Module 9 - Monte Carlo methods

Pawel Chilinski

January 21, 2014

Exercise 1.

Assume that the probability distribution of a phone call duration is the following:

duration [min.]	1	2	3	5	8	10	15
probability	0.1	0.15	0.25	0.2	0.15	0.1	0.05

- Generate 1000 observations pertaining to this distribution. Use function sample(). Compare the frequencies of the generated values with theoretical probabilities.

```
> vals<-c(1,2,3,5,8,10,15)
> probs<-c(0.1,0.15,0.25,0.2,0.15,0.1,0.05)
> set.seed(101)
> generated_sample <- sample(x=vals,size=1000,replace=T,prob=probs)
> barplot(table(generated_sample)/1000,xlab="duration[min.]",ylab="probability",border=F)
> names(probs)<-vals
> barplot(probs,border="green",add = TRUE)
> legend("topleft", legend=c("generated","theoretical"), col = c("gray","green"), lty =c(1,1),pch=rep(0,0))
```

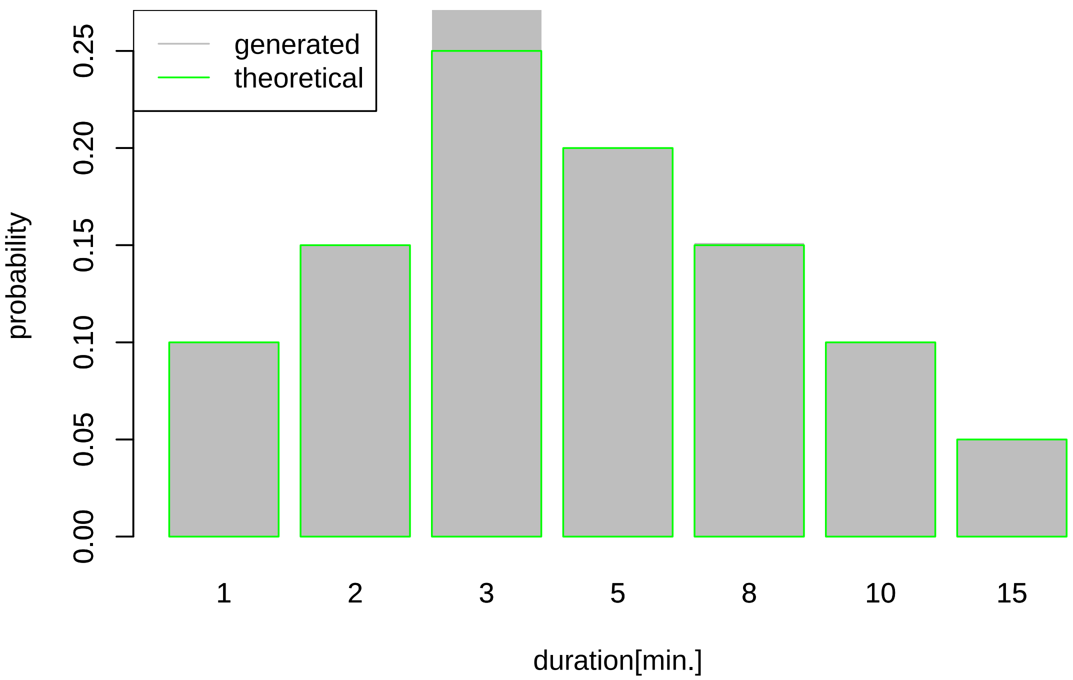

Figure 1: Generating sample and comparing with theoretical probabilities

- Use Monte Carlo method to calculate the probability of an event that a total duration of 10 independent phone calls exceeds one hour.
  To simulate this experiment I generate 1000 samples, where each sample is 10 calls which durations are simulated using sample function. I check what fraction of samples exceed 60 minute in total. This fraction is the estimate of the probability we are looking for.

```
> set.seed(101)
> sum(replicate(1000,sum(sample(x=vals,size=10,replace=T,prob=probs)),simplify = T)>60)/1000
```

```
[1] 0.221
```

Exercise 2.

Find an estimate of the following integral using Monte Carlo method:

$$\int_0^1 exp(-x^2/2)dx$$

based on a pseudo-random sample of size 5000. Calculate standard error of this approximation and construct confidence interval for θ. Is the actual value of the integral covered by the confidence interval constructed using Monte Carlo method?

$$\theta = \int_0^1 exp(-x^2/2)f(x)dx = E[exp(-U^2/2)] \text{ , where } U \sim U(0,1) \text{ and } f(x) = 1$$
$$\hat{\theta} = \frac{1}{m}(exp(-U_1^2/2) + ... + exp(-U_m^2/2))$$

Estimating:

```
> set.seed(101)
> m<-5000
> #simulation
> sample<-exp(-runif(m)^2/2)
> #estimate
> (estimate <- mean(sample))
```

```
[1] 0.8548243
```

Computing SE and confidence interval:

```
> #standard error
> (se<-sqrt(sum((sample-estimate)^2)/(m*(m-1))))
```

```
[1] 0.001726819
```

```
> #CI
> alfa<-0.05
> (ci<-c(estimate-qnorm(1-alfa/2)*se,estimate+qnorm(1-alfa/2)*se))
```

```
[1] 0.8514398 0.8582088
```

Because this integral doesn't have analytical solution I use r integrate function computation as real value:

```
> (val <- integrate(function(x){exp(-x^2/2)},0,1))
```

```
0.8556244 with absolute error < 9.5e-15
```

```
> val$value>ci[1] & val$value<ci[2]
```

```
[1] TRUE
```

So we can see that actual value of the integral is covered by the confidence interval constructed using Monte Carlo method.

Exercise 3.

Suppose that $X \sim N(1,1)$. Generate a pseudo-random sample of size 500 in order to estimate probability $P(exp(X) > 2)$ and standard error of your estimator.

```
> set.seed(101)
> m<-500
> #generating sample
> sample<-rnorm(m)
> #estimate
> (estimate <- sum(exp(sample)>2)/m)
```

```
[1] 0.228
```

```
> #standard error
> sqrt(estimate*(1-estimate)/m)
```

```
[1] 0.01876252
```

Exercise 4.

Data rivers available in the base package of R gives the lengths (in miles) of 141 'major' rivers in North America. Using bootstrap method (with 3000 samples) calculate standard error of the sample mean of rivers' lengths. Find percentile bootstrap confidence interval for an average length of a river.

```
> data(rivers)
> (rivers_mean <- mean(rivers))

[1] 591.1844

> n<-length(rivers)
> m<-3000
> set.seed(101)
> bootstrap_means<-replicate(m,mean(sample(rivers,n,replace=T)))
> # standard error
> (se <- sqrt(sum((bootstrap_means-mean(bootstrap_means))^2)/(m-1)))

[1] 41.1667

> #confidence interval
> alfa<-0.05
> (ci <- c(rivers_mean - quantile(bootstrap_means-rivers_mean,1-alfa/2),
+                                          rivers_mean - quantile(bootstrap_means-rivers_mean,alfa/2)))

   97.5%     2.5%
507.2961 664.9021
```

Exercise 5.

Using the quantile transformation method generate n = 200 random numbers pertaining to exponential distribution with parameter $\lambda = 1.5$. Draw histogram of the sample and compare it with theoretical density of exponential distribution.

To use quantile transformation method we have to find inverse of CDF for exponential distribution:

$$\begin{aligned}
&F(x) = 1 - e^{-\lambda x}\\
&1 - F = e^{-\lambda x}\\
&log(1 - F) = -\lambda x\\
&F^{-1} = -\frac{log(1 - F)}{\lambda}
\end{aligned}$$

So to generate random numbers pertaining to exponential distribution we have to generate from unform distribution and apply trnasformation:

$$-\frac{log(1 - U)}{\lambda}\text{ , where } U \sim U(0, 1)$$

```
> n<-200
> set.seed(101)
> sample<- -log(1-runif(n))/1.5
> hist(sample,freq=F,col="gray",border=F,breaks=20)
> x<-seq(0,5,0.1)
> lines(x,dexp(x, rate=1.5),type="l",pch=0,col="blue")
> legend("topright", legend=c("quantile transformation","theoretical density"), col = c("gray","blue"), lty =c
```

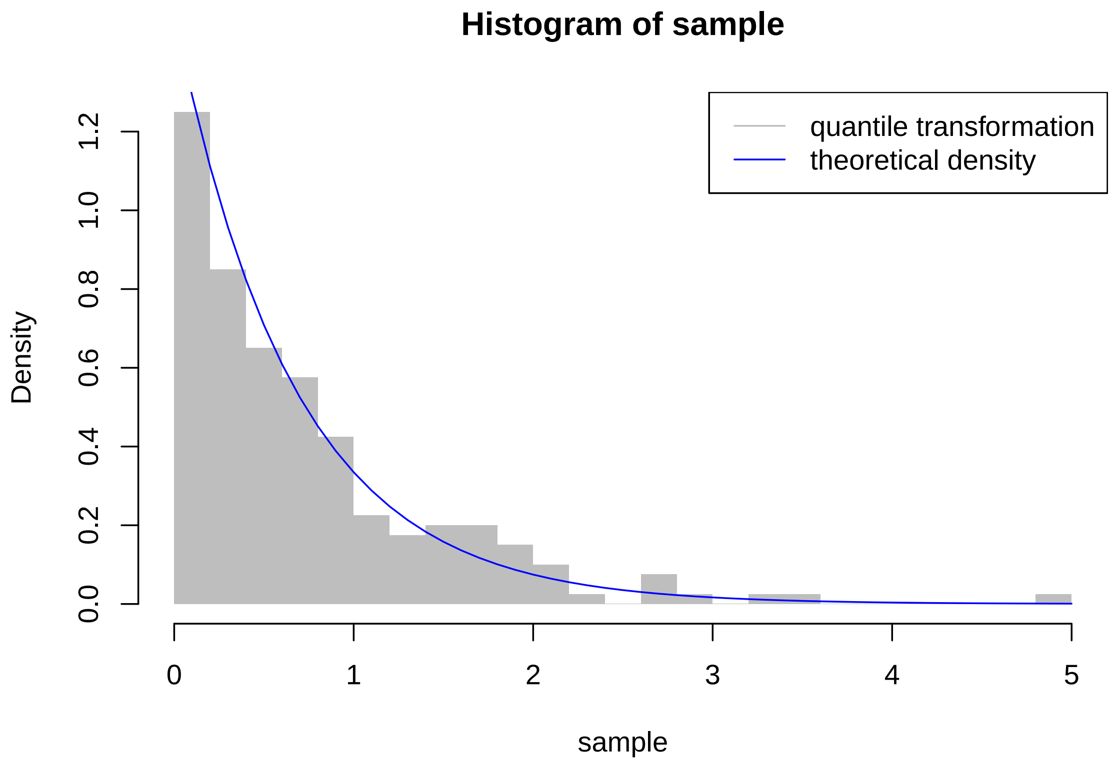

Figure 2: Generating sample using quantile transformation and comparing with theoretical density of exponential distribution