Computational statistics
Week 1

Léa Bresson

List of exercises:

- Exercise 1.22
- Exercise 2.2
- Exercise 2.7
- Exercise 2.19

1 Exercise 1.22

1.1 Suppose that $X \sim f(x|\theta)$, with prior distribution $\pi(\theta)$, an interest is in the estimation of the parameter $h(\theta)$.

a) Using the loss function $L(\delta, h(\theta))$, show that the estimator that minimizes the Bayes risk $\int\int L(\delta, h(\theta))f(x|\theta)\pi(\theta)dxd\theta$, is given by the estimator θ that minimizes (for each x)

$$\int L(\delta, h(\theta))\pi(\theta|x)d\theta.$$

Solution:

The Bayes risk can be written :

$$R_\pi = \int_\Theta \int_\chi L(\delta, h(\theta))f(x|\theta)\pi(\theta)dxd\theta$$

with $x \in \chi$ and $\theta \in \Theta$.

Suppose χ and Θ are σ-finite measure spaces. By definition, $L(\delta, h(\theta)) \geq 0$ so $L(\delta, h(\theta))f(x|\theta)\pi(\theta)$ is measurable and non-negative. In this context, we can use the Tonelli's theorem to invert the ordering of the two integrals.

$$R_\pi = \int_\chi \int_\Theta L(\delta, h(\theta))f(x|\theta)\pi(\theta)d\theta dx$$

The Bayes theorem (see slide 16 or page 12 of the book) states that:

$$f(x|\theta)\pi(\theta) = Z(x)\pi(\theta|x)$$

with $Z(x) = \int f(x|\theta)\pi(\theta)d\theta$ is the marginal density of X (or Bayesian evidence).

Hence,

$$\begin{aligned}
R_\pi &= \int_\chi \int_\Theta L(\delta, h(\theta))Z(x)\pi(\theta|x)d\theta dx \\
&= \int_\chi \phi(\pi\theta|x)Z(x)dx
\end{aligned}$$

with $\phi(\pi\theta|x) = \mathbb{E}\Big[L(\delta, h(\theta)|x\Big] = \int_\Theta L(\delta, h(\theta))\pi(\theta|x)d\theta$, the posterior loss.

We can see that minimizing the Bayes risk for each x is equivalent to minimizing the posterior loss $\phi(\pi\theta|x) = \int_\Theta L(\delta, h(\theta))\pi(\theta|x)d\theta$.

To conclude the estimator that minimizes the Bayes risk is given by the estimator θ that minimizes (for each x)

$$\int_\Theta L(\delta, h(\theta))\pi(\theta|x)d\theta.$$

b) For $L(\delta, \theta) = \|h(\theta) - \delta\|^2$, show that the Bayes estimator of $h(\theta)$ is $\delta^* = \mathbb{E}^*\Big[h(\theta)|x\Big]$.

Solution:

The Bayes estimator minimizes, for each x, $\phi(\pi\theta|x)$.

$$\begin{aligned}
\phi(\pi\theta|x) &= \int_\Theta L(\delta, h(\theta))\pi(\theta|x)d\theta \\
&= \mathbb{E}\Big[L(\delta, h(\theta)|x)\Big] \\
&= \mathbb{E}\Big[\ \|h(\theta) - \delta\|^2 \,|x\Big] \\
&= \mathbb{E}\Big[\ \|h(\theta)\|^2 + \|\delta\|^2 - 2 < \delta, h(\theta > |x\Big] \\
&= \mathbb{E}\Big[\ \|h(\theta)\|^2 + \delta^2 - 2 < \delta, h(\theta) > |x\Big] \\
&= \mathbb{E}\Big[\ \|h(\theta)\|^2 \,|x\Big] + \delta^2 - 2 < \delta, \mathbb{E}\Big[h(\theta)|x\Big] >
\end{aligned}$$

The minimum is such that:

$$\phi^{*'}(\pi\theta|x) = 2\delta^* - 2\mathbb{E}^*\Big[h(\theta)|x\Big] = 0$$

$$\Rightarrow \delta^* = \mathbb{E}^*\Big[h(\theta)|x\Big].$$

It is a minimum since $\phi^{*''}(\pi\theta|x) > 0$.

c) For $L(\delta,\theta) = |h(\theta) - \delta|$, show that the Bayes estimator of $h(\theta)$ is the median of the posterior distribution.

Solution:

The Bayes estimator minimizes, for each x, $\phi(\pi\theta|x)$.

$$\begin{aligned}
\phi(\pi\theta|x) &= \int_{\Theta} L(\delta, h(\theta))\pi(\theta|x)d\theta \\
&= \int_{\Theta} |h(\theta) - \delta|\pi(\theta|x)d\theta \\
&= \int_{-\infty}^{\delta} (\delta - h(\theta))\pi(\theta|x)d\theta + \int_{\delta}^{+\infty} (h(\theta) - \delta)\pi(\theta|x)d\theta \\
&= \delta \int_{-\infty}^{\delta} \pi(\theta|x)d\theta - \int_{-\infty}^{\delta} h(\theta)\pi(\theta|x)d\theta + \int_{\delta}^{+\infty} (h(\theta)\pi(\theta|x)d\theta - \delta \int_{\delta}^{+\infty} \pi(\theta|x)d\theta.
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \phi'(\pi\theta|x) &= \int_{-\infty}^{\delta} \pi(\theta|x)d\theta + h(\delta)\pi(\delta|x) - h(\delta)\pi(\delta|x) - h(\delta)\pi(\delta|x) - \int_{\delta}^{\infty} \pi(\theta|x)d\theta + h(\delta)\pi(\delta|x) \\
&= \int_{-\infty}^{\delta} \pi(\theta|x)d\theta - \int_{\delta}^{\infty} \pi(\theta|x)d\theta \\
&= \mathbb{P}(h(\theta) \leq \delta|x) - \mathbb{P}(h(\theta) > \delta|x) \\
&= 2\mathbb{P}(h(\theta) \leq \delta|x) - 1
\end{aligned}$$

Hence, $\phi^{*'}(\pi\theta|x) = 0$ if $\mathbb{P}(h(\theta) \leq \delta^*|x) = \frac{1}{2}$.

To conclude the Bayes estimator of $h(\theta)$ is such that $\mathbb{P}(h(\theta) \leq \delta^*|x) = \frac{1}{2}$, i.e., δ^* is the conditional median of $h(\theta)$ given x.

It is a minimum since $\phi^{*''}(\pi\theta|x) = 2\pi(\delta^*|x) \geq 0$.

2 Exercice 2.7

2.1 Establish the properties of the Box–Muller algorithm of Example 2.8. If U_1 and U_2 are iid $\mathcal{U}[0,1]$, show that:

a) The transforms $X_1 = \sqrt{-2log(U_1)}\cos(2\pi U_2)$, $X_2 = \sqrt{-2log(U_1)}\sin(2\pi U_2)$, are iid $\mathcal{N}(0,1)$.

Solution:

First of all, we can rewrite U_1 and U_2 as :

$$X_1^2 + X_2^2 = -2log(U_1) \Rightarrow U_1 = \exp\big[-\frac{1}{2}(X_1^2+X_2^2)\big]$$

and,

$$\frac{X_1}{X_2} = \tan(2\pi U_2) \Rightarrow U_2 = \frac{\arctan(\frac{X_1}{X_2})}{2\pi}.$$

Let us define $h : (U_1, U_2) \mapsto (X_1, X_2)$, an homeomorphism between $[0,1]^2$ and $\mathcal{R}^2$. We have $h^{-1} : (X_1, X_2) \mapsto (U_1, U_2)$ with : $h^{-1} : (X_1, X_2) = \Big\{\exp[-\frac{1}{2}(X_1^2+X_2^2)], \frac{\arctan(\frac{X_1}{X_2})}{2\pi}\Big\}$.

In this context, the joint density of (X_1, X_2) can be expressed in terms of the Jacobian of h^{-1} :

$$
\begin{aligned}
f_{X_1,X_2} &= |J_{h^{-1}}(X_1,X_2)| \\
&= \begin{vmatrix} \frac{\partial h_1^{-1}}{\partial x_1} & \frac{\partial h_1^{-1}}{\partial x_2} \\ \frac{\partial h_2^{-1}}{\partial x_1} & \frac{\partial h_2^{-1}}{\partial x_2} \end{vmatrix} \\
&= \begin{vmatrix} x_1 \exp\big[-\frac{X_1^2+X_2^2}{2}\big] & -x_2 \exp\big[-\frac{X_1^2+X_2^2}{2}\big] \\ \frac{1}{2\pi}\frac{-X_2}{X_1^2+X_2^2} & \frac{1}{2\pi}\frac{X_1}{X_1^2+X_2^2} \end{vmatrix} \\
&= \frac{1}{2\pi}\exp\big[-\frac{(X_1^2+X_2^2)}{2}\big]
\end{aligned}
$$

Hence, we have :

$$
\begin{aligned}
f_{X_1,X_2} &= \frac{1}{\sqrt{2\pi}}\exp\big[-\frac{x_1^2}{2}\big] \times \frac{1}{\sqrt{2\pi}}\exp\big[-\frac{x_2^2}{2}\big] \\
&= f_{X_1} \times f_{X_2}
\end{aligned} \tag{1}
$$

We can clearly see that X_1 et X_2 are iid $\mathcal{N}(0,1)$.

b) The polar coordinates are distributed as:
i): $r^2 = (X_1^2 + X_2^2) \sim \chi_2^2$,
ii): $\theta = \arctan \frac{X_2}{X_1} \sim \mathcal{U}[0, 2\pi]$.

Solution:

i): The distribution of the sum of the squares of 2 independent standard normal random variables is a chi-square with 2 degree of freedom.
In our case, X_1^2 and X_2^2 are independent since X_1 and X_2 are iid (question a). Therefore, r^2 is distributed as a chi-square random variable with 2 degree of freedom:

$$r^2 \sim \chi_2^2$$

Proof :

1. *Let's prove that the square of a standard normal random variable is a chi-square random variable with 1 degree of freedom.*

   Let X be a standard normal random variable with a probability density function $f_X(x)$. We define Z such that $Z = X^2$.

   - For $z < 0$, the distribution function of Z, $P(Z \leq z) = 0$ since Z is a square. Therefore, $f_Z(z) = 0$

   - For $z \geq 0$, the distribution function of Z is :

$$\begin{aligned} P(Z \leq z) &= P(X^2 \leq z) \\ &= P(-z^{\frac{1}{2}} \leq X \leq z^{\frac{1}{2}}) \\ &= \int_{-z^{\frac{1}{2}}}^{z^{\frac{1}{2}}} f_X(x)dx \end{aligned} \tag{2}$$

   Therefore, for $z \geq 0$, the density of Z is:

$$\begin{aligned} f_Z(z) &= \frac{dP(Z \leq z)}{dz} \\ &= \frac{\int_{-z^{\frac{1}{2}}}^{z^{\frac{1}{2}}} f_X(x)dx}{dz} \\ &= f_X(z^{\frac{1}{2}})\frac{dz^{\frac{1}{2}}}{dz} - f_X(-z^{\frac{1}{2}})\frac{d\text{-}z^{\frac{1}{2}}}{dz} \\ &= \frac{1}{\sqrt{2\pi}} \exp\Big[-\frac{1}{2}(z^{\frac{1}{2}})^2\Big] \times \frac{1}{2}z^{\frac{-1}{2}} - \frac{1}{\sqrt{2\pi}} \exp\Big[-\frac{1}{2}(-z^{\frac{1}{2}})^2\Big] \times \frac{-1}{2}z^{\frac{-1}{2}} \\ &= \frac{1}{\sqrt{2\pi}}z^{\frac{1}{2}} \exp\Big[-\frac{1}{2}z\Big] \\ &= \frac{1}{2^{\frac{-1}{2}}\Gamma(\frac{1}{2})}z^{\frac{1}{2}-1} \exp\Big[-\frac{1}{2}z\Big] \text{ because } \Gamma(\frac{1}{2}) = \sqrt{\pi}. \end{aligned} \tag{3}$$

We recognize the density function of a chi-square random variable with 1 degree of freedom.

2. *Let's prove that the sum of independent chi-square random variables is a chi-square random variable.*

   Let $Z = \sum_{i=1}^{k} Z_i$ and $M_Z(t)$ its moment generating function. Suppose that the random variable Z_i is a chi-square random variable with n_i degrees of freedom and all the random variables are independent, thus:

$$\begin{aligned} M_Z(t) &= \prod_{i=1}^{k} M_{Z_i}(t) \\ &= \prod_{i=1}^{k} (1-2t)^{\frac{-n_i}{2}} \\ &= (1-2t)^{-\sum_{i=1}^{k} \frac{n_i}{2}} \\ &= (1-2t)^{\frac{-n}{2}} \end{aligned} \tag{4}$$

   We can see that the moment generating function of Z is the moment generating function of a chi-square random variable with $n = \sum_{i=1}^{k} n_i$ degree of freedom.

3. *Combining the two facts above, we obtain that the sum of squares of 2 independent standard normal random variables is a chi-square random variable with 2 degree of freedom.*

ii): From question a, we know that $U_2 = \frac{\arctan(\frac{X_1}{X_2})}{2\pi} = \frac{\theta}{2\pi} \Rightarrow \theta = 2\pi U_2$.

Since $U_2 \sim \mathcal{U}[0,1]$, we have:

$$\theta \sim \mathcal{U}[0, 2\pi]$$

.

c) Establish that $\exp(-\frac{r^2}{2}) \sim \mathcal{U}[0,1]$, and so r^2 and θ can be simulated directly.

Solution:

From a), we know that: $X_1^2 + X_2^2 = -2log(U_1)$.

From b), we have that: $r^2 = X_1^2 + X_2^2$.

Hence, $r^2 = -2log(U_1) \Rightarrow \exp(-\frac{r^2}{2}) = U_1$ and consequently, $\exp(-\frac{r^2}{2}) \sim \mathcal{U}[0,1]$.

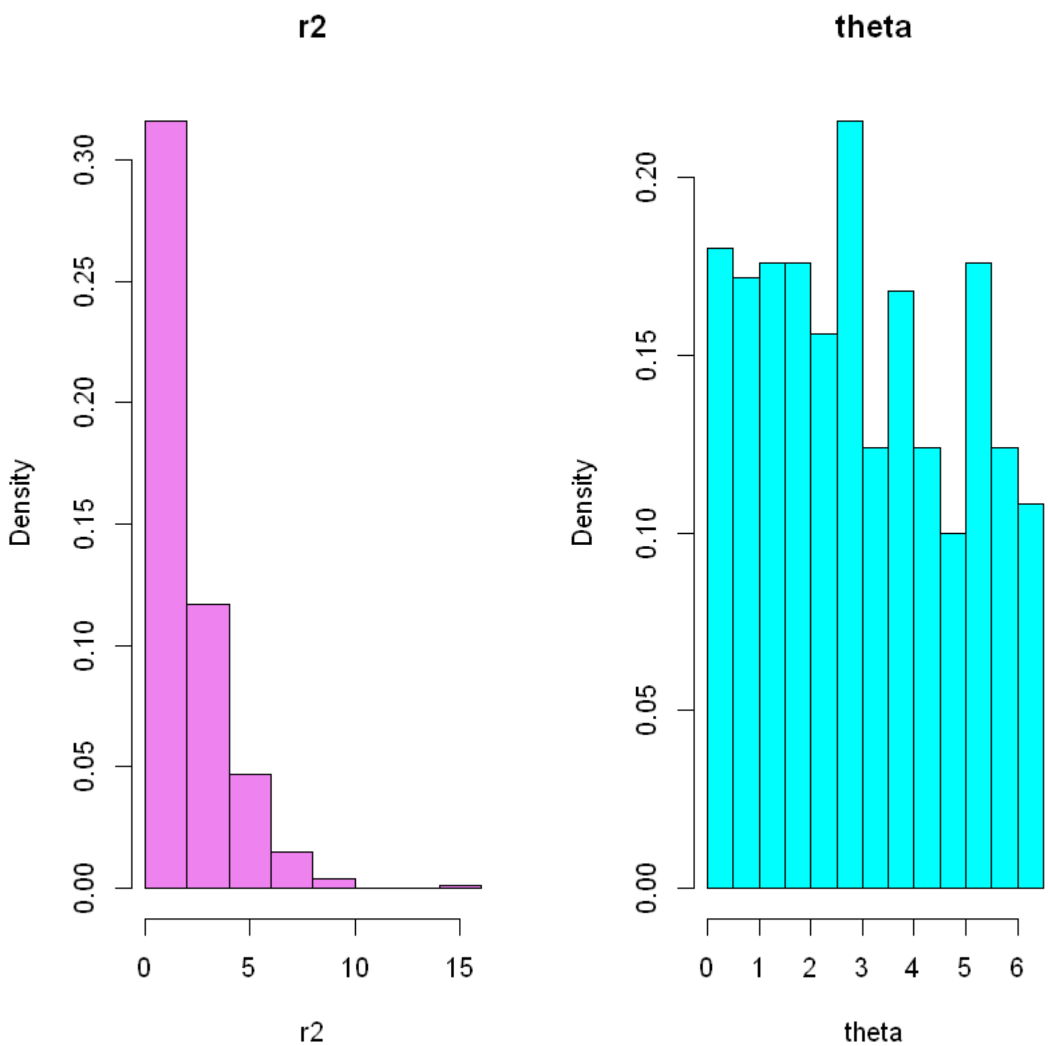
To conclude, we can simulate r^2 and θ directly from the uniform distribution on [0,1] since:

- $r^2 = -2log(U_1)$ and $\theta = 2\pi U_2$, and

- U_1 and U_2 are iid $\mathcal{U}[0,1]$.

Example of a simulation using R:

```
In [1]: U1 = runif(500,0,1)
        U2 = runif(500,0,1)
        r2 = -2*log(U1)
        theta = 2*pi*U2
        par(mfrow=c(1,2))
        hist(r2, freq=F, col = "violet", main="r2")
        hist(theta, freq=F, col= "cyan", main="theta")
```



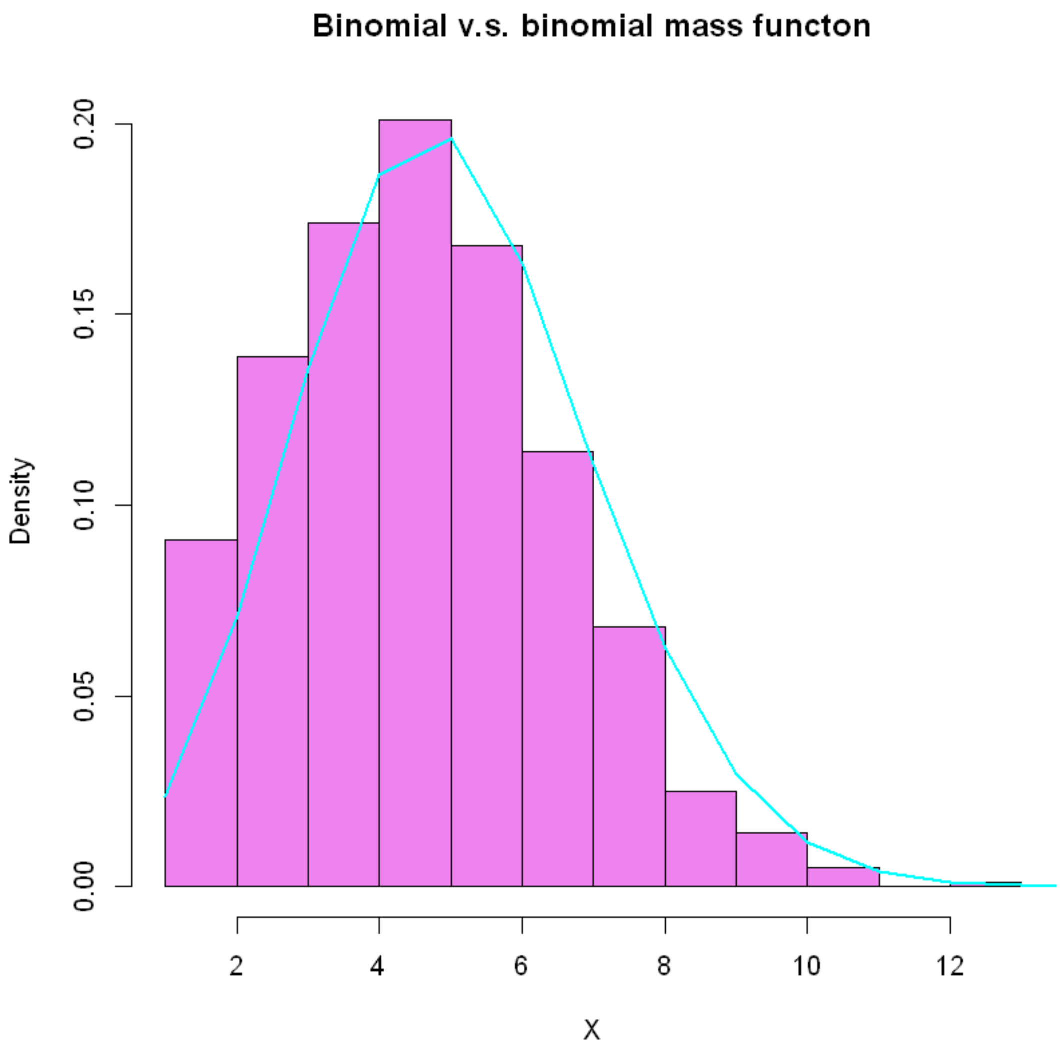
3 Exercice 2.2

3.1 a) Generate a binomial Bin(n, p) random variable with n = 25 and p = .2. Make a histogram and compare it to the binomial mass function, and to the R binomial generator.

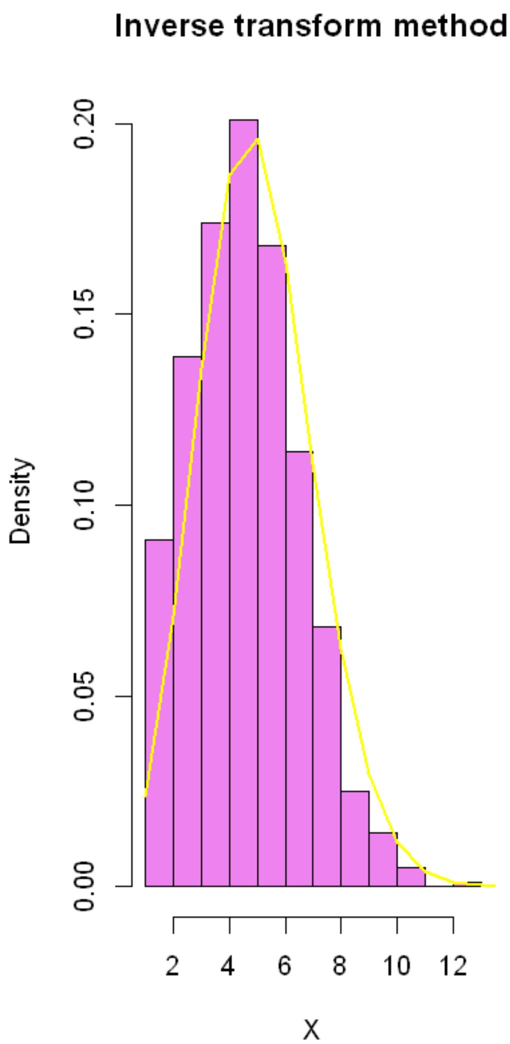
Comment: We have seen in lecture 1 that if $U \sim U_{[0,1]}$, then the random variable $F^{-1}(U)$ has the distribution F (Probability Integral Transform). It is the starting point of the inverse transform method that I am going to use: to generate a random variable $x \sim F$, I generate a random variable $u \sim U_{[0,1]}$ and then I use the transform $x = F^{-1}(u)$.

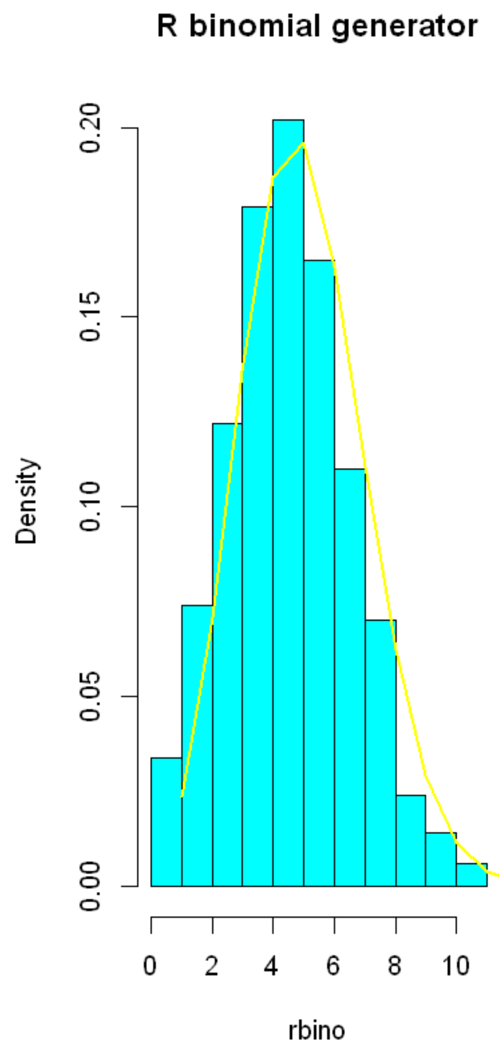
In [2]:
```
n = 25
p = 0.2

N = 1000 # number of simulations

# 1) I use the inverse transform method for generating binomial random variables
Binomial = function(N,n,p){
    proba = pbinom(0:n,n,p)
    X = rep(0,N)
    for (i in 1:N){
        u = runif(1)
        X[i] = 0 + sum(proba<u)
    }
    return(X)
}

# 2) I use the R binomial generator for generating binomial random variables
rbino = rbinom(N, n, p)
```

In [3]:
```
# Histogram : binomial (inverse transform method) v.s. binomial mass function

X = Binomial(N,n,p)
hist(X, col="violet", freq=F, main="Binomial v.s. binomial mass functon")
# Mass function
lines(1:n, dbinom(1:n,n,p), lwd = 2, col="cyan")
```




```
In [4]: # Histogram : binomial (inverse transform method) v.s. the R binomial generator

        par(mfrow=c(1,2))
        hist(X,freq=F, col="violet", main="Inverse transform method")
        lines(1:n, dbinom(1:n,n,p), lwd = 2, col="yellow")

        hist(rbino,freq=F, col="cyan", main="R binomial generator")
        lines(1:n, dbinom(1:n,n,p), lwd = 2, col="yellow")
```





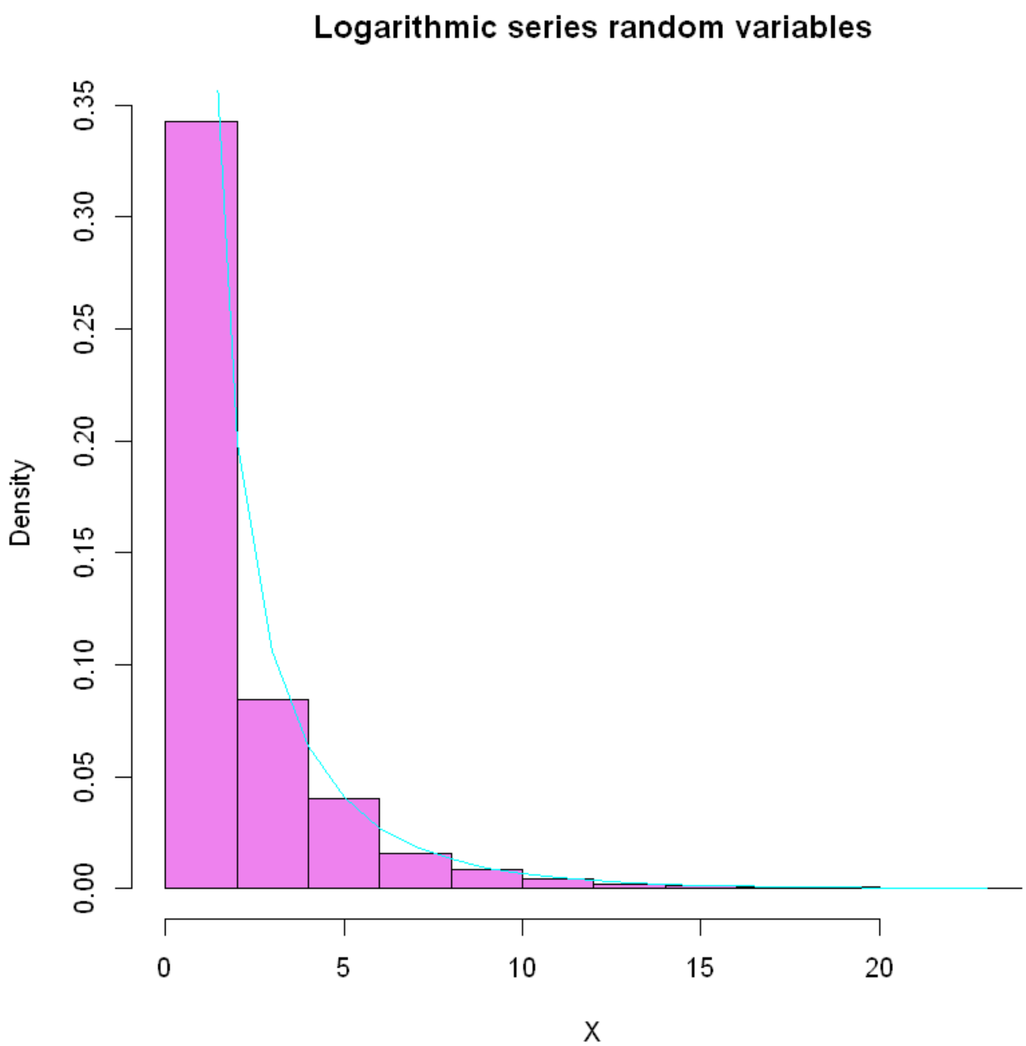
Comment: The histograms are very similar. The transform method gives satisfactory results.

3.2 b) Generate 5,000 logarithmic series random variables. Make a histogram and plot the mass function.

In [5]:
```
# I use the inverse transform method:

logarithmic = function(N,p){
    x = seq(1,35,1)
    mass = (-(1-p)^x)/(x*log(p))
    mass_f = cumsum(mass)
    X = rep(0,N)
```

```
    for (i in 1:N){
        unif = runif(1)
        X[i] = x[1] + sum(mass_f < unif)
    }
    return (X)
}
```

```
In [6]: N = 5000
        p = 0.2
        X = logarithmic(N,p)
        hist(X, freq=F, col = "violet", main="Logarithmic series random variables")
        X = seq(min(X), max(X), 1)
        lines((-(1-p)^X)/(X*log(p)), lwd = 1.5, col="cyan")
```



Conclusion: To conclude, the inverse transform method gives satisfactory method with the binomial or logarithmic series random variables.

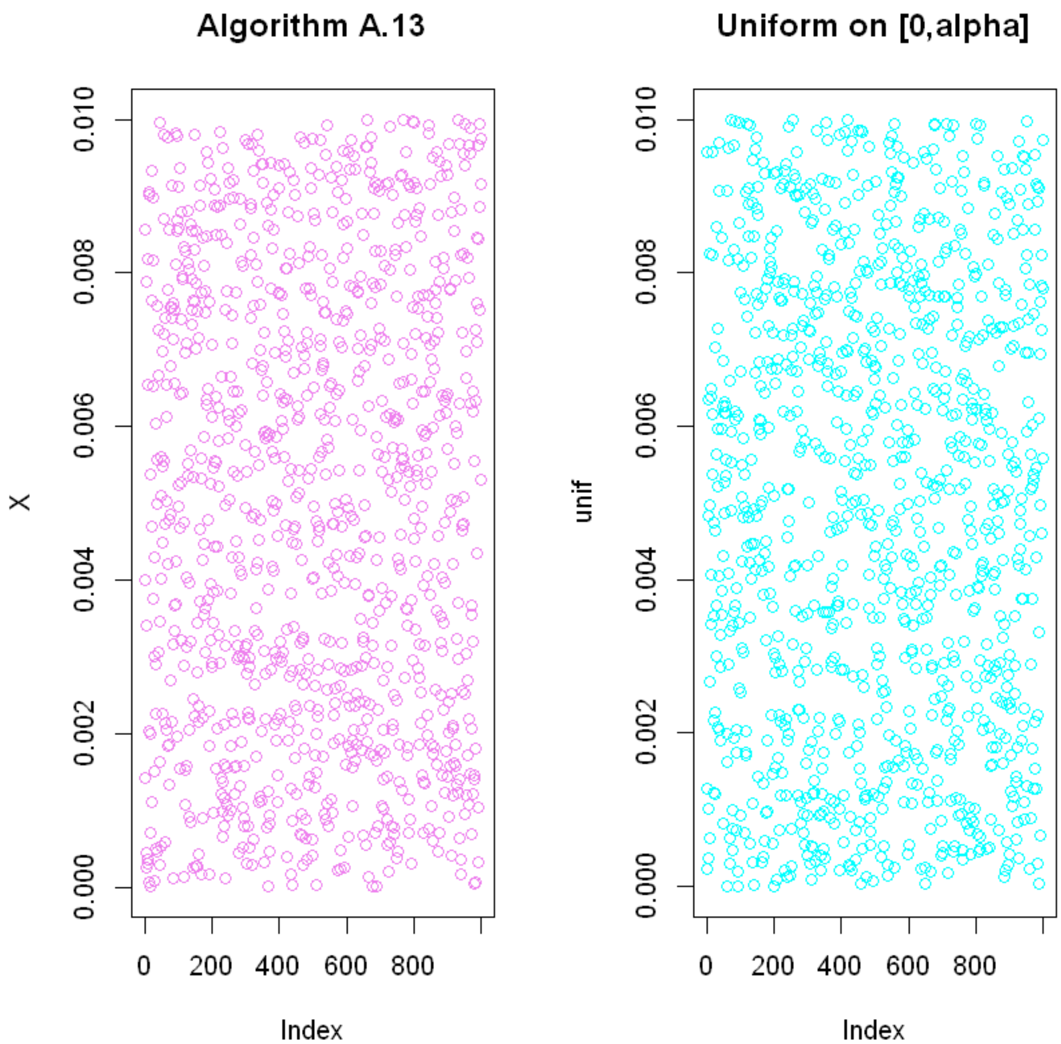
4 Exercice 2.19

4.1 For $\alpha \in [0,1]$, show that the algorithm A.13 produces a simulation from $U([0,\alpha])$. Compare it with the transform αU, $U \sim U(0,1)$ for values of α close to 0 and close to 1.

```
In [7]: # Algorithm A.13 :
        Generateur1 = function(N, alpha){
            X = seq(1,N) # initialisation
            for (i in 1: N){
                    #X[i] = runif(1)
                        while (X[i] >= alpha){
                            X[i] = runif(1)
                        }
                    }
            return(X)
        }
```

1. Comparison between the algorithm A.13 and simulations from $U([0,\alpha])$:

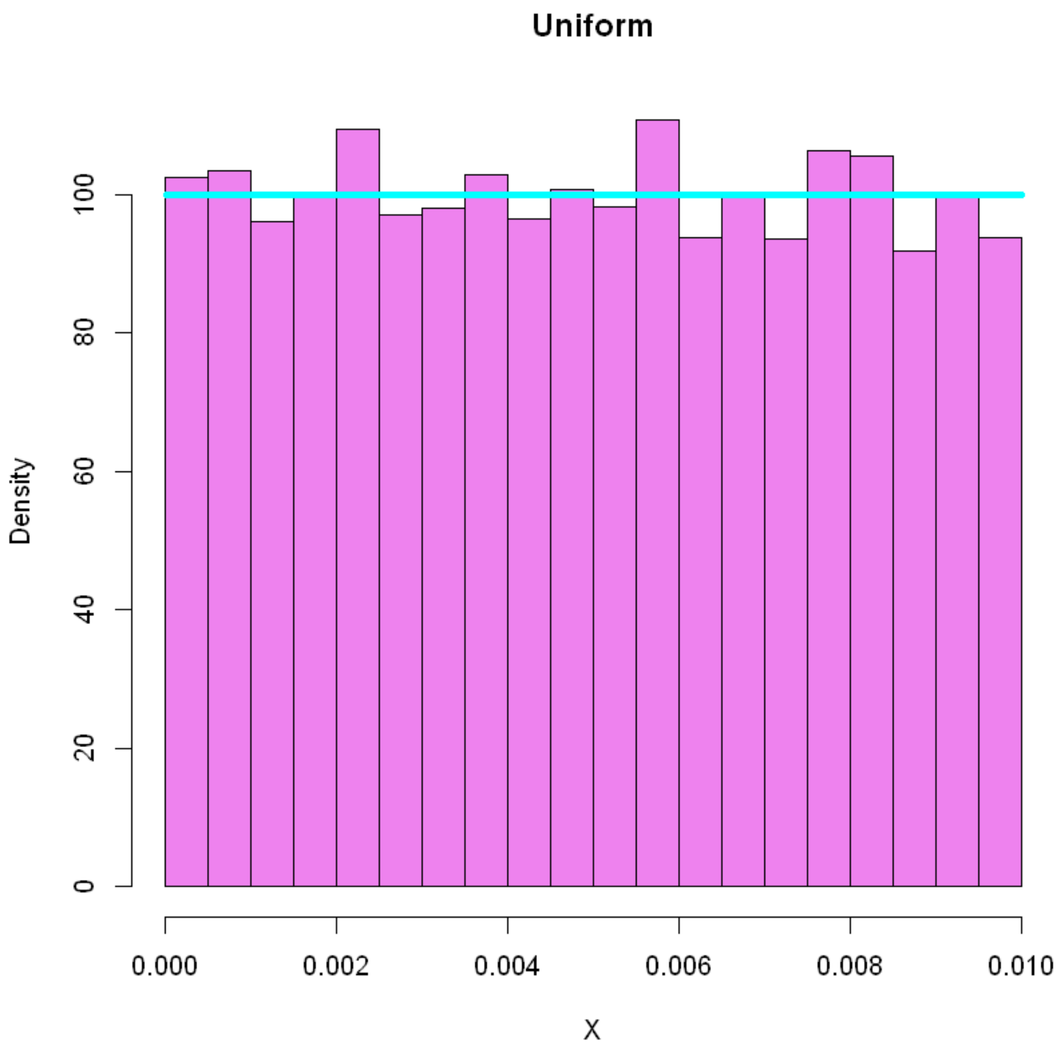
```
In [8]: N = 1000 # Number of simulations
        alpha = 0.01
        #alpha=runif(1) # Results are the same using alpha = runif(1)
        X = Generateur1(N, alpha)
        u = seq(min(X), max(X), length.out = N)
        unif = runif(u,0,alpha)
        par(mfrow=c(1,2))
        plot(X, col ='violet', main="Algorithm A.13")
        plot(unif,col='cyan', main="Uniform on [0,alpha]")
```



The graphs are very similar.

We are now going to compare the histogram given by the algorithm A.13 with the probability density function of a random variable following a uniform distribution on $[0, \alpha]$.

```
In [9]: N = 10^4 # We increase the number of simulations in order to have a better view
        X = Generateur1(N, alpha)
        u = seq(min(X), max(X), length.out = N)
        hist(X, freq=F, col = "violet", main="Uniform")
        lines(u, dunif(u,0,alpha), lwd = 4, col="cyan")
```



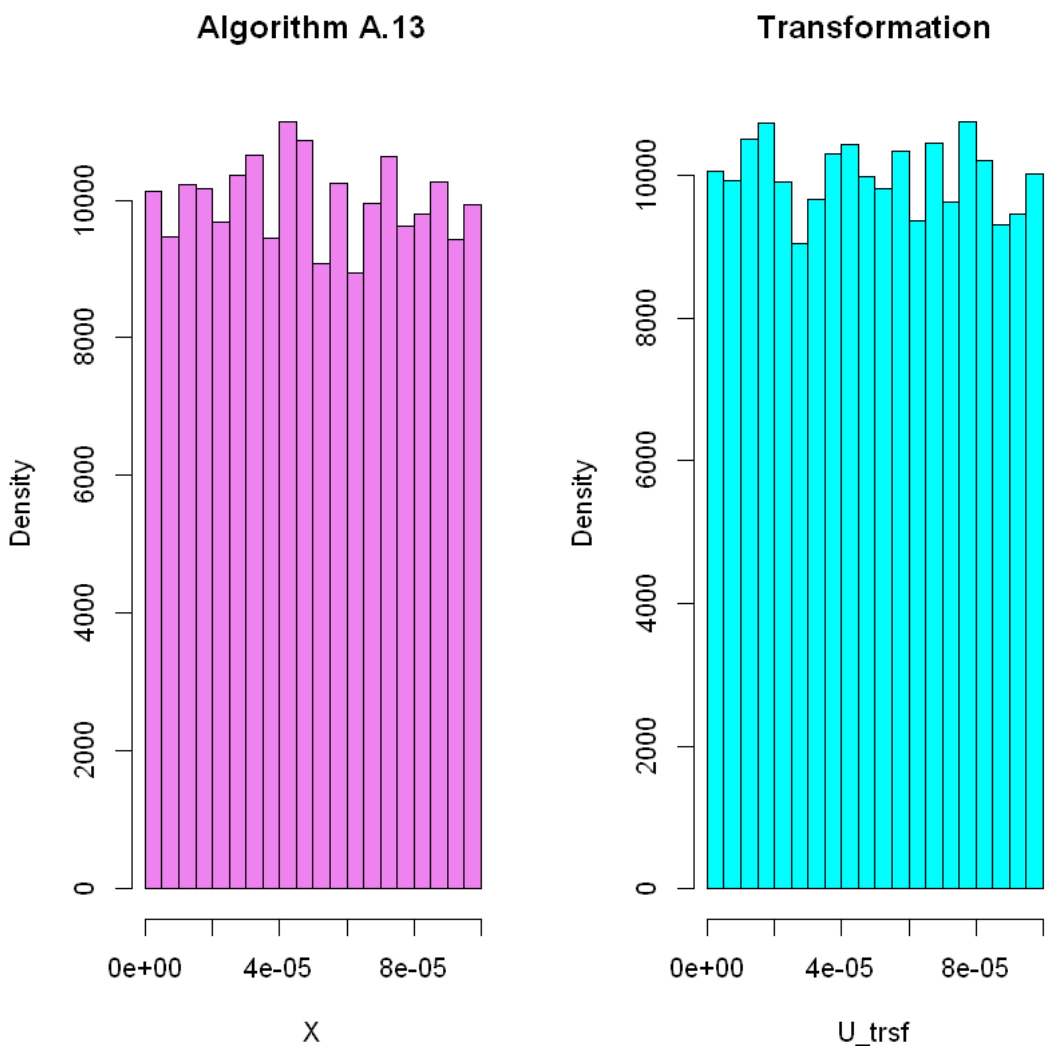
Comment: To conclude, the algorithm A.13 produces a simulation from $U([0,\alpha])$. In fact, the algorithm picks the random variables with uniform distribution on $[0,1]$ in the range of $[0,\alpha]$ where they are still uniformly distributed.

Note that results are the same using αU, $U \sim U(0,1)$ (i.e. alpha=runif(1) in the code above) instead of alpha = 0.01 (arbitrary value).

2) Comparison of the algorithm A.13 with the transform αU, $U \sim U(0,1)$ for values of α close to 0 and close to 1

Case 1: α close to 0

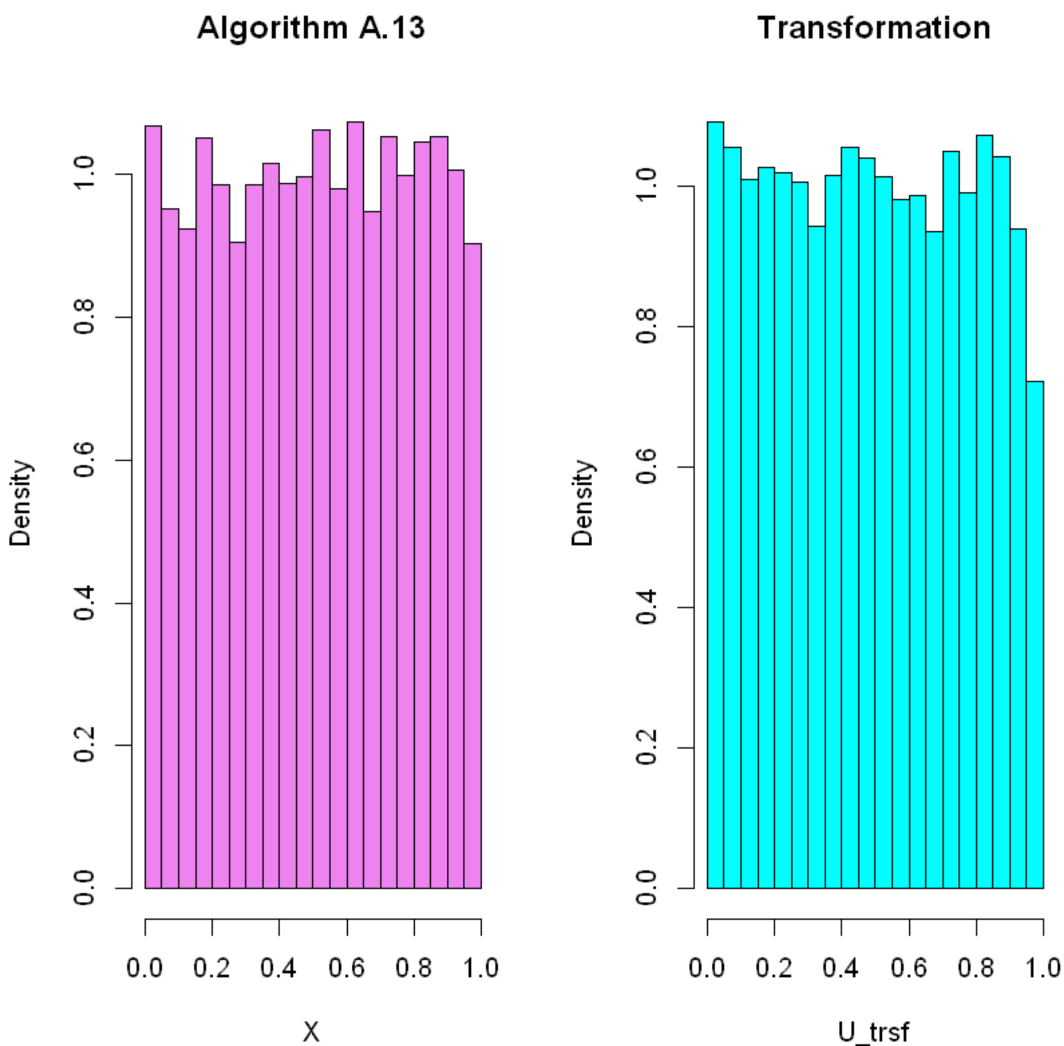
```
In [10]: alpha = 0.0001
         X = Generateur1(N, alpha)
         U_trsf = runif(N) * alpha  # Transform
         Y = seq(min(U_trsf), max(U_trsf), length.out = 1000)

         par(mfrow=c(1,2))
         hist(X, freq=F, col = "violet", main="Algorithm A.13")
         hist(U_trsf, freq=F, col= "cyan", main="Transformation")
```



Case 2: α close to 1

```
In [11]: alpha = 0.99
          X = Generateur1(N, alpha)
```

```
U_trsf = runif(N) * alpha  # Transform
Y = seq(min(U_trsf), max(U_trsf), length.out = 1000)

par(mfrow=c(1,2))
hist(X, freq=F, col = "violet", main="Algorithm A.13")
hist(U_trsf, freq=F, col= "cyan", main="Transformation")
```



Comment: For both cases (α close to 0 or 1), the histograms from the algorithm A.13 and from the transformation are very similar. It shows that the transform αU with $U \sim U(0,1)$ produces a simulation from $U(0,\alpha)$.

Link with the accept-reject algorithm : The simulation of the target density f is done using an instrumental density g. In our case f is the $U(0,1)$ and g is the $U(0,\alpha)$. We have $\frac{f(X)}{Mg(X)} = \alpha$ if

$M = 1$. It means that all the points in $[0, \alpha]$ are accepted.