Hypothesis testing

Introduction to hypothesis testing

Exercise 1: 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

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

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 ϕ_1 is the probability to reject H_0 (i.e. $X \neq 5$) when H_0 is true i.e. when $p = 0.5$:

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

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

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

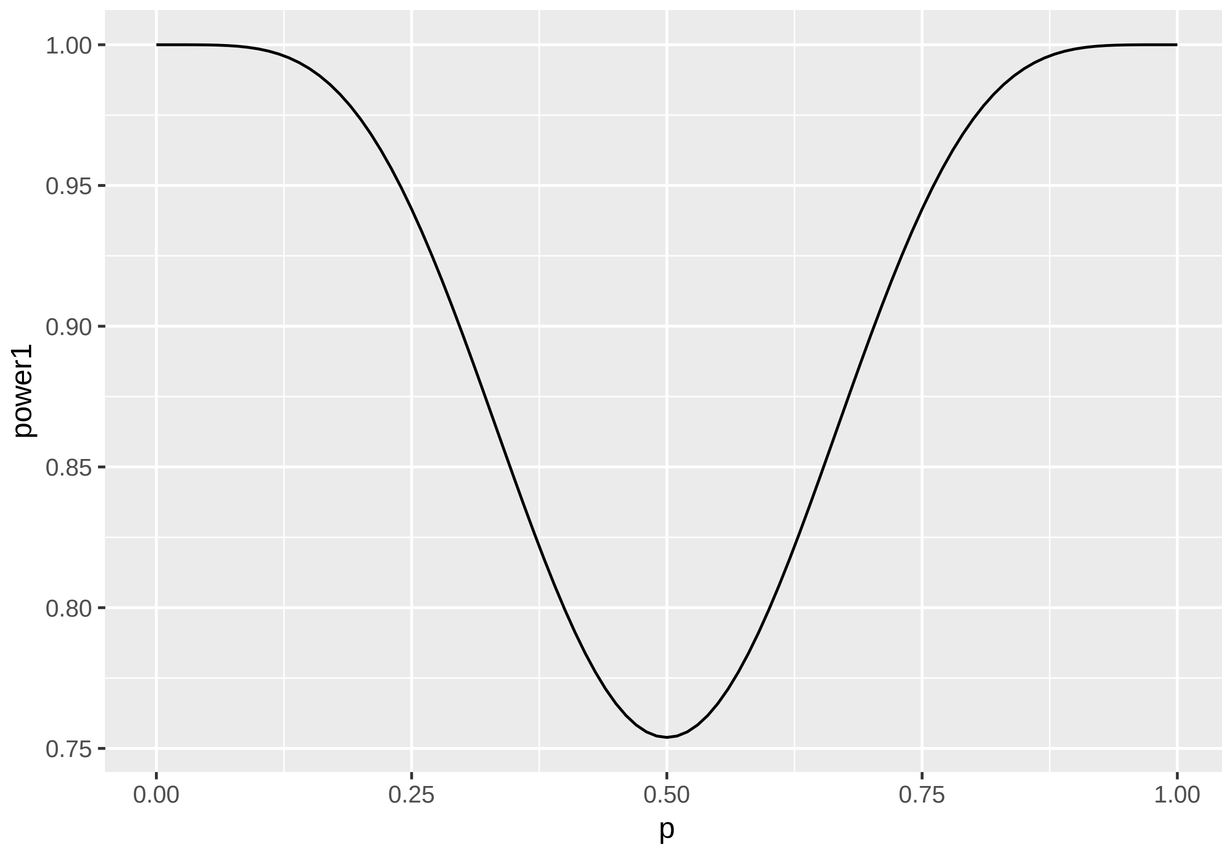
The power function of ϕ_1 is

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

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. $\text{Card}(\Theta_0) = 1$) $\alpha_{\phi_1} = \beta_{\phi_1}(1/2)$.

```r
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 ϕ_2 is the probability to reject H_0 (i.e. $X \in \{0,1,9,10\}$) when H_0 is true i.e. when $p = 0.5$:

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

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

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

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

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

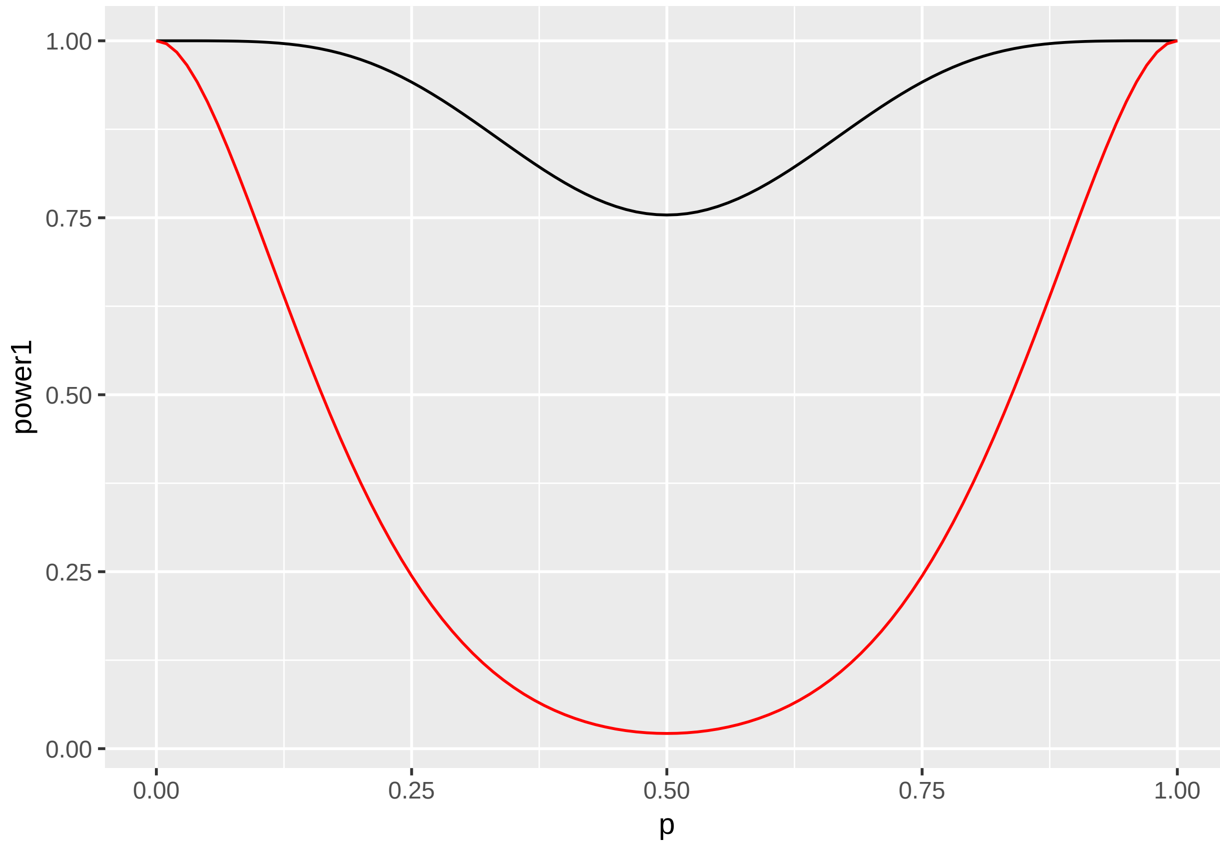
The power function of ϕ_2 is

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

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

```
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

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

The significance level of this test ϕ_3 is the probability to reject H_0 (i.e. $X \in \{0,2,10\}$) when H_0 is true i.e. when $p = 0.5$:

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

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

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

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

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

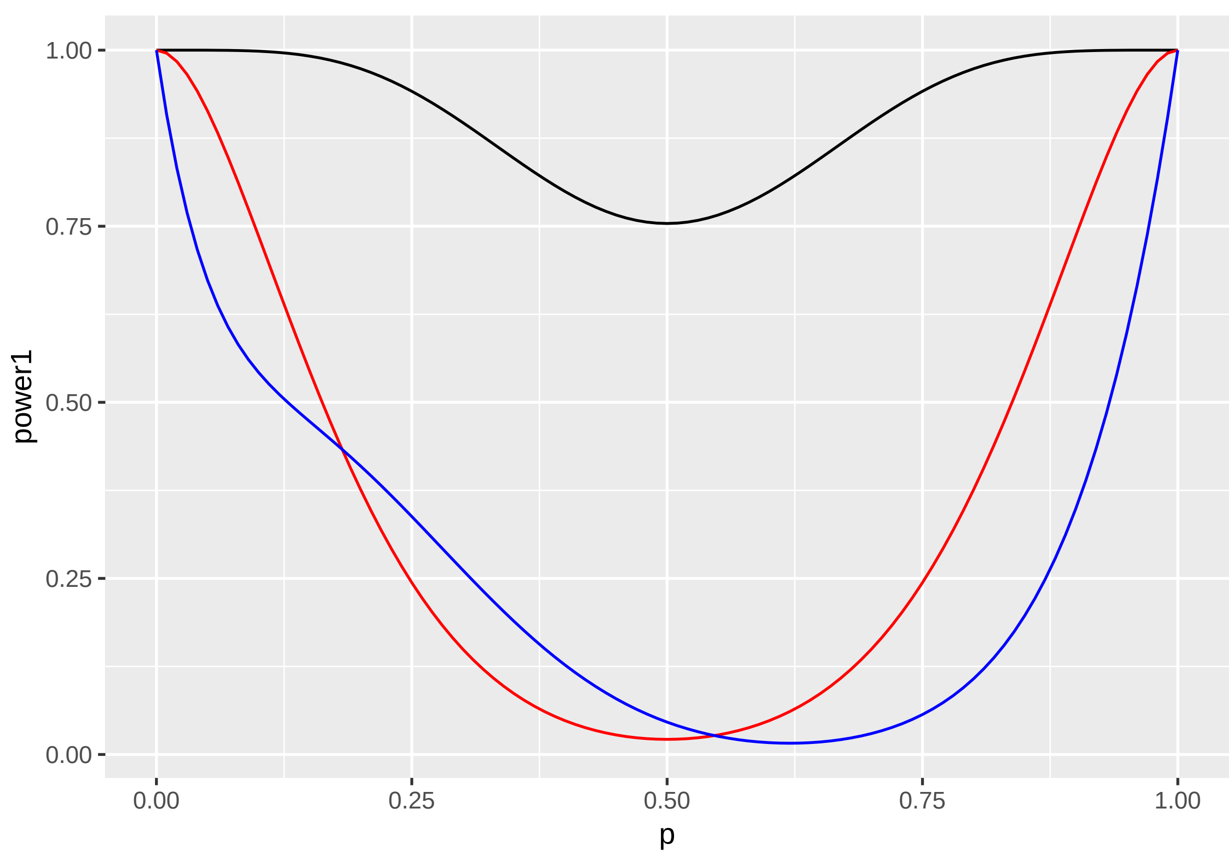
The power function of ϕ_3 is

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

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

```
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

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

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

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

```
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

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

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 δ ($\epsilon = \delta/30$).

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

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

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

```
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 δ 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

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

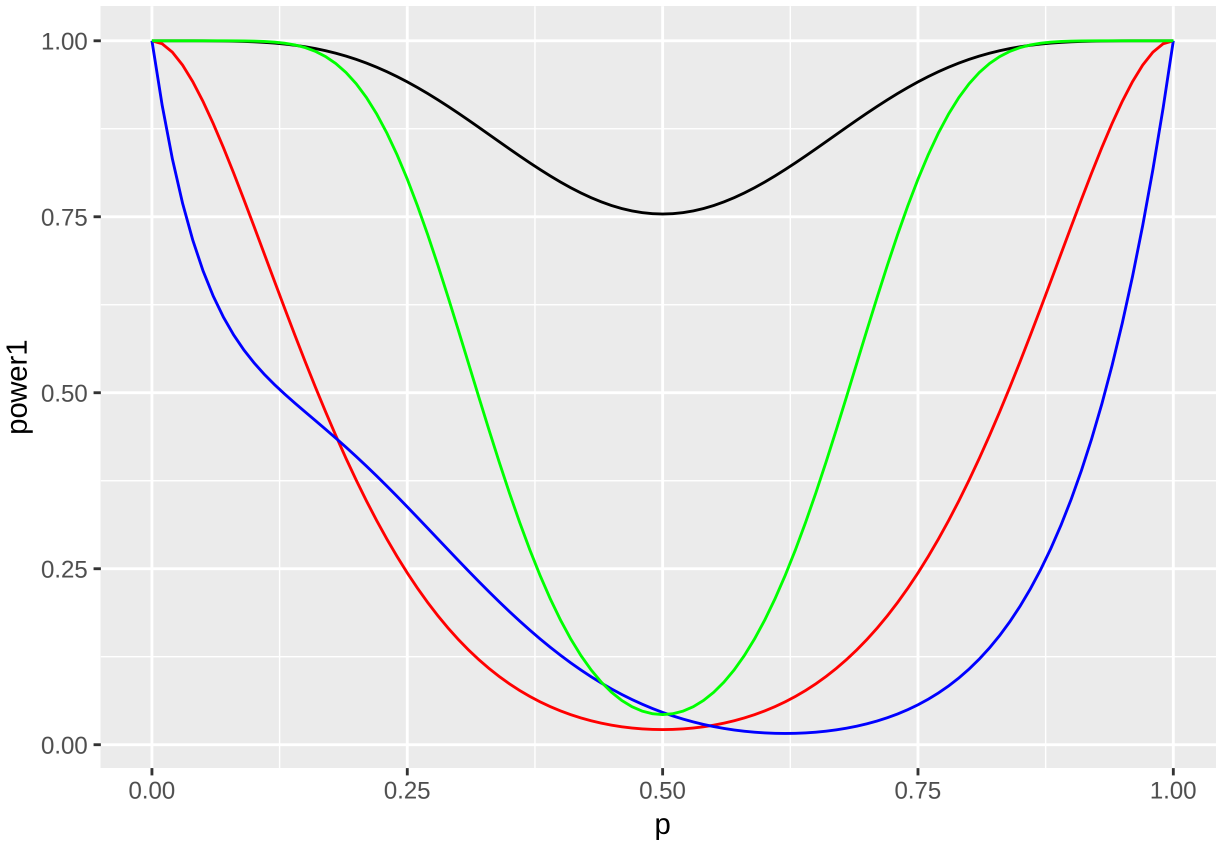
The power function of ϕ_4 is

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

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

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

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

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

```r
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?

ϕ_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. ϕ_4 is then the test we prefer.

Choice of the hypotheses

Exercise 2: 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 a positive 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 $(x_1, \dots x_n)$ is a realization of $(X_1, \dots, X_n)$, where X_i are i.i.d. from the Bernoulli distribution $B(\theta)$ where θ 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 θ represents the average score among children with ADD and σ 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 σ 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 $H_1 \ : \ p \neq (1/10, \dots, 1/10)$.