# PC7 – Hypothesis testing

*Octobre 22 2018*

Unlike point estimators or confidence interval which give numerical values, we may be interested in answering yes/no questions or choose between two contradictorial theories. Hypothesis testing answer these kind of questions. The two theories (or the yes and the no), called hypotheses, won't have the same role. One will be called the null hypothesis $H_0$ and the other the alternative hypothesis $H_1$. The way we choose between the two hypotheses is conceptually similar to the way a jury deliberates in a court trial. $H_0$ is going to be the defendant: just as the latter is presumed innocent until "proven" guilty, so is the null hypothesis "accepted" unless the data argue overwhelmingly to the contrary. In short, $H_1$ should be the hypothesis you want to prove.

## Introduction to hypothesis testing

### Exercise 1: Hypothesis testing and doping controls

During a sport meeting, J.C. and S.R. are subject to an unnannounced 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.

1. The **first step** for solving a hypothesis testing is to write the **statistical model**. Precise the statistical model for the doping control of J.C. and S.R..

Let $x_{JC}$ and $x_{SR}$ the measured hematocrit levels of J.C. and S.R., we assume that they are realizations of $X_{JC}$ and $X_{SR}$ Gaussian random variables distributed from $\mathcal{N}(\tau_{JC}, 4)$ and $\mathcal{N}(\tau_{SR}, 4)$, where $\tau_{JC}$ and $\tau_{SR}$ are the unknown true levels of hematocrit in J.C.'s and S.R.'s blood. He have observed $x_{JC} = 48$ and $x_{SR} = 50$.

2. The **second step** is to choose and write the **hypotheses**. They are formally written as equations on the parameters of the model. To choose between $H_0$ and $H_1$, we apply the same rule as in a court trial with $H_0$ being typically a statement reflecting the status quo. And we give to $H_0$ the benefit of the doubt. Write the hypotheses for the doping control.

We want to know if J.C. and S.R. are doped. To do so we want to know if their hematocrit levels are normal. We want to give the benefit of the doubt to the innoncence of the athletes. So $H_0$ is going to represent the fact that the athlete is not doped, i.e. $\tau = 45$ and $H_1$ the fact that the athlete is doped $\tau > 45$. So the two hypotheses we want to test are

- $H_0^{SR}\ :\ \tau_{SR} = 45$ against $H_1^{SR}\ :\ \tau_{SR} > 45$,
- $H_0^{JC}\ :\ \tau_{JC} = 45$ against $H_1^{JC}\ :\ \tau_{JC} > 45$.

3. The hypotheses are written thanks to parameters of the model, but as always in statistics, these parameters are not known. We then use an estimator to approximate these parameters. The **third step** consists in choosing an **estimator** (a test statistic) for the parameter which is tested and precise its distribution under $H_0$ (and $H_1$ if you can). Propose an estimator for the parameter you want to test for J.C. and S.R..

- The MLE of $\tau_{JC}$ is $X_{JC}$. Under $H_0^{JC}$, $X_{JC} \sim \mathcal{N}(45, 4)$, under $H_1^{JC}$, $X_{JC} \sim \mathcal{N}(\tau_{JC}, 4)$ for some $\tau_{JC} > 45$.
- and the MLE of $\tau_{SR}$ is $X_{SR}$. Under $H_0^{SR}$, $X_{SR} \sim \mathcal{N}(45, 4)$, under $H_1^{SR}$, $X_{SR} \sim \mathcal{N}(\tau_{SR}, 4)$ for some $\tau_{SR} > 45$.

4. The **fourth step** is to choose the shape of the **rejection region**. Looking at the hypotheses, you can choose in which qualitative case an estimator can be considered as too extreme and tilts the balance to the side of the culpability of the defendant. Propose a shape for the rejection region.

We want to reject the innocence of an athlete if the hematocrit level is too high so that the rejection region should be right-sided:

- $[\tau_c^{JC}, +\infty)$ for J.C.,
- $[\tau_c^{SR}, +\infty)$ for S.R..

5. The **fifth step** is to compute the **boundaries** of the rejection region using the chosen **significance level**. In the following, we want to choose $\tau_c$ such that we reject $H_0$ if the measured rate is larger than this threshold $\tau_c$.

a. You propose to reject $H_0$ as soon as the measured rate is larger than 45 ($\tau_c = 45$). What is the probability of wrongly concluding that an athlete is doped?

Wrongly concluding that an athlete is doped means that actually the athlete is not doped, i.e. $\tau = 45$ but we have rejected $H_0$, i.e. the observation $X$ is larger than 45. We know that when $\tau = 45$, $X \sim \mathcal{N}(45, 4)$, so that the probability of wrongly concluding that an athlete is doped is

$$\mathbb{P}_{H_0}(X > 45) = \mathbb{P}_{X\sim\mathcal{N}(45,4)}(X > 45) = \frac{1}{2}.$$

This probability is huge!

b. The previous probability of accusing an innocent person is far too huge: no jury would convict a defendant knowing it had such a chance of sending an innocent person to jail! So you decide to change the rule and propose to reject $H_0$ as soon as the measured rate is larger than 60 ($\tau_c = 60$). What is the probability of wrongly concluding that an athlete is doped?

Wrongly concluding that an athlete is doped means that actually the athlete is not doped, i.e. $\tau = 45$ but we have rejected $H_0$, i.e. the observation $X$ is larger than 60. We know that when $\tau = 45$, $X \sim \mathcal{N}(45, 4)$, so that the probability of wrongly concluding that an athlete is doped is

$$\mathbb{P}_{H_0}(X > 60) = \mathbb{P}_{X\sim\mathcal{N}(45,4)}(X > 60) \simeq 3.10^{-14}.$$

```
1-pnorm(60,mean=45,sd=2)
```

```
## [1] 3.18634e-14
```

This probability is very small!

c. Setting $\tau_c$ that large err in the other direction by giving the null hypothesis too much benefit of the doubt. We then decide to choose $\tau_c$ such that the probability of wrongly concluding that an athlete is doped, i.e. the probability of the type II risk, equals 0.05. Find such threshold $\tau_c$. 0.05, usually denoted $\alpha$ is called the level of significance of the test, it the probability to reject $H_0$ while it's true.

Wrongly concluding that an athlete is doped means that actually the athlete is not doped, i.e. $\tau = 45$ but we have rejected $H_0$, i.e. the observation $X$ is larger than $\tau_c$. We know that when $\tau = 45$, $X \sim \mathcal{N}(45, 4)$, so that the probability of wrongly concluding that an athlete is doped is

$$\mathbb{P}_{H_0}(X > \tau_c) = \mathbb{P}_{X\sim\mathcal{N}(45,4)}\left(\frac{X-45}{2} > \frac{\tau_c - 45}{2}\right) = 1 - F_{\mathcal{N}(0,1)}\left(\frac{\tau_c - 45}{2}\right).$$

We want the previous probability to equal $\alpha = 0.05$ so we choose $\tau_c$ such that $\frac{\tau_c - 45}{2} = q_{1-\alpha}$, where $q_{1-\alpha}$ is the $1-\alpha$ quantile of the standard distribution. Then, we choose $\tau_c = 45 + 2q_{1-\alpha} \simeq 48.29$.

```
45+ 2* qnorm(0.95)
```

```
## [1] 48.28971
```

6. The **sixth step** is to write the test, i.e. summarizing the previous steps by precising the **rule of rejection for** $H_0$ and then **conclude considering the observed values**.

a. Summarize the previous conclusions into a decision rule.

We 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 $H_0$ it otherwise.

Then the hypostheses test is

$$\phi(X) = \begin{cases} 0 & \text{if } X > \tau_c = 45 + 2q_{1-\alpha} \\ 1 & \text{otherwise} \end{cases}.$$

b. What are your conclusions for J.C. and S.R.

The observed value of J.C. is $48 < 45 + 2q_{1-\alpha}$ so that we don't reject $H_0^{JC}$. The observed value of S.R. is $50 > 45 + 2q_{1-\alpha}$ so that we reject $H_0^{SR}$.

c. When you observe a measure leading to the rejection of $H_0$, it doesn't prove that the athlete is doped. In other words, it must be remembered that rejecting $H_0$ does not prove that $H_0$ is false, any more than a jury's decision to convict guarantees that the defendant is guilty. It just means that if the true rate is 45, measurements of this rate as large or larger than $\tau_c$ are expected to occur only 5% of the time. Because of that small probability, a reasonable conclusion when the measurement of this rate is as large or larger than $\tau_c$ is that the true rate is larger than 45. Check that with your rule you wrongly reject $H_0$ with an average frequency of 0.05.

```
n <- 10000
x <- rnorm(n,45,2)
hyp <- (x>45+ 2* qnorm(0.95))
1/n*sum(hyp)
```

```
## [1] 0.0506
```

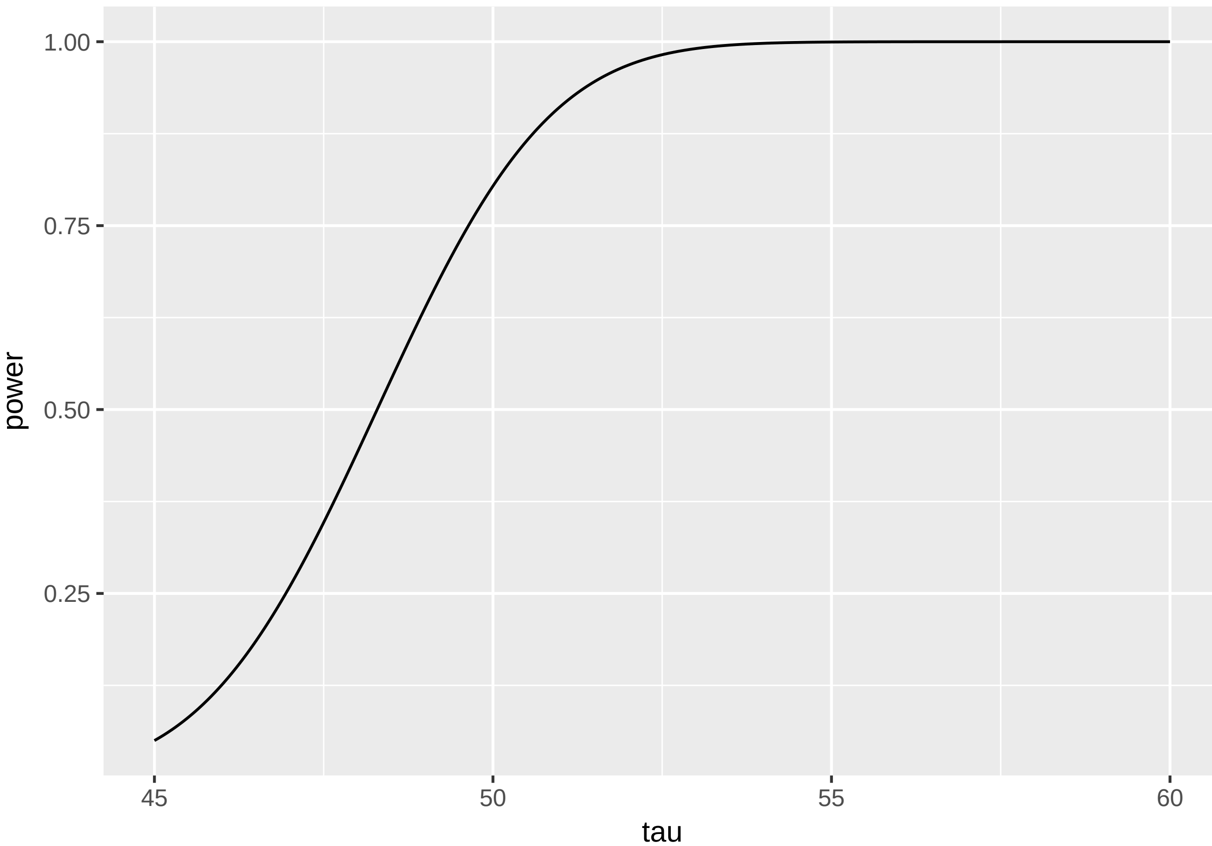
7. As a seventh step, you could check the power of the test at the end. You could also compute the p-value (PC8) for the observed value.

a. Plot the power function associated to your hypotheses test.

The power function is defined as follows

$$\begin{aligned} \tau \mapsto \beta_\phi(\tau) &= \mathbb{P}_{H_0}(\text{reject } H_0) = \mathbb{P}_{X\sim\mathcal{N}(\tau,4)}(X > 45 + 2q_{1-\alpha}) \\ &= \mathbb{P}_{X\sim\mathcal{N}(\tau,4)}\left(\frac{X-\tau}{2} > \frac{45 + 2q_{1-\alpha} - \tau}{2}\right) \\ &= 1 - F_{\mathcal{N}(0,1)}\left(\frac{45 + 2q_{1-\alpha} - \tau}{2}\right) \end{aligned}$$

```
tau <-seq(45, 60,by=0.01)
power <- data.frame(tau=tau, power = (1 - pnorm((45+ 2* qnorm(0.95)-tau)/2)))
ggplot(data=power, aes(x=tau, y=power))+geom_line()
```




b. What is the probability of detecting an abnormal hematocrit level when the true level is 50?

This probability is

$$\mathbb{P}_{X\sim\mathcal{N}(50,4)}(\text{reject } H_0) = \beta_\phi(50) = 1 - F_{\mathcal{N}(0,1)}\left(\frac{45 + 2q_{1-\alpha} - 50}{2}\right) \simeq 0.80$$

```
tau <- 50
1-pnorm((45+ 2* qnorm(0.95)-tau)/2)
```

```
## [1] 0.8037649
```

Note that the choices for steps 3 and 4 are optimal (best power) using Neyman-Pearson theory in the case of simple hyotheses. We will discuss it in PC8.

**Exercise 2: Hypothesis testing and biased coin**

You are arguing with a friend about the movie you are about to watch. This friend decides to make this choice by tossing a coin. Before doing so, you want to check that the coin is biased or not.

1. You propose to flip the coin 10 times and keep the coin if the number of heads is 5. Write the statistical model and the hypotheses you are testing. What is the rejection region and the acceptance region? Compute the significance level of this test and plot its power function.

Let $x$ be the observed number of heads obtained after 10 flips. We assume $x$ is the realization of a random variable $X$ which is distributed from a Binomial distribution $Bin(10, p)$ where $p$ is unknown. The statistical model is

$$\left(\left(\{0, 1, \ldots, 10\}, \mathcal{P}\left(\{0, 1, \ldots, 10\}\right)\right), \{Bin(10, p), p \in [0, 1]\}\right).$$

We want to test $H_0 \ : \ p = 0.5$ against $H_1 \ : \ p \neq 0.5$. The acceptance region of the considered test is $\{5\}$ and the rejection region is $\{0, 1, 2, 3, 4, 6, 7, 8, 9, 10\}$. The test by itself is

$$\phi_1(X) = \begin{cases} 0 & \text{if } X = 5 \\ 1 & \text{otherwise} \end{cases}.$$

The significance level of this test $\phi_1$ is the probability to reject $H_0$ (i.e. $X \neq 5$) when $H_0$ is true i.e. when $p = 0.5$:

$$\alpha_{\phi_1} = \mathbb{P}_{X\sim Bin(10,0.5)}(X \neq 5) \simeq 0.75$$

```
1-dbinom(x=5,size=10,prob=1/2)
```

```
## [1] 0.7539062
```

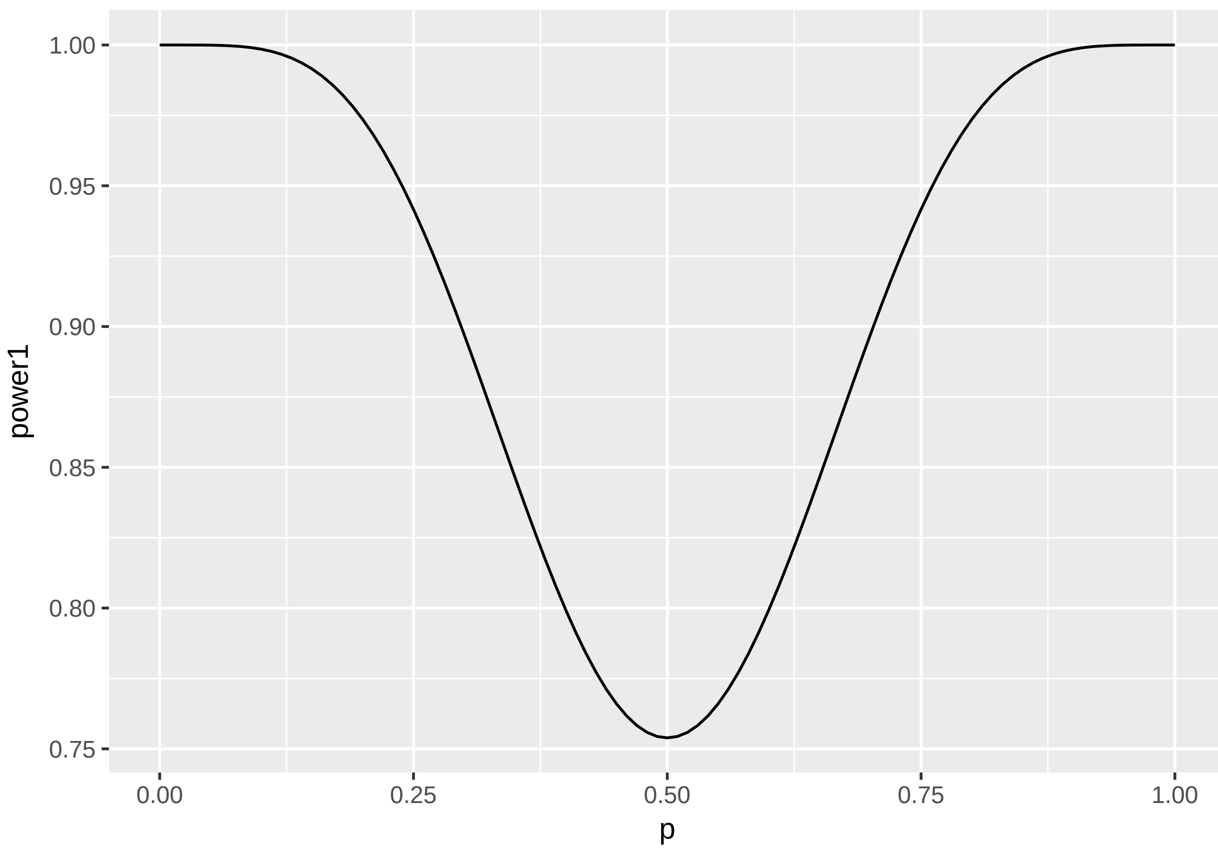
The power function of $\phi_1$ is

$$p \mapsto \beta_{\phi_1}(p) = \mathbb{P}_{X\sim Bin(10,p)}(\text{reject } H_0) = \mathbb{P}_{X\sim Bin(10,p)}(X \neq 5).$$

Note that the significance level equals to the power function at the parameter associated to $H_0$ (if $H_0 : \theta \in \Theta_0$ is simple, i.e. $\text{Card}(\Theta_0) = 1$) $\alpha_{\phi_1} = \beta_{\phi_1}(1/2)$.

```
p <- seq(0,1,by=0.01)
power1 <- sapply(p, function(pr) 1-dbinom(x=5,size=10,prob=pr))
data <- data.frame(p=p, power1=power1)

library(ggplot2)
pl <- ggplot(data=data, aes (x=p, y=power1)) + geom_line()
pl
```



2. You repeated the previous strategy twice on two of your four coins and you rejected the two first coins. Since the rejection rate under $H_0$ (0.75) is too high, you decide to change your strategy. You propose another rule for the two last coins: if the number of heads among 10 tosses is smaller or equal to 1 or larger or equal to 9 then you use another coin. What is the rejection region and the acceptance region? Compute the significance level of this test. Plot the power function. What is the probability that you accept the coin if it is actually biased with a probability of 3/4 (and 1/4) to obtain a head?

The acceptance region of the considered test is $\{2,3,4,5,6,7,8\}$ and the rejection region is $\{0,1,9,10\}$. The test by itself is

$$\phi_2(X) = \begin{cases} 0 & \text{if } X \in \{2,3,4,5,6,7,8\} \\ 1 & \text{otherwise} \end{cases}.$$

The significance level of this test $\phi_2$ is the probability to reject $H_0$ (i.e. $X \in \{0,1,9,10\}$) when $H_0$ is true i.e. when $p = 0.5$:

$$\alpha_{\phi_2} = \mathbb{P}_{X\sim Bin(10,0.5)}\left(X \in \{0,1,9,10\}\right) \simeq 0.02$$

```r
1-pbinom(q=8,size=10,prob=1/2)+pbinom(q=1,size=10,prob=1/2)
```

```
## [1] 0.02148438
```

```r
1-sum(dbinom(x=c(2,3,4,5,6,7,8),size=10,prob=1/2))
```

```
## [1] 0.02148437
```

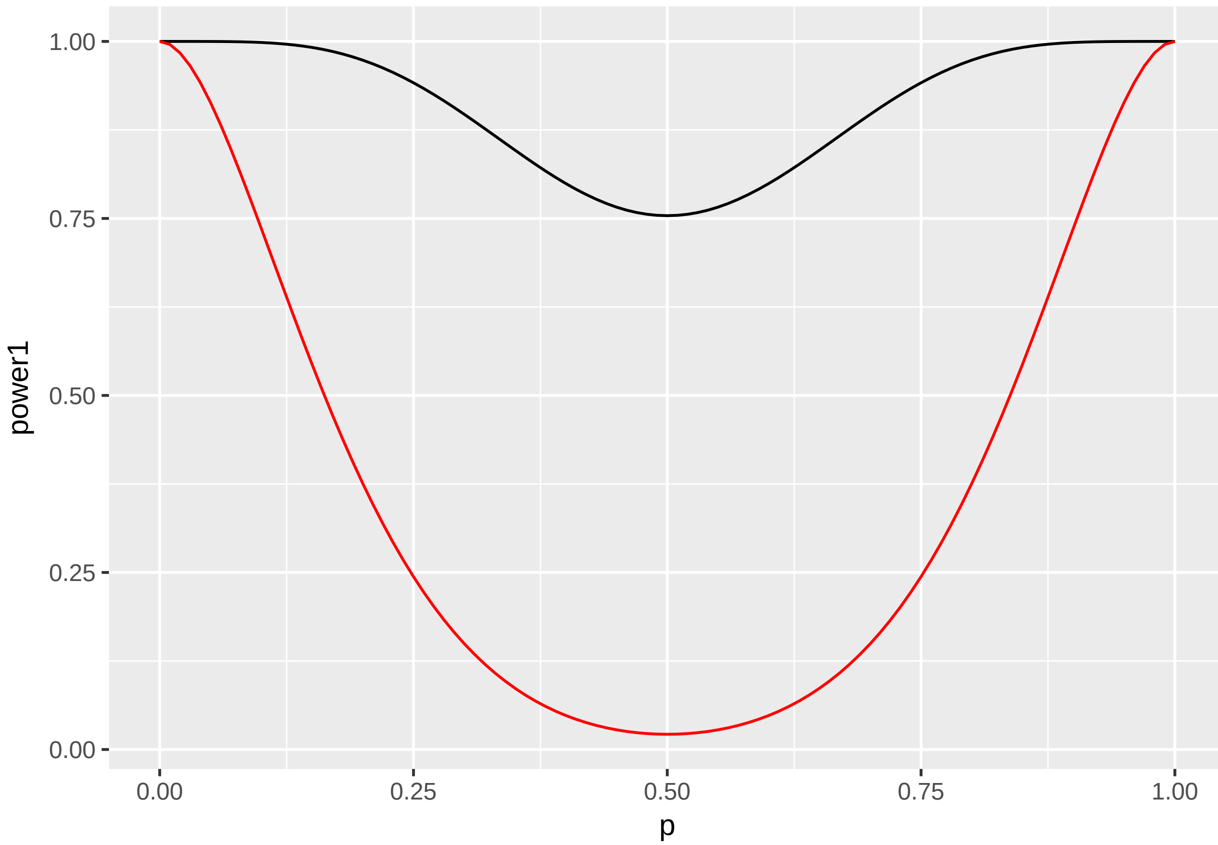
The power function of $\phi_2$ is

$$p \mapsto \beta_{\phi_2}(p) = \mathbb{P}_{X\sim Bin(10,p)}(\text{reject } H_0) = \mathbb{P}_{X\sim Bin(10,p)}(X \in \{0,1,9,10\}).$$

Again $\alpha_{\phi_2} = \beta_{\phi_2}(1/2)$.

```r
power2 <- sapply(p, function(pr) sum(dbinom(x=c(0,1,9,10),size=10,prob=pr)))
data$power2 <- power2

pl <-pl + geom_line(aes(x=p,y=power2),color="red")
pl
```




The probability that you accept the coin if it is actually biased with $p = 3/4$ is

$$\mathbb{P}_{X\sim Bin(10,3/4)}(\text{accept } H_0) = \mathbb{P}_{X\sim Bin(10,3/4)}(X \notin \{0,1,9,10\}) \simeq 0.76.$$

The probability that you accept the coin if it is actually biased with $p = 1/4$ is

$$\mathbb{P}_{X\sim Bin(10,1/4)}(\text{accept } H_0) = \mathbb{P}_{X\sim Bin(10,1/4)}(X \notin \{0,1,9,10\}) \simeq 0.76.$$

```
1-sum(dbinom(x=c(0,1,9,10),size=10,prob=3/4))
```

```
## [1] 0.7559452
```

```
1-sum(dbinom(x=c(0,1,9,10),size=10,prob=1/4))
```

```
## [1] 0.7559452
```

3. Your friend disagrees with your rule because he is superstitious and hates number 2. He would prefer to accept the coin if there are 1, 3, 4, 5, 6, 7, 8 or 9 heads. What is the rejection region and the acceptance region? Compute the significance level of this test. Plot the power function. What is the probability that you accept the coin if it is actually biased with a probability of 3/4 (and 1/4) to obtain a head?

The acceptance region of the considered test is $\{1,3,4,5,6,7,8,9\}$ and the rejection region is $\{0,2,10\}$. The test by itself is

$$\phi_3(X) = \begin{cases} 0 & \text{if } X \in \{1,3,4,5,6,7,8,9\} \\ 1 & \text{otherwise} \end{cases}.$$

The significance level of this test $\phi_3$ is the probability to reject $H_0$ (i.e. $X \in \{0,2,10\}$) when $H_0$ is true i.e. when $p = 0.5$:

$$\alpha_{\phi_3} = \mathbb{P}_{X\sim Bin(10,0.5)}\left(X \in \{0,2,10\}\right) \simeq 0.05$$

```r
sum(dbinom(x=c(0,2,10),size=10,prob=1/2))
```

```
## [1] 0.04589844
```

```r
1-sum(dbinom(x=c(1,3,4,5,6,7,8,9),size=10,prob=1/2))
```

```
## [1] 0.04589844
```

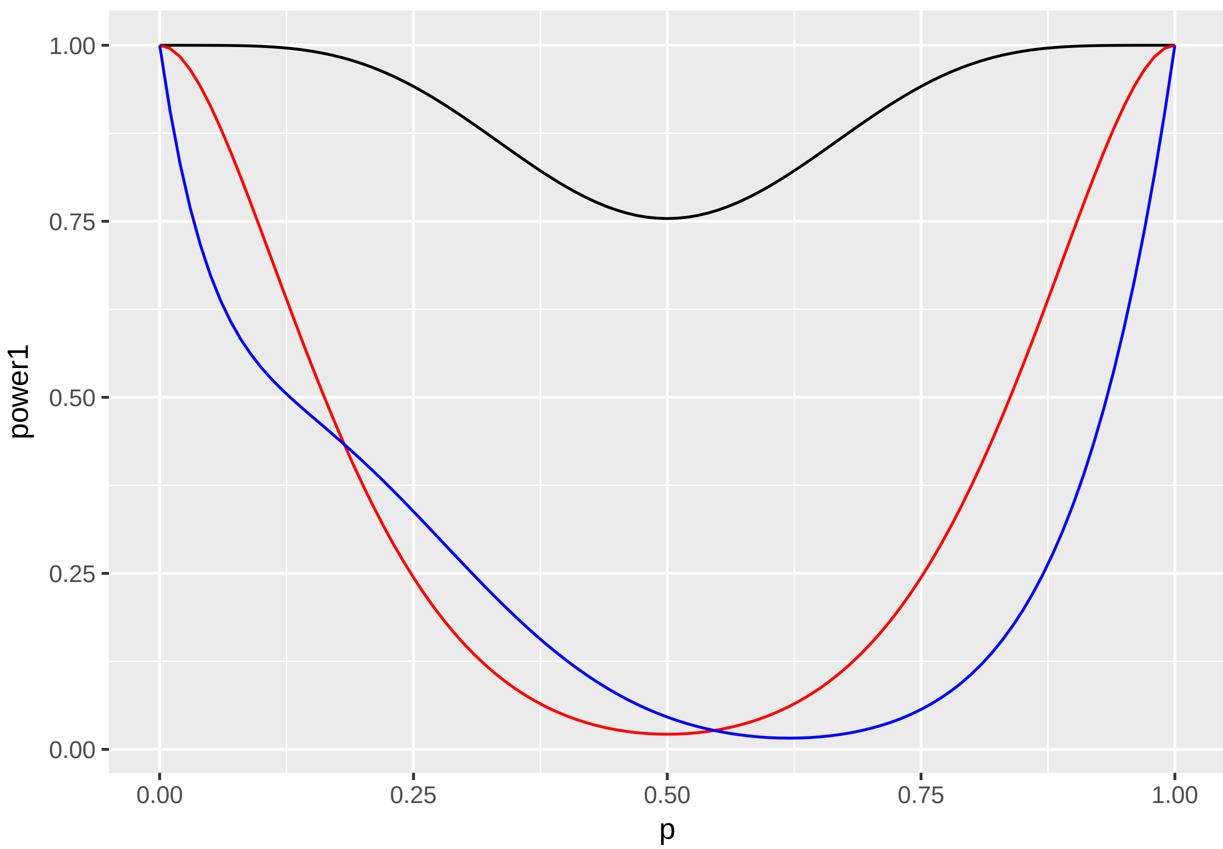
The power function of $\phi_3$ is

$$p \mapsto \beta_{\phi_3}(p) = \mathbb{P}_{X\sim Bin(10,p)}(\text{reject } H_0) = \mathbb{P}_{X\sim Bin(10,p)}(X \in \{0,2,10\}).$$

Again $\alpha_{\phi_3} = \beta_{\phi_3}(1/2)$.

```r
power3 <- sapply(p, function(pr) sum(dbinom(x=c(0,2,10),size=10,prob=pr)))
data$power3 <- power3

pl <- pl + geom_line(aes(x=p,y=power3),color="blue")
pl
```




The probability that you accept the coin if it is actually biased with $p = 3/4$ is

$$\mathbb{P}_{X\sim Bin(10,3/4)}(\text{accept } H_0) = \mathbb{P}_{X\sim Bin(10,3/4)}(X \notin \{0,2,10\}) \simeq 0.94.$$

The probability that you accept the coin if it is actually biased with $p = 1/4$ is

$$\mathbb{P}_{X\sim Bin(10,1/4)}(\text{accept } H_0) = \mathbb{P}_{X\sim Bin(10,1/4)}(X \notin \{0,2,10\}) \simeq 0.66.$$

```r
1-sum(dbinom(x=c(0,2,10),size=10,prob=3/4))
```

```
## [1] 0.9432993
```

```r
1-sum(dbinom(x=c(0,2,10),size=10,prob=1/4))
```

```
## [1] 0.662118
```

4. You finally propose to throw 30 times the coin and reject the coin if the number of heads $X$ is such that $|X-15|>\delta$ for some $\delta>0$. Propose a hypothesis testing at level smaller than and as close as possible to 0.05. Plot its power. What is the probability that you accept the coin if it is actually biased and with a probability of 3/4 (and 1/4) to obtain a head? What is your conclusion if you obtain 4 heads?

Step1: the statistical model is now

$$((\{0,1,\ldots,30\},\mathcal{P}(\{0,1,\ldots,30\})),\{Bin(30,p),p\in[0,1]\}).$$

Step 2: we want to test the same hypotheses: $H_0\ :\ p=0.5$ against $H_1\ :\ p\neq 0.5$.

Step 3: the MLE of $p$ is $X/30$.

Step 4: we want to reject $H_0$ when the estimator of $p$ is too far from 1/2, i.e. $|X/30-1/2|>\epsilon$, i.e. $|X-15|>\delta$ for some $\delta$ ($\epsilon=\delta/30$).

Step 5: The significance level of the test which rejects $H_0$ when $|X-15|>\delta$ is

$$\alpha_4=\mathbb{P}_{X\sim Bin(10,0.5)}(|X-15|>\delta)=\mathbb{P}_{X\sim Bin(10,0.5)}(X>15+\delta)+\mathbb{P}_{X\sim Bin(10,0.5)}(X<15-\delta).$$

Here are its values for $\delta\in\{0,1,\ldots,14\}$:

```r
n <- 30
s=0:(n/2-1)
r <- sapply(s, function(x) (1-pbinom(q=15+x,size=n,prob=1/2)+pbinom(q=15-x-1,size=n,prob=1/2)))
r
```

```
##  [1] 8.555356e-01 5.846647e-01 3.615946e-01 2.004884e-01 9.873715e-02
##  [6] 4.277395e-02 1.612480e-02 5.222879e-03 1.430906e-03 3.249142e-04
## [11] 5.947612e-05 8.430332e-06 8.679926e-07 5.774200e-08 1.862645e-09
```

The best $\delta$ is then $\delta=5$ so that the aceptance region is $\{10,11,12,13,14,15,16,17,18,19,20\}$, the rejection region is $\{0,1,2,3,4,5,6,7,8,9,21,22,23,24,25,26,27,28,29,30\}$ and the significance level is $\alpha_{\phi_4}=0.042$. The test by itself is

$$\phi_4(X)=\begin{cases}0 & \text{if } X\in\{10,11,12,13,14,15,16,17,18,19,20\}\\ 1 & \text{otherwise}\end{cases}.$$

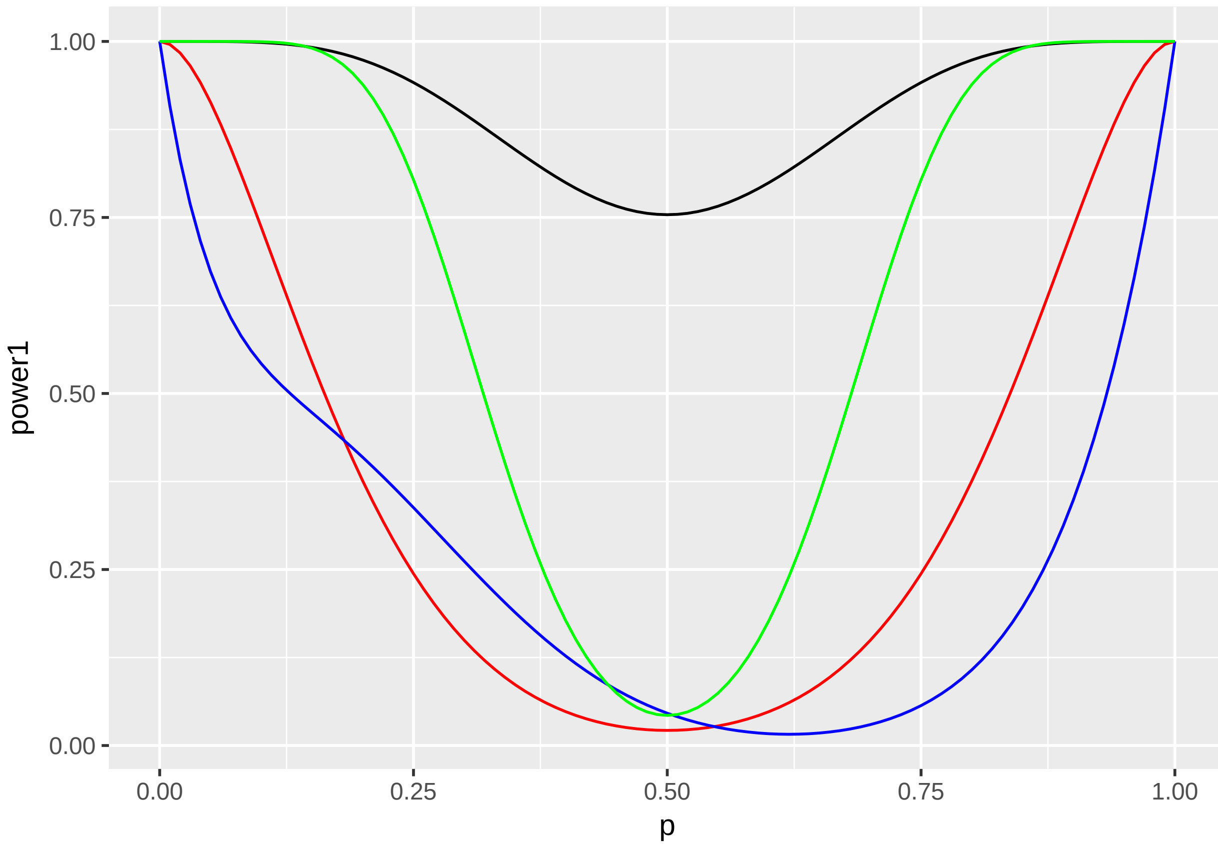
The power function of $\phi_4$ is

$$p\mapsto\beta_{\phi_4}(p)=\mathbb{P}_{X\sim Bin(10,p)}(\text{reject } H_0)=1-\mathbb{P}_{X\sim Bin(10,p)}(X\in\{10,11,12,13,14,15,16,17,18,19,20\}).$$

Again $\alpha_{\phi_4}=\beta_{\phi_4}(1/2)$.

```r
power4 <- sapply(p, function(pr) 1-sum(dbinom(x=c(10,11,12,13,14,15,16,17,18,19,20),size=30,prob=pr)))
data$power4 <- power4

pl <- pl + geom_line(aes(x=p,y=power4),color="green")
pl
```




The probability that you accept the coin if it is actually biased with $p = 3/4$ is

$$\mathbb{P}_{X\sim Bin(10,3/4)}(\text{accept } H_0) = \mathbb{P}_{X\sim Bin(10,3/4)}(X \in \{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}) \simeq 0.20.$$

The probability that you accept the coin if it is actually biased with $p = 1/4$ is

$$\mathbb{P}_{X\sim Bin(10,1/4)}(\text{accept } H_0) = \mathbb{P}_{X\sim Bin(10,1/4)}(X \in \{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}) \simeq 0.20.$$

```
sum(dbinom(x=10:20,size=30,prob=3/4))
```

```
## [1] 0.1965931
```

```
sum(dbinom(x=10:20,size=30,prob=1/4))
```

```
## [1] 0.1965931
```

If among 30 flips we obtain 4 heads, i.e. the observed value is $x = 4$, then we reject $H_0$.

5. Comparing the power function of the four previous decision rules, which test would you choose and why?

$\phi_1$ has a significance level which is too high meaning, we do not give enough benefit to the doubt. We can see in the plot of its power function that the power, at $p = 1/2$, is far from being below the usual significance level $\alpha = 0.05$.

The three other tests have a significance level which is smaller than 0.05. The best test is the one which has the highest power for other values than $p_0 = 1/2$, which means that in average it does reject more $H_0$ when $H_0$ is false, i.e. the probability of a type II error is smaller. $\phi_4$ is then the test we prefer.

## Choice of the hypotheses

### Exercise 3: Test for an organic certification

To obtain an "organic" certification, a manufacturer of organic products has to guarantee that each batch contains less than 1% of GMO. Then he collects $n = 25$ products by batch and test if they contain less than 1% of GMO. Let $X_i$ be the logarithm of the GMO level (in %) of the product $i$.

**Model** : We assume that the $X_i$ are independent and are distributed from a Gaussian distribution $\mathcal{N}(\theta, 1)$.

1. As to the manufacturer, the GMO level is less than 1% unless the contrary is proved. He wants to test the hypothesis $H_0 \ : \ \theta \leq 0$ against $H_1 \ : \ \theta > 0$. The two hypotheses are composite and not simple. He wants that the test fails with probability less than 5% when $\theta \leq 0$ Compute a threshold $a_{25,5}$ such that

$$\sup_{\theta \leq 0} \mathbb{P}_\theta(\bar{X}_{25} > a_{25,5}) = 0.05,$$

where $\bar{X}_{25} = \frac{1}{25}\sum_{i=1}^{25} X_i$. You can use that $\mathbb{P}(Z > 1.645) \simeq O.O5$, when $Z \sim \mathcal{N}(0,1)$.

We know that, under $\mathbb{P}_\theta = \mathcal{N}(\theta, 1)$ the distribution of $\bar{X}_{25}$ is $\mathcal{N}(\theta, 1/25)$ so that $5\left(\bar{X}_{25} - \theta\right) \sim \mathcal{N}(0,1)$.

Then

$$\mathbb{P}_\theta(\bar{X}_{25} > a_{25,5}) = \mathbb{P}_\theta\left(5\left(\bar{X}_{25} - \theta\right) > 5(a_{25,5} - \theta)\right) = \mathbb{P}_{Z\sim\mathcal{N}(0,1)}\left(Z > 5(a_{25,5} - \theta)\right) = 1 - F_{\mathcal{N}(0,1)}\left(5(a_{25,5} - \theta)\right).$$

Since $\theta \mapsto 1 - F_{\mathcal{N}(0,1)}\left(5(a_{25,5} - \theta)\right)$ is increasing with $\theta$, then

$$\sup_{\theta \leq 0} \mathbb{P}_\theta(\bar{X}_{25} > a_{25,5}) = 1 - F_{\mathcal{N}(0,1)}\left(5(a_{25,5} - 0)\right) = 1 - F_{\mathcal{N}(0,1)}(5a_{25,5}).$$

So we choose $a_{25,5}$ such that $5a_{25,5} = q_{1-\alpha}$, where $q_{1-\alpha}$ is the $1-\alpha$-quantile of a standard distribution. With $\alpha = 0.05$, we obtain $a_{25,5} = q_{0.95}/5 \simeq 0.33$.

```r
qnorm(0.95)/5
```

```
## [1] 0.3289707
```

2. An anti-GMO organization wants to be sure that the GMO level in organic products are indeed less than 1%. In particular, it wants to know if the test succeeds in eliminating products which contains more than 50% the legal maximum. What is the probability that the test does not reject $H_0$ when the GMO level is 1.5%?

We want to know the probability that the test does not reject $H_0$ when the GMO level is 1.5%, i.e. $\theta = \ln(1.5)$.

$$\begin{aligned}\mathbb{P}_\theta(\bar{X}_{25} \leqslant a_{25,5}) &= \mathbb{P}_\theta\left(5\left(\bar{X}_{25} - \theta\right) \leqslant 5(a_{25,5} - \theta)\right) = \mathbb{P}_{Z\sim\mathcal{N}(0,1)}\left(Z \leqslant 5(a_{25,5} - \theta)\right) \\ &= F_{\mathcal{N}(0,1)}\left(5(a_{25,5} - \theta)\right) = F_{\mathcal{N}(0,1)}\left(q_{0.95} - 5\ln(1.5)\right) \simeq 0.35.\end{aligned}$$

```r
pnorm( qnorm(0.95) -5*log(3/2) )
```

```
## [1] 0.3510557
```

The probability of not rejecting $H_0$ when $\theta = \ln(3/2)$ is 35%.

3. Shocked by the previous result, the organization advocates for a test which really proves that the GMO level is less than 1%. The organization supporters think that the GMO level is greater than 1% unless the contrary is proved. Then $H_0$ is $\theta > 0$ and $H_1$ is $\theta \leq 0$. Propose a hypothesis test $H_0$ against $H_1$ such that the probability that the test wrongly rejects $H_0$ is less than 0.05.

We are going to choose a left hand sided rejection region i.e. $(-\infty, -\delta)$. We choose $\delta$ such that

$$\sup_{\theta > 0} \mathbb{P}_\theta(\bar{X}_{25} < -\delta) = 0.05.$$

Using question 1 (by symmetry), $\delta = a_{25,5}$.

4. Regulatory agencies accept the hypothesis test proposed by the manufacturer ($H_0 : \theta \leq 0$ against $H_1 : \theta > 0$) but require to detect at least 80% of the cases where the GMO level exceeds 10% of the legal maximum. The significance level is still 5%, what should the manufacturer do to adhere to the law?

He has to increase the size $n$ of the sample to increase the power of the test. We search for $s$ and $n$ such that

$$\sup_{\theta \leqslant 0} \mathbb{P}_\theta(\bar{X}_n > s) = 0.05$$

and

$$\inf_{\theta \geqslant \ln(1+1/10)} \mathbb{P}_\theta(\bar{X}_n > s) \geqslant 0.8.$$

We first search for $s$ such that the first equality is true

$$5\% = \sup_{\theta \leqslant 0} \mathbb{P}_\theta(\bar{X}_n > s) = \sup_{\theta \leqslant 0} \mathbb{P}_{Z \sim \mathcal{N}(0,1)} \left(Z > \sqrt{n}(s - \theta)\right) = 1 - F_{\mathcal{N}(0,1)} \left(\sqrt{n} s\right).$$

So we choose $s$ such that $\sqrt{n}s = q_{1-\alpha}$, i.e. $s = q_{1-\alpha}/\sqrt{n}$. And

$$\inf_{\theta \geqslant \ln(1+1/10)} \mathbb{P}_\theta(\bar{X}_n > s) = \inf_{\theta \geqslant \ln(11/10)} \mathbb{P}_{Z \sim \mathcal{N}(0,1)} \left(Z > \sqrt{n}(s - \theta)\right) = 1 - F_{\mathcal{N}(0,1)} \left(\sqrt{n}(s - \ln(11/10))\right).$$

So that we search for $n$ such that $\sqrt{n}(s - \ln(11/10)) \leqslant q_{0.2}$, i.e. $q_{1-\alpha} - \sqrt{n}\ln(11/10) \leqslant q_{0.2}$, i.e. $n \geqslant \left(\frac{q_{1-\alpha} - q_{0.2}}{\ln(11/10)}\right)^2 \simeq 680.1$.

```
qnorm(0.95)
```

```
## [1] 1.644854
```

```
qnorm(0.2)
```

```
## [1] -0.8416212
```

```
(qnorm(0.95)-qnorm(0.2))/log(11/10)
```

```
## [1] 26.08824
```

```
((qnorm(0.95)-qnorm(0.2))/log(11/10))^2
```

```
## [1] 680.5963
```

**Exercise 4: Various experiments**

Formalize the hypotheses and statistical models in the following cases:

1. Defeated in his most recent attempt to win a congressional seat because of a large gender gap, a politician has spent the last two years speaking out in favor of women's rights issues. A newly released poll claims to have contacted a random sample of 120 of the politician's current supporters and found that 72 were men. In the election that he lost, exit polls indicated that 65% of those who voted for him were men. The politician wants to know if his campaign has has aositive effect on women.

Let $x_i = 1$ if the i-th supporter of the politician is a man and $x_i = 0$ otherwise for $i \leqslant n = 120$. From the survey, $\sum_{i=1}^{120} x_i = 72$. We assume that $(x_1, \ldots x_n)$ is a realization of $(X_1, \ldots, X_n)$, where $X_i$ are i.i.d. from the Bernoulli distribution $B(\theta)$ where $\theta$ represents the proportion of men among all the supporters of the politician. So that the statistical model is

$$\left(\{0,1\}^n, \mathcal{P}\left(\{0,1\}^n\right), \left\{B(\theta)^{\otimes n}, \theta \in [0,1]\right\}\right).$$

The status quo is that there is no change so that he wants to test $H_0 : \theta = 0.65$ against $H_1 : \theta < 0.65$, i.e. the proportion of women has increased (equivalently, the proportion of men has increased).

2. A herbalist is experimenting juices extracted from berries and roots that may have the ability to affect the Stanford-Binet IQ scores of students afflicted with mild cases of attention deficit disorder (ADD). A random sample of twenty-two children diagnosed with the condition have been drinking Brain-Blaster daily for two months. Past experience suggests that children with ADD score an average of 95 on the IQ test with a standard deviation of 15.

The studied population is all children with ADD. Let $x_i$ be the IQ score of the i-th children $i \leqslant n = 22$. We assume that $(x_1, \dots x_n)$ is a realization of $(X_1, \dots, X_n)$, where $X_i$ are i.i.d. from some distribution $P_{\theta,\sigma}$ where $\theta$ represents the average score among children with ADD and $\sigma$ their standard deviation. So that the statistical model is

$$\left(\mathbb{R}^n, \mathcal{B}\left(\mathbb{R}^n\right), \left\{P_\theta^{\otimes n}, \theta \in \mathbb{R}, \sigma \in \mathbb{R}_+\right\}\right).$$

The herbalist has to prove that his juice is improving the IQ score of children. So that he wants to test $H_0 \ : \ \theta = 95$ against $H_1 \ : \ \theta > 95$. If $H_0$ is not rejected, he could also test if $\sigma$ has changed: he could also test $H_0 \ : \ \sigma = 15$ against $H_1 \ : \ \sigma \neq 15$.

3. A company sold a sampler producing uniform random digits. Its client suspects the sampler to be biased towards small digits and wants to complain. The client pays an external control society to test this sampler. They obtain the following sample $1, 2, 1, 2, 0, 8, 6, 1, 2, 4, 5, 1, 2, 4, 8, 4, 4, 3, 0, 2$.

Let $x_i$ be the i-th sampled digit for $i \leq n = 20$. We assume that $(x_1, \dots x_n)$ is a realization of $(X_1, \dots, X_n)$, where $X_i$ are i.i.d. from the distribution $P_p$ where $p \in \Delta_{10} := \{p = (p_1, \dots, p_{10}) \in [0,1]^{10} : \ \sum_{i=1}^{10} p_i = 1\}$ and $P_p$ is defined as $P_p(X = j) = p_j$ for all $j \leq 10$. So that the statistical model is

$$\left(\{0, 1, 2, \dots, 9\}^n, \mathcal{P}\left(\{0, 1, 2, \dots, 9\}^n\right), \left\{P_p^{\otimes n}, p \in \Delta_{10}\right\}\right).$$

The client has to prove that the company does not provide a uniform sampler. The probability to reject the fact that the sampler is uniform if the sampler is really uniform should be small. So that he wants to test $H_0 \ : \ p = (1/10, \dots, 1/10)$ against $H_1 \ : \ p \neq (1/10, \dots, 1/10)$.

## Gaussian models

### Exercise 5: Gaussian hypothesis testing

Let

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

be Gaussian observations.

1. We first assume that $\sigma$ is known and we want to test $H_0$: $\mu = \mu^*$ versus $H_1$: $\mu \neq \mu^*$ for $\mu^* in \mathbb{R}$.

a. Propose a test at level 0.05 (steps 1 to 6 of Exercise 1) and create an R function **rej_reg_1** that takes as input a vector of observations $x = (x_1, \dots, x_n)$, the variance $\sigma^2$, the parameter $\mu^*$ and a level of signficance $\alpha \in (0, 1)$, and outputs a vector of size four containing the lower and upper bound of the aceptance region, the statistic of the test and the result of the test: 0 if $H_0$ is not rejected and 1 if $H_0$ is rejected.

Step 1: The statistical model is

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

Step 2: The two hypotheses we want to test are $H_0 \ : \ \mu = \mu^*$ against $H_1 \ : \ \mu \neq \mu_*$. This is the so called one sample, two-sided U-test.

Step 3: The MLE of $\mu$ is $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Under $H_0$: $\bar{X}_n \sim \mathcal{N}(\mu^*, \sigma^2/n)$. Under $H_1$: $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$ for some $\mu \neq \mu^*$.

Step 4: We want to reject the null hypothesis when the estimator of the mean is too far from $\mu^*$. So that we choose an aceptance region of the following shape: $[\mu^* - \delta, \mu^* + \delta]$.

Step 5: We choose $\delta$ to obtain a test with significance level $\alpha$, i.e. such that

$$\alpha = \mathbb{P}_{H_0}(|\bar{X}_n - \mu^*| > \delta) = 2\mathbb{P}_{\bar{X}_n \sim \mathcal{N}(\mu^*, \sigma^2/n)}\left(\sqrt{n}\frac{\bar{X}_n - \mu^*}{\sigma} > \frac{\sqrt{n}\delta}{\sigma}\right) = 2 - 2F_{\mathcal{N}(0,1)}\left(\frac{\sqrt{n}\delta}{\sigma}\right).$$

So that, we choose $\delta$ such that $\frac{\sqrt{n}\delta}{\sigma} = q_{1-\alpha/2}$ so $\delta = \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}$.

Step 6: Finally our test rejects $H_0$ when $|\bar{X}_n - \mu^*| > \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}$. Formally,

$$\phi(X_1, \ldots, X_n) = \begin{cases} 0 & \text{if } \bar{X}_n \in \left[\mu^* - \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}, \mu^* + \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}\right] \\ 1 & \text{otherwise} \end{cases}.$$

```r
rej_rec_1 <- function(x, sigma2, alpha, mu0){
  q_alpha <- qnorm(1-alpha/2)
  n <- length(x)
  mean <- mean(x)
  phi <- ((mean < mu0-sqrt(sigma2/n)*q_alpha)||(mean >mu0+sqrt(sigma2/n)*q_alpha))
  test <- c(mu0-sqrt(sigma2/n)*q_alpha, mu0+sqrt(sigma2/n)*q_alpha,mean, phi)
  return(test)
}
```

```r
mu0 <- 0
sigma2 <- 4
alpha <- 0.05
n<-10
x <- rnorm(10,mu0,sqrt(sigma2))
rej_rec_1(x, sigma2, alpha, mu0)
```

```
## [1] -1.2395901  1.2395901 -0.1306374  0.0000000
```

b. Check the significance level of the test through simulations when $\mu^* = 0$, $\sigma = 2$, $\alpha = 0.05$ and $n = 10$.

```r
mu0 <- 0
sigma2 <- 4
alpha <- 0.05
n<- 10

samples <- lapply(1:1e3, function(i) rnorm(n, mu0, sqrt(sigma2)))
tests <- lapply(samples, function(x) rej_rec_1(x, sigma2, alpha, mu0))
test <- sapply(tests, function(te) te[4])
sum(test)/1e3
```

```
## [1] 0.049
```

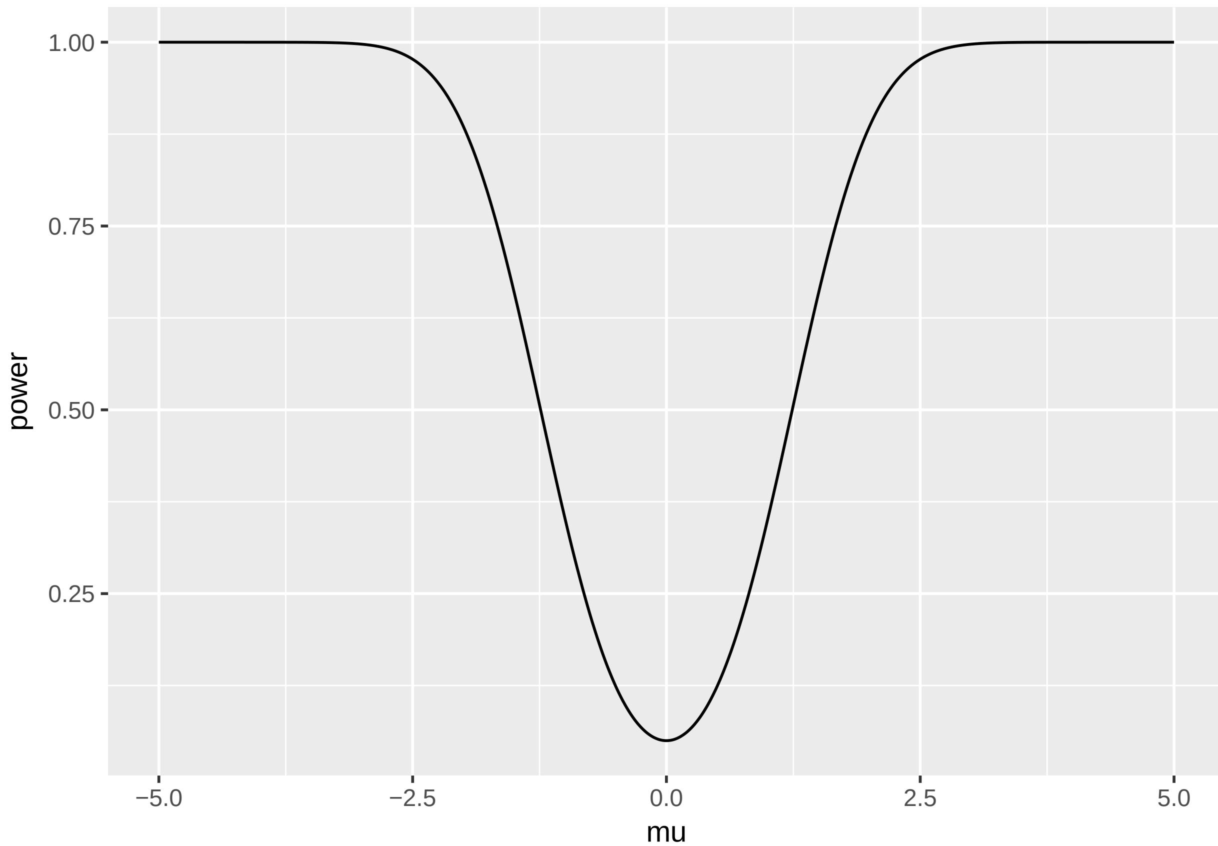
c. Plot the power function of this test under the same condition.

The power function of this test is

$$\mu \mapsto \beta_\phi(\mu) = \mathbb{P}_{X_i \sim \mathcal{N}(\mu, \sigma^2)}(\text{reject } H_0) = \mathbb{P}_{X_i \sim \mathcal{N}(\mu, \sigma^2)}\left(\bar{X}_n \notin \left[\mu^* - \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}, \mu^* + \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}\right]\right)$$

$$\begin{aligned}\beta_\phi(\mu) &= \mathbb{P}_{X_i \sim \mathcal{N}(\mu, \sigma^2)}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} > \frac{\mu^* - \mu}{\sigma/\sqrt{n}} + q_{1-\alpha/2}\right) + \mathbb{P}_{X_i \sim \mathcal{N}(\mu, \sigma^2)}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < \frac{\mu^* - \mu}{\sigma/\sqrt{n}} - q_{1-\alpha/2}\right) \\ &= 1 - F_{\mathcal{N}(0,1)}\left(\frac{\mu^* - \mu}{\sigma/\sqrt{n}} + q_{1-\alpha/2}\right) + F_{\mathcal{N}(0,1)}\left(\frac{\mu^* - \mu}{\sigma/\sqrt{n}} - q_{1-\alpha/2}\right).\end{aligned}$$

```
mu0 <- 0
sigma2 <- 4
alpha <- 0.05
n<- 10
mu <- seq(-5,5,by=0.01)
q <- qnorm(1-alpha/2)
power <-1-pnorm( sqrt(n/sigma2)*(mu0-mu)+ q )+pnorm( sqrt(n/sigma2)*(mu0-mu)- q )
data_power <- data.frame(mu=mu, power=power)
ggplot(data=data_power) + geom_line(aes(x=mu,y=power))
```




2. We do not assume that $\sigma$ is known anymore and we want to test $H_0$: $\mu = \mu^*$ versus $H_1$: $\mu \neq \mu^*$ for $\mu^* \in \mathbb{R}$.
   a. Propose a test at level 0.05 (steps 1 to 6 of Exercise 1) and create an R function **rej_reg_2** that takes as input a vector of observations $x = (x_1, \ldots, x_n)$, the parameter $\mu^*$ and a level of signficance $\alpha \in (0, 1)$, and outputs a vector of size four containing the lower and upper bound of the acceptation region, the statistic of the test and the result of the test: 0 if $H_0$ is not rejected and 1 if $H_0$ is rejected.

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 $H_1 \ : \ \mu \neq \mu_*$. This is the so called one sample, two-sided t-test.

Step 3: The MLE of $\mu$ is $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Under $H_0$: $\bar{X}_n \sim \mathcal{N}(0, \sigma^2/n)$, this distribution depends on $\sigma$ which is not known. This statistics can't 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 estimated mean is too far from $\mu^*$. So we choose an aceptance region of the following shape: $[-\delta, \delta]$.

Step 5: We choose $\delta$ to obtain a test with significance level $\alpha$, i.e. such that

$$\alpha = \mathbb{P}_{H_0}(|T_n| > \delta) = 2\mathbb{P}_{T_n \sim \mathcal{T}(n-1)}\left(T_n > \delta\right) = 2 - 2F_{\mathcal{T}(n-1)}\left(\delta\right).$$

So that, we choose $\delta = t_{1-\alpha/2}(n-1)$ which is the $1-\alpha/2$ quantile of a Student distribution with $n-1$ degrees of freedom.

Step 6: Finally our test rejects $H_0$ when $|T_n| > t_{1-\alpha/2}(n-1)$. Formally,

$$\phi_2(X_1, \dots, X_n) = \begin{cases} 0 & \text{if } T_n \in \left[-t_{1-\alpha/2}(n-1), t_{1-\alpha/2}(n-1)\right] \\ 1 & \text{otherwise} \end{cases}.$$

```
rej_rec_2 <- function(x, alpha, mu0){
  n <- length(x)
  t_alpha <- qt(p=1-alpha/2,df=n-1)
  mean <- mean(x)
  sd <- sd(x)
  t <- sqrt(n)*(mean-mu0)/sd
  phi <- ((t < -t_alpha)||(t>t_alpha))
  test <- c(-t_alpha, t_alpha,t, phi)
  return(test)
}
```

```
mu0 <- 0
sigma2 <- 4
alpha <- 0.05
n<-10
x <- rnorm(10,mu0,sqrt(sigma2))
rej_rec_2(x,  alpha, mu0)
```

```
## [1] -2.262157  2.262157 -2.291846  1.000000
```

b. Check the significance level of the test through simulations when $\sigma = 2$, $\alpha = 0.05$ and $n = 10$.

```
mu0 <- 0
sigma2 <- 4
alpha <- 0.05
n<- 10

samples <- lapply(1:1e3, function(i) rnorm(n, mu0, sqrt(sigma2)))
tests <- lapply(samples, function(x) rej_rec_2(x, alpha, mu0))
test <- sapply(tests, function(te) te[4])
sum(test)/1e3
```

```
## [1] 0.049
```

3. The athlete S.R. of the first exercise contests the control. He goes through another control where 10 measures of some other level of his blood are done. A usual measure is 0, and doping product can increase or decrease this level. The measures are assumed to be Gaussian but its standard deviation is not known. The following measurements are observed: 2.1, 1.4, -0.1, -0.2, 1.2, -0.7, 0.7, 1.5, 0.5, 0.3 . What is the conclusion of the control for S.R.?

Step 1: Let $x_i$ be the i-th measure. We assume that $(x_1, x_2, \ldots, x_{10})$ is a realization of the random vector $(X_1, \ldots, X_{10})$ where $X_i \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$. The statistical model is then

$$\left(\mathbb{R}^{10}, \mathcal{B}(\mathbb{R}^{10}, \{\mathcal{N}^{\otimes 10}(\mu, \sigma^2); (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+\}\right).$$

$\mu$ represents the true biological level of S.R. and $\sigma$ the standard deviation of all the measurements.

Steps 2, 3, 4 and 5 are already done in question 2.

Step 6: We reject $H_0$ if $T_n$ is not in $\left[-t_{1-\alpha/2}(n-1), t_{1-\alpha/2}(n-1)\right]$. Using the function rej_rec_2, we reject $H_0$ if $T_n$ is not in $[-\delta., \delta]$, whre $\delta \simeq 2.26$, the observed value of the statistics is around 2.41 and we reject $H_0$.

```
x=c( 2.1, 1.4, -0.1, -0.2, 1.2, -0.7, 0.7, 1.5, 0.5, 0.3)
mu0 <- 0
alpha <- 0.05
rej_rec_2(x,alpha,mu0)
```

```
## [1] -2.262157  2.262157  2.412598  1.000000
```

In practise, use the function **t.test** already coded in R!

## Other models

**Exercise 6: Urn**

A urn contains $N$ balls, numbered from 1 to $N$ where $N \in \{99, 101\}$ is unknown. The following balls, which were placed back in the urn after each drawing, were drawn:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 52 | 48 | 60 | 17 | 03 | 25 | 98 | 15 | 51 | 12 | 38 | 76 |
| 25 | 64 | 87 | 23 | 19 | 05 | 82 | 46 | 73 | 09 | 50 | 48 | 34. |

Propose a rule to test $H_0$: $N = 99$ versus $H_1$: $N = 101$, following all the steps described in Exercise 1. Besides compute the probability of type II error of your decision rule.

**Exercise 6: Are financial professionals unscrupulous?**

Roger read in the newspaper that 20% of financial professionals think they need to break the law to succeed: http://www.slate.fr/story/101785/wall-street-enfeindre-loi-reussir

He thinks that this percentage is underestimated and wants to investigate and to test his hypothesis. He only has two financial professionals in his address book; He is going to ask them the following question: "Do you think you need to break the law to succeed?".

1. Assuming that Roger's conclusion only depends on the answers at this question, which decision rule can he establish? What is the probability that Roger is wrong if the percentage of financial professionals is indeed 20%?

2. Roger wants to wrongly conclude that the percentage is strictly greater than 20% with probability 5%. He realizes that he does not need to survey anyone to realize this constraint: he just has to put 19 black chips and 1 red chip in his pocket. Why? What is the drawback of this method?

3. Roger now chooses a card in a deck of 32 cards and write the result before surveying the two financial professionals he knows. Why did he choose this method? What is probability that Roger's conclusion is wrong if the percentage of financial professionals is indeed 20%?

4. Roger feels that such a decision should not be taken on a random basis. What should he do?