T_n \leqslant \delta\right) = F_{\mathcal{T}(n-1)}\left(\delta\right).$$

So that, we choose $\delta = t_\alpha$ which is the α quantile of a Student distribution with $n-1$ degrees of freedom.

Step 6: Finally our test rejects H_0 when $T_n \leqslant t_\alpha$. Formally,

$$\phi(X_1, \dots, X_n) = \begin{cases} 0 & \text{if } T_n \in (t_\alpha, +\infty) \\ 1 & \text{otherwise} \end{cases}.$$

Step 7: Let t_{obs} be the observed value of the test statistic. Then the p-value is

$$\mathbb{P}_{H_0}(T_n \leqslant t_{obs}) = \mathbb{P}_{T_n \sim \mathcal{T}(n-1)}\left(T_n \leqslant t_{obs}\right) = F_{\mathcal{T}(n-1)}\left(t_{obs}\right).$$

```
rej_rec_2 <- function(x, alpha, mu0){
  n <- length(x)
  t_alpha <- qt(p=alpha,df=n-1)
  mean <- mean(x)
  sd <- sd(x)
  t <- sqrt(n)*(mean-mu0)/sd
  phi <- (t <= t_alpha)
  pvalue <- pt(t,df=n-1)
  test <- c(t_alpha, t, phi, pvalue)
  return(test)
}
```

```
mu0 <- 0
mu <- -2
sigma2 <- 4
alpha <- 0.05
n<-100
x <- rnorm(n,mu,sqrt(sigma2))
rej_rec_2(x,  alpha, mu0)
```

```
## [1] -1.660391e+00 -1.119221e+01  1.000000e+00  1.400648e-19
```

b. Check that it is indeed equivalent to the rejection of H_0 when the test statistic is in the rejection region and when the p-value is smaller than the significance level through simulations when $\sigma = 2$, $\alpha = 0.05$, $\mu^* = 0$, $\mu \in \{0, -0.5\}$ and $n = 100$.

```
mu0 <- 0
mu <-  0
sigma2 <- 4
alpha <- 0.05
n <- 100
```

```r
samples <- lapply(1:1e3, function(i) rnorm(n, mu, sqrt(sigma2)))
tests <- lapply(samples, function(x) rej_rec_2(x, alpha, mu0))
test <- lapply(tests, function(te) c(te[3],(te[4]<=0.05)))
test_equ <- sapply(test, function(te) (te[1]==te[2]))
1e3-sum(test_equ)
```

```
## [1] 0
```

```r
reject <- sapply(tests, function(te) te[3])
sum(reject)/1e3
```

```
## [1] 0.047
```

```r
mu0 <- 0
mu <-  -0.5
sigma2 <- 4
alpha <- 0.05
n<- 100

samples <- lapply(1:1e3, function(i) rnorm(n, mu, sqrt(sigma2)))
tests <- lapply(samples, function(x) rej_rec_2(x, alpha, mu0))
test <- lapply(tests, function(te) c(te[3],(te[4]<=0.05)))
test_equ <- sapply(test, function(te) (te[1]==te[2]))
1e3-sum(test_equ)
```

```
## [1] 0
```

```r
reject <- sapply(tests, function(te) te[3])
sum(reject)/1e3
```

```
## [1] 0.789
```

c. Check that under H_0, the p-values are distributed uniformly on $(0,1)$ through simulations when $\sigma = 2$, $\alpha = 0.05$, $\mu^* = 2$ and $n = 20$.

Similarly to Exercise 2 of PC5, one can see that if the c.d.f. F_X is continuous and strictly increasing,

$$F_X(X) \sim \mathcal{U}([0,1]).$$

Indeed for any $t \in [0,1]$,

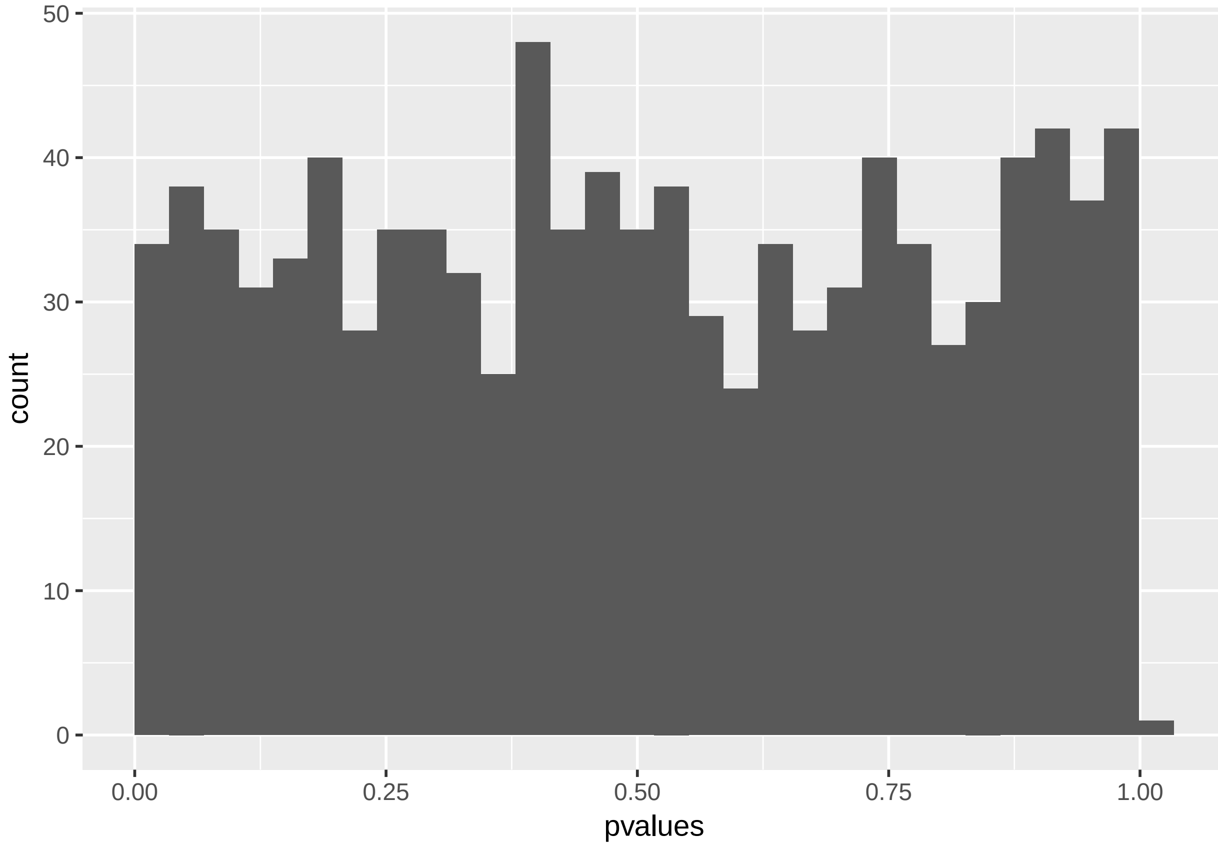
$$\mathbb{P}\left(F_X(X) \leqslant t\right) = \mathbb{P}\left(F_X^{-1}(F_X(X)) \leqslant F_X^{-1}(t)\right) = \mathbb{P}\left(X \leqslant F_X^{-1}(t)\right) = F_X(F_X^{-1}(t)) = t.$$

We can check this property numerically:

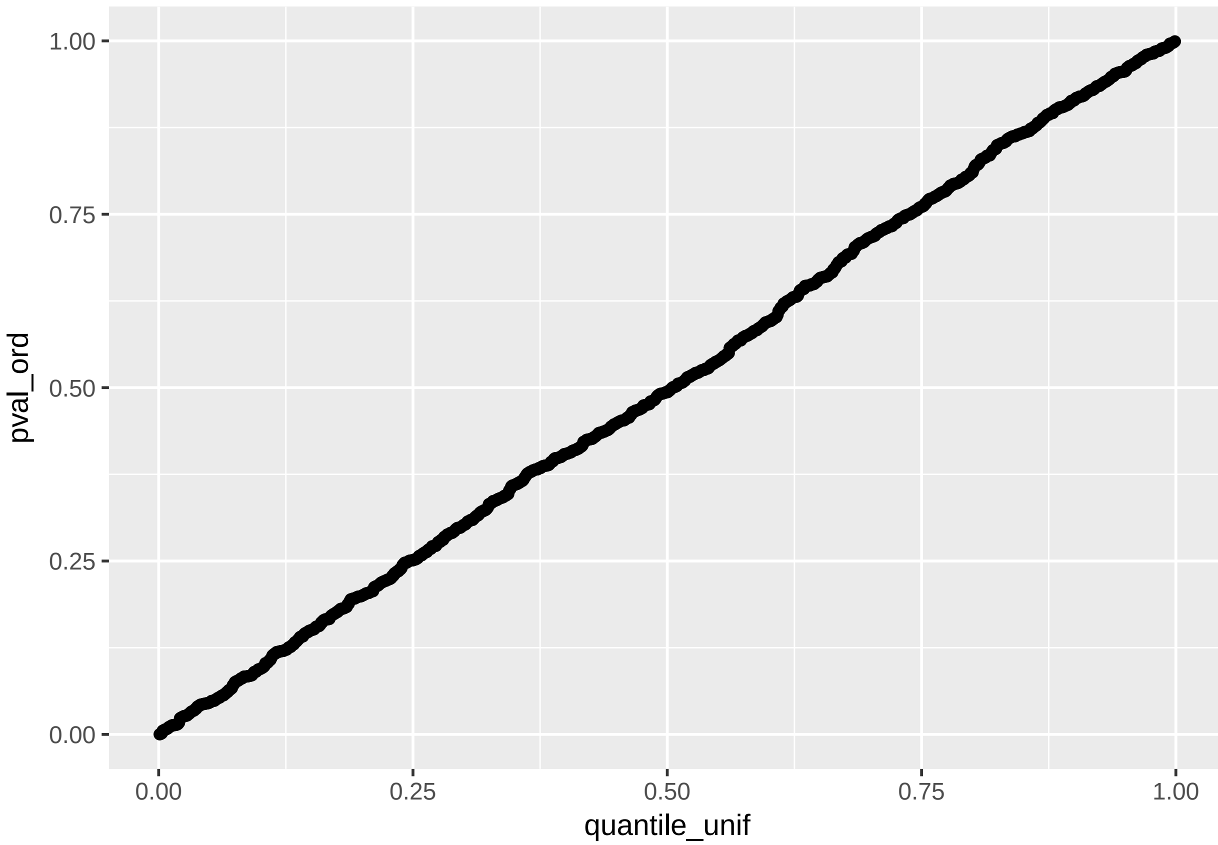
```r
mu0 <- 2
mu <-  mu0
sigma2 <- 4
alpha <- 0.05
n <- 20

samples <- lapply(1:1e3, function(i) rnorm(n, mu, sqrt(sigma2)))
tests <- lapply(samples, function(x) rej_rec_2(x, alpha, mu0))
pvalues <- sapply(tests, function(te) te[4])
data_pvalues <-data.frame(pvalues=pvalues)
library(ggplot2)
ggplot(data=data_pvalues)+geom_histogram(aes(x=pvalues),boundary=0)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
data_pvalues$pval_ord <- sort(pvalues)
N <- 1e3
data_pvalues$quantile_unif <- punif(1/(N+1)*(1:N))
ggplot(data=data_pvalues)+geom_point(aes(x=quantile_unif,y=pval_ord))
```



Likelihood ratio test and Neyman-Pearson

In PC7, we saw that the power function was a good way to compare two hypothesis tests with the same significance level. Here we present a test, namely the likelihood ratio test, which is automatically the best, i.e. with the smallest type II error for a fixed significance level when the hypotheses to test are simple, i.e. $\text{Card}(\Theta_0) = \text{Card}(\Theta_0) = 1$ (Neyman-Pearson Lemma).

Exercise 3: Binomial example

Imagine that there are two coins. The probability of heads of coin 1 is 0.4 and the probability of heads of coin 2 is 0.5. A friend is tossing one of them 5 times but you don't know which coin it is and you would like to find it out.

1. Write the statistical model.

The statistical model is

$$(\{0, 1, \dots, 5\}, \mathcal{P}(\{0, 1, \dots, 5\}), \{Bin(5, p) : p \in \{0.4, 0.5\}\})$$

2. In a dataframe, write the likelihood to obtain x heads for $x \in \{0, 1, \dots, 5\}$ for the both dice.

The likelihood is

$$\mathcal{L}(X; p) = \binom{5}{X} p^X (1-p)^{5-X}.$$

```
d<-data.frame()
d<-rbind(d, dbinom(0:5,size=5,prob=0.4),dbinom(0:5,size=5,prob=0.5))
names(d) <- c("lik0", "lik1" , "lik2", "lik3", "lik4" , "lik5" )
round(d,3)
```

```
##    lik0  lik1  lik2  lik3  lik4  lik5
## 1 0.078 0.259 0.346 0.230 0.077 0.010
## 2 0.031 0.156 0.312 0.312 0.156 0.031
```

3. Compute the likelihood ratio $\mathcal{L}(p = 0.5, x)/\mathcal{L}(p = 0.4, x)$ for $x \in \{0, 1, \ldots, 5\}$. What does it mean if $\mathcal{L}(p = 0.5, x)/\mathcal{L}(p = 0.4, x) = 2$? What do you observe for the likelihood ratio?

```r
round(d[2,]/d[1,],3)
```

```
##    lik0  lik1  lik2  lik3  lik4  lik5
## 2 0.402 0.603 0.904 1.356 2.035 3.052
```

$\mathcal{L}(p = 0.5, x)/\mathcal{L}(p = 0.4, x) = 2$ means that it is twice (2 times) more likely to observe x heads among 5 tosses when the probability to obtain a head is 0.5 than when the probability to obtain a head is 0.4.

The likelihood ratio is increasing when x increases. That is to say, when the number of obtained heads increases, it is more likely that the dice corresponds to $p = 0.5$.

4. Show that $x \in \{0, 1, \ldots, 5\} \mapsto \mathcal{L}(p = 0.5, x)/\mathcal{L}(p = 0.4, x)$ is increasing.

$$\frac{\mathcal{L}(p = 0.5, x)}{\mathcal{L}(p = 0.4, x)} = \frac{\binom{5}{x} 0.5^x 0.5^{5-x}}{\binom{5}{x} 0.4^x 0.6^{5-x}} = \left(\frac{0.6 * 0.5}{0.4 * 0.5}\right)^x \frac{0.5^5}{0.6^5} = \left(\frac{6}{4}\right)^x \frac{0.5^5}{0.6^5}.$$

Since $6/4 > 1$, $\mathcal{L}(p = 0.4, x)/\mathcal{L}(p = 0.5, x)$ is increasing when x is increasing.

5. Enumerate all the possible rejection regions to test $H_0 :\ p = 0.4$ against $H_1 :\ p = 0.5$, where p is the probability to obtain a head.

There are $2^6 = 64$ possible rejection regions: for each integer among $\{0, 1, \ldots, 5\}$, we can choose if it is in the rejection region or not.

Possible rejection region: $\emptyset$, $\{0\}$, $\{1\}$,$\{0,1\}$, $\{2\}$, $\{0,2\}$, $\{1,2\}$, $\{0,1,2\}$, $\{3\}$, $\{0,3\}$, $\{1,3\}$, $\{0,1,3\}$, $\{2,3\}$, $\{0,2,3\}$, $\{1,2,3\}$, $\{0,1,2,3\}$, $\{4\}$, $\{0,4\}$, $\{1,4\}$,$\{0,1,4\}$, $\{2,4\}$, $\{0,2,4\}$, $\{1,2,4\}$, $\{0,1,2,4\}$, $\{3,4\}$, $\{0,3,4\}$, $\{1,3,4\}$, $\{0,1,3,4\}$, $\{2,3,4\}$, $\{0,2,3,4\}$, $\{1,2,3,4\}$, $\{0,1,2,3,4\}$, $\{5\}$, $\{0,5\}$, $\{1,5\}$,$\{0,1,5\}$, $\{2,5\}$, $\{0,2,5\}$, $\{1,2,5\}$, $\{0,1,2,5\}$, $\{3,5\}$, $\{0,3,5\}$, $\{1,3,5\}$, $\{0,1,3,5\}$, $\{2,3,5\}$, $\{0,2,3,5\}$, $\{1,2,3,5\}$, $\{0,1,2,3,5\}$, $\{4,5\}$, $\{0,4,5\}$, $\{1,4,5\}$,$\{0,1,4,5\}$, $\{2,4,5\}$, $\{0,2,4,5\}$, $\{1,2,4,5\}$, $\{0,1,2,4,5\}$, $\{3,4,5\}$, $\{0,3,4,5\}$, $\{1,3,4,5\}$, $\{0,1,3,4,5\}$, $\{2,3,4,5\}$, $\{0,2,3,4,5\}$, $\{1,2,3,4,5\}$, $\{0,1,2,3,4,5\}$.

6. Compute the significiance level and the power of all possible tests. Among the tests with a level smaller than 0.0103, 0.0871 and 0.3175 which one has the largest power?

```r
enum_tests <- data.frame("0"=c(rep(0,32),rep(1,32)),
                         "1"=rep(c(rep(0,16),rep(1,16)),2),
                         "2"=rep(c(rep(0,8),rep(1,8)),4),
                         "3"=rep(c(rep(0,4),rep(1,4)),8),
                         "4"=rep(c(rep(0,2),rep(1,2)),16),
                         "5"=rep(c(0,1),32))
enum_tests$level <- enum_tests$X0*dbinom(0,size=5,prob=0.4) +
  enum_tests$X1*dbinom(1,size=5,prob=0.4) +
  enum_tests$X2*dbinom(2,size=5,prob=0.4) +
  enum_tests$X3*dbinom(3,size=5,prob=0.4) +
  enum_tests$X4*dbinom(4,size=5,prob=0.4) +
  enum_tests$X5*dbinom(5,size=5,prob=0.4)
enum_tests$power <- enum_tests$X0*dbinom(0,size=5,prob=0.5) +
  enum_tests$X1*dbinom(1,size=5,prob=0.5) +
  enum_tests$X2*dbinom(2,size=5,prob=0.5) +
  enum_tests$X3*dbinom(3,size=5,prob=0.5) +
```

```r
  enum_tests$X4*dbinom(4,size=5,prob=0.5) +
  enum_tests$X5*dbinom(5,size=5,prob=0.5)
enum_tests
```

```
##    X0 X1 X2 X3 X4 X5   level   power
## 1   0  0  0  0  0  0 0.00000 0.00000
## 2   0  0  0  0  0  1 0.01024 0.03125
## 3   0  0  0  0  1  0 0.07680 0.15625
## 4   0  0  0  0  1  1 0.08704 0.18750
## 5   0  0  0  1  0  0 0.23040 0.31250
## 6   0  0  0  1  0  1 0.24064 0.34375
## 7   0  0  0  1  1  0 0.30720 0.46875
## 8   0  0  0  1  1  1 0.31744 0.50000
## 9   0  0  1  0  0  0 0.34560 0.31250
## 10  0  0  1  0  0  1 0.35584 0.34375
## 11  0  0  1  0  1  0 0.42240 0.46875
## 12  0  0  1  0  1  1 0.43264 0.50000
## 13  0  0  1  1  0  0 0.57600 0.62500
## 14  0  0  1  1  0  1 0.58624 0.65625
## 15  0  0  1  1  1  0 0.65280 0.78125
## 16  0  0  1  1  1  1 0.66304 0.81250
## 17  0  1  0  0  0  0 0.25920 0.15625
## 18  0  1  0  0  0  1 0.26944 0.18750
## 19  0  1  0  0  1  0 0.33600 0.31250
## 20  0  1  0  0  1  1 0.34624 0.34375
## 21  0  1  0  1  0  0 0.48960 0.46875
## 22  0  1  0  1  0  1 0.49984 0.50000
## 23  0  1  0  1  1  0 0.56640 0.62500
## 24  0  1  0  1  1  1 0.57664 0.65625
## 25  0  1  1  0  0  0 0.60480 0.46875
## 26  0  1  1  0  0  1 0.61504 0.50000
## 27  0  1  1  0  1  0 0.68160 0.62500
## 28  0  1  1  0  1  1 0.69184 0.65625
## 29  0  1  1  1  0  0 0.83520 0.78125
## 30  0  1  1  1  0  1 0.84544 0.81250
## 31  0  1  1  1  1  0 0.91200 0.93750
## 32  0  1  1  1  1  1 0.92224 0.96875
## 33  1  0  0  0  0  0 0.07776 0.03125
## 34  1  0  0  0  0  1 0.08800 0.06250
## 35  1  0  0  0  1  0 0.15456 0.18750
## 36  1  0  0  0  1  1 0.16480 0.21875
## 37  1  0  0  1  0  0 0.30816 0.34375
## 38  1  0  0  1  0  1 0.31840 0.37500
## 39  1  0  0  1  1  0 0.38496 0.50000
## 40  1  0  0  1  1  1 0.39520 0.53125
## 41  1  0  1  0  0  0 0.42336 0.34375
## 42  1  0  1  0  0  1 0.43360 0.37500
## 43  1  0  1  0  1  0 0.50016 0.50000
## 44  1  0  1  0  1  1 0.51040 0.53125
## 45  1  0  1  1  0  0 0.65376 0.65625
## 46  1  0  1  1  0  1 0.66400 0.68750
## 47  1  0  1  1  1  0 0.73056 0.81250
## 48  1  0  1  1  1  1 0.74080 0.84375
## 49  1  1  0  0  0  0 0.33696 0.18750
```

```
## 50  1  1  0  0  0  1 0.34720 0.21875
## 51  1  1  0  0  1  0 0.41376 0.34375
## 52  1  1  0  0  1  1 0.42400 0.37500
## 53  1  1  0  1  0  0 0.56736 0.50000
## 54  1  1  0  1  0  1 0.57760 0.53125
## 55  1  1  0  1  1  0 0.64416 0.65625
## 56  1  1  0  1  1  1 0.65440 0.68750
## 57  1  1  1  0  0  0 0.68256 0.50000
## 58  1  1  1  0  0  1 0.69280 0.53125
## 59  1  1  1  0  1  0 0.75936 0.65625
## 60  1  1  1  0  1  1 0.76960 0.68750
## 61  1  1  1  1  0  0 0.91296 0.81250
## 62  1  1  1  1  0  1 0.92320 0.84375
## 63  1  1  1  1  1  0 0.98976 0.96875
## 64  1  1  1  1  1  1 1.00000 1.00000
```

In the following we are going to consider pure and randomized tests. A pure test has this form

$$\phi(X) = \begin{cases} 1 & \text{if } X \text{ is in the rejection region} \\ 0 & \text{otherwise} \end{cases}.$$

```
enum_tests[enum_tests$level<=0.0103,]
```

```
##   X0 X1 X2 X3 X4 X5   level   power
## 1  0  0  0  0  0  0 0.00000 0.00000
## 2  0  0  0  0  0  1 0.01024 0.03125
```

Among the pure tests with significance level smaller than 0.0103, the test which rejects H_0 when $x = 5$ has the largest power.

```
enum_tests[enum_tests$level<=0.0871,]
```

```
##    X0 X1 X2 X3 X4 X5   level   power
## 1   0  0  0  0  0  0 0.00000 0.00000
## 2   0  0  0  0  0  1 0.01024 0.03125
## 3   0  0  0  0  1  0 0.07680 0.15625
## 4   0  0  0  0  1  1 0.08704 0.18750
## 33  1  0  0  0  0  0 0.07776 0.03125
```

Among the pure tests with significance level smaller than 0.0871, the test which rejects H_0 when $x \geqslant 4$ has the largest power.

```
enum_tests[enum_tests$level<=0.3175,]
```

```
##    X0 X1 X2 X3 X4 X5   level   power
## 1   0  0  0  0  0  0 0.00000 0.00000
## 2   0  0  0  0  0  1 0.01024 0.03125
## 3   0  0  0  0  1  0 0.07680 0.15625
## 4   0  0  0  0  1  1 0.08704 0.18750
## 5   0  0  0  1  0  0 0.23040 0.31250
## 6   0  0  0  1  0  1 0.24064 0.34375
## 7   0  0  0  1  1  0 0.30720 0.46875
## 8   0  0  0  1  1  1 0.31744 0.50000
## 17  0  1  0  0  0  0 0.25920 0.15625
## 18  0  1  0  0  0  1 0.26944 0.18750
## 33  1  0  0  0  0  0 0.07776 0.03125
## 34  1  0  0  0  0  1 0.08800 0.06250
## 35  1  0  0  0  1  0 0.15456 0.18750
```

```
## 36   1  0  0  0  1  1 0.16480 0.21875
## 37   1  0  0  1  0  0 0.30816 0.34375
```

Among the pure tests with significance level smaller than 0.3175, the test which rejects H_0 when $x \geqslant 3$ has the largest power.

7. Among the tests with a level smaller than 0.25 which one has the largest power? Compute the level and the power of the randomized test

$$\phi(X) = \begin{cases} 1 & \text{if } X > 3 \\ \frac{0.25-\mathbb{P}_{Z\sim Bin(5,0.4)}(Z>3)}{\mathbb{P}_{Z\sim Bin(5,0.4)}(Z=3)} & \text{if } X = 3 \\ 0 & \text{otherwise} \end{cases}.$$

A randomized test is a (measurable) function on $\mathcal{X}$ in $[0,1]$. We reject H_0 with $\phi(X)$ probability.

```
enum_tests[enum_tests$level<=0.25,]
```

```
##     X0 X1 X2 X3 X4 X5   level   power
## 1    0  0  0  0  0  0 0.00000 0.00000
## 2    0  0  0  0  0  1 0.01024 0.03125
## 3    0  0  0  0  1  0 0.07680 0.15625
## 4    0  0  0  0  1  1 0.08704 0.18750
## 5    0  0  0  1  0  0 0.23040 0.31250
## 6    0  0  0  1  0  1 0.24064 0.34375
## 33   1  0  0  0  0  0 0.07776 0.03125
## 34   1  0  0  0  0  1 0.08800 0.06250
## 35   1  0  0  0  1  0 0.15456 0.18750
## 36   1  0  0  0  1  1 0.16480 0.21875
```

Among the pure tests with significance level smaller than 0.25, the test which rejects H_0 when $x \in \{3,5\}$ has the largest power (around 0.34).

The level of the randomized test ϕ is

$$\mathbb{E}_{X\sim Bin(5,0.4)}(\phi) = \mathbb{P}_{X\sim Bin(5,0.4)}(X>3) + \frac{0.25-\mathbb{P}_{Z\sim Bin(5,0.4)}(Z>3)}{\mathbb{P}_{Z\sim Bin(5,0.4)}(Z=3)}\mathbb{P}_{X\sim Bin(5,0.4)}(X=3) = 0.25.$$

Its power is

$$\mathbb{E}_{X\sim Bin(5,0.5)}(\phi) = \mathbb{P}_{X\sim Bin(5,0.5)}(X>3) + \frac{0.25-\mathbb{P}_{Z\sim Bin(5,0.4)}(Z>3)}{\mathbb{P}_{Z\sim Bin(5,0.4)}(Z=3)}\mathbb{P}_{X\sim Bin(5,0.5)}(X=3) \simeq 0.41 > 0.34.$$

```
1-pbinom(3,size=5,prob=0.5) +
  (0.25-1+pbinom(3,size=5,prob=0.4))/(dbinom(3,size=5,prob=0.4)) * dbinom(3,size=5,prob=0.5)
```

```
## [1] 0.4085286
```

This randomized test is more powerful than the best pure test. Actually ϕ is the most powerful test at level at most 0.25 using Neyman-Pearson Theorem.

8. Give the most powerful test at level 0.05 to test $H_0: \ p = 0.4$ against $H_1: \ p = 0.5$.

Using Neyman-Pearson Theorem, the likelihood ration test is uniformly most powerful. The likelihood ratio test at level $\alpha = 0.05$ is

$$\tilde{\phi}_{0.05}(X) = \begin{cases} 1 & \text{if } \frac{\mathcal{L}(p=0.5,X)}{\mathcal{L}(p=0.4,X)} > c_{0.05} \\ \gamma_{0.05} & \text{if } \frac{\mathcal{L}(p=0.5,X)}{\mathcal{L}(X)} = c_{0.05} \\ 0 & \text{otherwise} \end{cases},$$

where $\gamma_{0.05} \in (0,1)$ and $c_{0.05}$ has to be chosen such that the significance level equals to 0.05, is among the most powerful test at level 0.05.

Since the likelihood ratio $x \mapsto \tilde{\phi}_{0.05}(x)$ is increasing with x, the previous test is equivalent to

$$\phi_{0.05}(X) = \begin{cases} 1 & \text{if } X > k_{0.05} \\ \gamma_{0.05} & \text{if } X = k_{0.05} \\ 0 & \text{otherwise} \end{cases},$$

where $\gamma_{0.05}$ and $k_{0.05}$ has to be chosen such that the significance level equals to 0.05.

The significance level of $\phi_{0.05}$ is

$$\mathbb{E}_{X\sim Bin(5,0.4)}(\phi_{0.05}) = \mathbb{P}_{X\sim Bin(5,0.4)}(X > k_{0.05}) + \gamma_{0.05}\mathbb{P}_{X\sim Bin(5,0.4)}(X = k_{0.05}).$$

We first choose the largest $k_{0.05}$ such that $\mathbb{P}_{X\sim Bin(5,0.4)}(X > k_{0.05}) \leqslant 0.05$ and then solve for $\gamma_{0.05}$,

$$0.05 = \mathbb{P}_{X\sim Bin(5,0.4)}(X > k_{0.05}) + \gamma_{0.05}\mathbb{P}_{X\sim Bin(5,0.4)}(X = k_{0.05}).$$

Since $\mathbb{P}_{X\sim Bin(5,0.4)}(X > 5) = 0$, $\mathbb{P}_{X\sim Bin(5,0.4)}(X > 4) \simeq 0.01$, $\mathbb{P}_{X\sim Bin(5,0.4)}(X > 3) \simeq 0.08$, $\mathbb{P}_{X\sim Bin(5,0.4)}(X > 2) \simeq 0.32$, $\mathbb{P}_{X\sim Bin(5,0.4)}(X > 1) \simeq 0.66$, $\mathbb{P}_{X\sim Bin(5,0.4)}(X > 0) \simeq 0.92$, $\mathbb{P}_{X\sim Bin(5,0.4)}(X > -1) \simeq 1$, then $k_{0.05}$ has to equal 4.

```
1-pbinom(seq(5,0,by=-1),size=5,prob=0.4)
```

```
## [1] 0.00000 0.01024 0.08704 0.31744 0.66304 0.92224
```

Then $\gamma_{0.05} = \frac{0.05-\mathbb{P}_{Z\sim Bin(5,0.4)}(Z>4)}{\mathbb{P}_{Z\sim Bin(5,0.4)}(Z=4)} \simeq 0.52$.

```
(0.05 - 1+pbinom(4,size=5,prob=0.4))/(dbinom(4,size=5,prob=0.4))
```

```
## [1] 0.5177083
```

Exercise 4: Swimming pool

A swimming pool seller wants to compare two products which kills bacteria. Both products guarantee that 95% of bacteria are eliminated. Yet the pH may become more basic so more harmful for users. Thus he does an experiment with two swimming pools in his shop. He puts the product A in pool 1 and product B in pool 2. He has a pH meter but he doubts its reliability. Therefore he does 10 measurements in each swimming pool.

In pool 1, he observes $x_1, ..., x_{10}$ with respective values

7.33 ; 6.17 ; 7.46 ; 8.13 ; 6.68 ; 6.76 ; 7.97 ; 6.76 6.81 ; 8.40 .

The empirical mean is $\bar{x}_{10} = 7.247$ and the empirical variance (the one divided by $n-1$) is $\sqrt{\sigma_x^2} = 0.73$.

In pool 2, he observes $y_1, ..., y_{10}$ with respective values

10.40 ; 7.27 ; 8.99 ; 7.28 ; 9.18 ; 9.10 ; 7.96 ; 7.71 ; 9.59 ; 9.61

The empirical mean is $\bar{y}_{10} = 8.709$ and the empirical variance (the one divided by $n-1$) is $\sqrt{\sigma_y^2} = 1.08$.

The seller does not have any preference beforehand and wants to give the best advice to his customers. He wants to clearly say "prefer A", "prefer B" or "do as you wish". We know that $pH = 7$ is neutral for the skin while $pH = 9$ is basic and harmful.

1. Assume that the observations are realizations of independent Gaussian random variables with variance σ^2 and mean m_1 for pool 1, and m_2 for pool 2. Give the statistical model considering a Gaussian vector. Compute the mean and the covariance matrix.
2. Assume that $\sigma^2 = 1$ is known.

a. Give a likelihood ratio test $H_0 \ : \ m_1 = m_2 = 7$ against $H_1 \ : \ m_1 = 7, \ m_2 = 9$ at level α. Simplify the result as much as possible. Compute the p-value.

b. Explain why this test is uniformly most powerful among the hypothesis tests at level α which test the same hypotheses.

3. We still assume that $\sigma^2 = 1$ is known, test $H_0 \ m_1 = m_2$ against $H_1 \ : \ m_1 < m_2$ at level α.

4. Now, we assume that σ^2 is unknown (more realistic for this problem statement).

a. Give the distribution of $Z = \bar{Y} - \bar{X}$ and of $W = 9\sigma_x^2 + 9\sigma_y^2$. We admit that W and Z are independent.

b. Give a hypothesis test $H_0 \ : \ m_1 = m_2$ against $H_1 \ : \ m_1 < m_2$ at level α. Compute the p-value.

c. What should be the conclusion of the seller?

UMP hypotheses tests for composite hypotheses

Exercise 5: Test for the variance in Gaussian models

Let $(X_1, \dots, X_n)$ be a n-sample of a normal distribution $\mathcal{N}(\mu_0, \sigma^2)$, where μ_0 is known and σ^2 is unknown.

1. Give the statistical model.

The vector of observations $Z = (X_1, \dots, X_n)$ is in $\mathbb{R}^n$ (with the sigma-field $\mathcal{B}(\mathbb{R}^n)$). The unknown parameter here is $\theta = \sigma^2 \in \Theta := \mathbb{R}_+$. And the distribution P_θ, parametrized by $\theta = \sigma^2$, of Z has the following density function (w.r.t the Lebesgue-measure):

$$z = (x_1, \dots, x_n) \to p(\sigma^2, z) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu_0)^2}{2\sigma^2}\right)$$

2. Compute explicitly the likelihood ratio test of size at most α for testing $H_0 \ : \ \sigma^2 = \sigma_0^2$ against $H_1 \ : \ \sigma^2 = \sigma_1^2$, where $\sigma_0^2 < \sigma_1^2$.

The likelihood ratio is

$$r(z) = \frac{p(\sigma_1^2, z)}{p(\sigma_0^2, z)} = \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left(\frac{1}{2}\Big(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\Big) \sum_{i=1}^{n} (x_i - \mu_0)^2\right).$$

The likelihood ration test, in the case of continuous observations has the following form:

$$\phi_\alpha(z) = \begin{cases} 1 & \text{if } r(z) > c_\alpha \\ 0 & \text{otherwise} \end{cases},$$

where c_α is chosen to get a level α of the test.

As $\sigma_0^2 < \sigma_1^2$, r is an increasing function of $\sum_{i=1}^n (x_i - \mu_0)^2$. So the likelihood ratio test rejects H_0 when $\sum_{i=1}^n (x_i - \mu_0)^2$ is large, i.e. $\sum_{i=1}^n (x_i - \mu_0)^2 > c$ for some constant c, we do not reject H_0 otherwise. We compute explicitly c by controlling the type I error:

$$\alpha = P_{\sigma_0^2}\left(\sum_{i=1}^{n} (X_i - \mu_0)^2 > c\right) = P_{\sigma_0^2}\left(\sum_{i=1}^{n} \frac{(X_i - \mu_0)^2}{\sigma_0^2} > \frac{c}{\sigma_0^2}\right) = P_{S \sim \chi^2(n)}\left(S > \frac{c}{\sigma_0^2}\right).$$

So that $\frac{c}{\sigma_0^2} = \chi^2_{n,1-\alpha}$. And the likelihood ratio test is

$$\phi_\alpha(z) = \begin{cases} 1 & \text{if } \sum_{i=1}^n (x_i - \mu_0)^2 > \sigma_0^2 \chi^2_{n,1-\alpha} \\ 0 & \text{otherwise} \end{cases}.$$

a. Is this test UMP(α) for testing $H_0: \ \sigma^2 = \sigma_0^2$ against $H_1: \ \sigma^2 = \sigma_1^2$ when $\sigma_0^2 < \sigma_1^2$?

Yes it is a UMP(α) for testing $H_0: \ \sigma^2 = \sigma_0^2$ against $H_1: \ \sigma^2 = \sigma_1^2$ using Neyman-Pearson Lemma.

b. Is this test UMP(α) for testing $H_0': \ \sigma^2 \le \sigma_0^2$ against $H_1': \ \sigma^2 > \sigma_0^2$?

We first compute the power function associated to this test:

$$\begin{aligned}\beta_{\phi_\alpha}(\sigma^2) = \mathbb{E}_{\sigma^2}(\phi_\alpha) = \mathbb{P}_{\sigma^2}\left(\sum_{i=1}^n (x_i - \mu_0)^2 > \sigma_0^2 \chi^2_{n,1-\alpha}\right) = P_{\sigma^2}\left(\sum_{i=1}^n \frac{(X_i - \mu_0)^2}{\sigma^2} > \frac{\sigma_0^2 \chi^2_{n,1-\alpha}}{\sigma^2}\right) \\ = P_{S \sim \chi^2(n)}\left(S > \frac{\sigma_0^2 \chi^2_{n,1-\alpha}}{\sigma^2}\right).\end{aligned}$$

To prove that ϕ_α is UMP(α) we want to prove that (1) $\sup_{\sigma^2 \le \sigma_0^2} \mathbb{E}_{\sigma^2}(\phi_\alpha(Z)) \le \alpha$, (2) for all test ψ such that $\sup_{\sigma^2 \le \sigma_0^2} \mathbb{E}_\theta(\psi(Z)) \le \alpha$ then for all $\sigma^2 > \sigma_0^2$

$$\mathbb{E}_{\sigma^2}(\psi(Z)) \le \mathbb{E}_{\sigma^2}(\phi_\alpha(Z)).$$

Using that the power function β_{ϕ_α} of ϕ_α is an increasing function of σ^2, we have $\sup_{\sigma^2 \le \sigma_0^2} \mathbb{E}_{\sigma^2}(\phi_\alpha(Z)) = \mathbb{E}_{\sigma_0^2}(\phi_\alpha(Z)) = \alpha$. So (1) is true.

Let ψ be a test such that $\sup_{\sigma^2 \le \sigma_0^2} \mathbb{E}_\theta(\psi(Z)) \le \alpha$ and $\sigma^2 > \sigma_0^2$. Notice that ϕ_α defined in the previous question does not depend on σ^2 as long as $\sigma^2 > \sigma_0^2$. Using Neyman-Pearson Lemma, for any fixed $\sigma^2 > \sigma_0^2$, for all test ψ such that $\sup_{\sigma^2 \le \sigma_0^2} \mathbb{E}_\theta(\psi(Z)) \le \alpha$ then

$$\mathbb{E}_{\sigma^2}(\psi(Z)) \le \mathbb{E}_{\sigma^2}(\phi_\alpha(Z)).$$

This inequality is true for any $\sigma^2 > \sigma_0^2$ so that (2) is true.

In toher words, ϕ_α is UMP(α) for testing $H_0': \ \sigma^2 \le \sigma_0^2$ against $H_1': \ \sigma^2 > \sigma_0^2$.

Exercise 6: Likelihood ratio test in exponential models

Let $(X_1, \ldots, X_n)$ n be n i.i.d. random variables distributed from an exponential distribution with parameter $\theta \in \Theta = (0, +\infty) =: \mathbb{R}_+^*$,

$$p(\theta; x) = \theta^{-1} \exp(-x/\theta) \mathbb{1}_{\mathbb{R}_+}(x)\,.$$

Let $\theta_0 > 0$. We want to test

$$H_0: \ \theta \le \theta_0 \ \text{ against } \ H_1: \ \theta > \theta_0\,.$$

We first consider

$$H_0: \ \theta = \theta_0 \ \text{ against } \ H_1: \ \theta = \theta_1\,,$$

where $\theta_1 > \theta_0$.

1. Show that the likelihood ratio is an increasing function of some statistic, determine what the statistic is.

The likelihood ratio is

$$r(z) = \frac{\prod_{i=1}^n p(\theta_1, x_i)}{\prod_{i=1}^n p(\theta_0, x_i)} = \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left(\sum_{i=1}^n x_i \left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\right),$$

for $z = (x_1, \ldots, x_n) \in \left(\mathbb{R}_+^*\right)^n$. As $\theta_0 < \theta_1$, r is an increasing function of the statistics $\sum_{i=1}^n x_i$.

2. Give a UMP(α) test. The constants, in the definition of the test, have to be explicitly given. You can use that if $X_i \sim \text{Gamma}(a_i, b)$ independently and c is a real constant then $\sum_i X_i \sim \text{Gamma}(\sum_i a_i, b)$ and $cX_i \sim \text{Gamma}(a_i, b/c)$, that a $\exp(\lambda)$ distribution is a $\text{Gamma}(1, 1/\lambda)$ distribution and that a $\text{Gamma}(n, 1/2)$ distribution is a $\chi^2(2n)$ distribution.

Using Neyman-Pearson Lemma, a UMP(α) test is

$$\phi_\alpha(z) = \begin{cases} 1 & \frac{p_1(z)}{p_0(z)} > c_\alpha \\ \gamma_\alpha & \frac{p_1(z)}{p_0(z)} = c_\alpha \\ 0 & \frac{p_1(z)}{p_0(z)} < c_\alpha \end{cases}$$

where c_α is chosen such that the type I error is α:

$$\mathbb{E}_{\theta_0}(\phi_\alpha(Z)) = \alpha.$$

As $\sum_i X_i$ is a random variable which admits a density w.r.t. the Lebesgue measure and using Question 1, then the UMP(α) test has the from:

$$\phi_\alpha(Z) = \begin{cases} 1 & \text{if } \sum_i X_i > c_\alpha \\ 0 & \text{otherwise.} \end{cases}$$

We now compute c_α. Under $\mathbb{P}_{\theta_0}$, $\sum_i X_i \sim \text{Gamma}(n, 1/\theta_0)$, so $(2\sum_i X_i)/\theta_0 \sim \chi^2(2n)$. Then

$$\mathbb{E}_{\theta_0}(\phi_\alpha(Z)) = \mathbb{P}_{\theta_0}\left(\sum_i X_i \geq c_\alpha\right) = P_{Y\sim\chi^2(2n)}(Y \geq 2c_\alpha/\theta_0) = \alpha$$

when $c_\alpha = \theta_0\chi^2_{2n,1-\alpha}/2$.

3. Compute the power function associated to this test.

The power function is

$$\beta_\phi(\theta) = \mathbb{E}_\theta(\phi(Z)) = \mathbb{P}_\theta\left(\sum_i X_i \geq \frac{\theta_0\chi^2_{2n,1-\alpha}}{2}\right) = \mathbb{P}_{Y\sim\chi^2(2n)}\left(Y \geq \frac{\theta_0\chi^2_{2n,1-\alpha}}{\theta}\right).$$

4. Is the test built in Question 2 also a UMP(α) test for testing

$$H_0: \ \theta \leq \theta_0 \ \text{ against } \ H_1: \ \theta = \theta_1 \,.$$

A test ϕ^* is UMP(α) if (1) $\sup_{\theta\in\Theta_0} \mathbb{E}_\theta(\phi^*(Z)) \leq \alpha$, (2) for all test ψ such that $\sup_{\theta\in\Theta_0} \mathbb{E}_\theta(\psi(Z)) \leq \alpha$ then for all $\theta \in \Theta_1$

$$\mathbb{E}_\theta(\psi(Z)) \leq \mathbb{E}_\theta(\phi^*(Z)).$$

Using Question 3, the power function of ϕ is an increasing function of θ. So that

$$\sup_{\theta\in\Theta_0} \mathbb{E}_\theta(\phi(Z)) = \mathbb{E}_{\theta_0}(\phi(Z)) = \alpha$$

and (1) is true. Moreover using Neyman-Pearson Lemma, (2) is also true.

Then, yes, ϕ is a UMP(α) test for

$$H_0: \ \theta \leq \theta_0 \ \text{ against } \ H_1: \ \theta = \theta_1 \,.$$

5. Is the test built in Question 2 also a UMP(α) test for

$$H_0: \ \theta = \theta_0 \ \text{ against } \ H_1: \ \theta > \theta_0 \,.$$

(1) is true. Moreover the test ϕ does not depend on θ_1, provided that $\theta_1 > \theta_0$. So using the Neyman-Pearson Lemma, ϕ is the most powerful test for

$$H_0: \ \theta = \theta_0 \ \text{ against } \ H_1: \ \theta = \theta_1 \,.$$

for all $\theta_1 > \theta_0$. Then (2) is also true.

And yes, ϕ is a UMP(α) test for

$$H_0 : \ \theta = \theta_0 \ \text{ against } \ H_1 : \ \theta > \theta_0 \,.$$

6. Is the test built in Question 2 also a UMP(α) test for

$$H_0 : \ \theta \leq \theta_0 \ \text{ against } \ H_1 : \ \theta > \theta_0 \,.$$

Using the arguments of question 4 and 5, ϕ is a UMP(α) test for

$$H_0 : \ \theta \leq \theta_0 \ \text{ against } \ H_1 : \ \theta > \theta_0 \,.$$