Non-life — Assignment NL1

Niels Keizer* and Robert Jan Sopers†

September 17, 2016

Generating multinormal and multi-student r.v.'s

Q1

First we run the following script:

```
> set.seed(1)
> sum(duplicated(runif(1e6))) ## = 120
[1] 120
> sum(duplicated(rnorm(1e8))) ## = 0
[1] 0
```

The function `duplicated` returns a logical array where unique numbers are marked with 0 and duplicates are marked with 1 (the first occurence of the number is marked with a 0). Summming this array thus gives the total number of duplicates. The uniform distribution gives 120 duplicates in a much smaller sample size than the normal distribution, which gives 0 duplicates.

In the assignment, the expected number of different numbers is derived to be

$$\mathbb{E}[N_n] = \frac{1-f^n}{1-f}$$

The number of duplicates is then given by $n - \mathbb{E}[N_n]$ We run the following script:

```
> m <- 2^32;n <- 1e6
> f <- 1 - 1/m
> num_dup_unif <- n - (1-f^n)/(1-f)
> num_dup_unif
[1] 116.3988
```

The expected result of 116.4 is quite close to the generated result. The outcome of 120 is therefore consistent with the assumption that `runif` produces different values uniformly.

*Student number: 10910492
†Student number: 99999999

The resolution of `rnorm` is somewhere in the 2^{50}'s. Directly calculating $1 - f^n$ will give 1, because f is so close to 1. First, we use to approximation given in the assignment.

$$f^n = \left(1 - \frac{1}{m}\right)^n \approx 1 - \frac{n}{m} + \frac{n^2}{2m^2}$$

Inserting this into our equation for the number of different numbers gives

$$\frac{1-f^n}{1-f} \approx \frac{1 - 1 + \frac{n}{m} - \frac{n^2}{2m^2}}{1 - \left(1 - \frac{1}{m}\right)} = \frac{\frac{n}{m} - \frac{n^2}{2m^2}}{\frac{1}{m}} = n - \frac{n^2}{2m}$$

This results in the following equation for the expected number of duplicates.

$$n - \left(n - \frac{n^2}{2m}\right) = \frac{n^2}{2m}$$

Next we check in `R` if the number of duplicates is consistent with values for m of 10^{15}, 10^{16}, 10^{17} or 10^{18}, when n is 10^8

```
> n_norm <- 1e8
> m_norm <- c(1e15,1e16,1e17,1e18)
> num_dup_norm <- n_norm^2/(2*m_norm)
> num_dup_norm
[1] 5.000 0.500 0.050 0.005
```

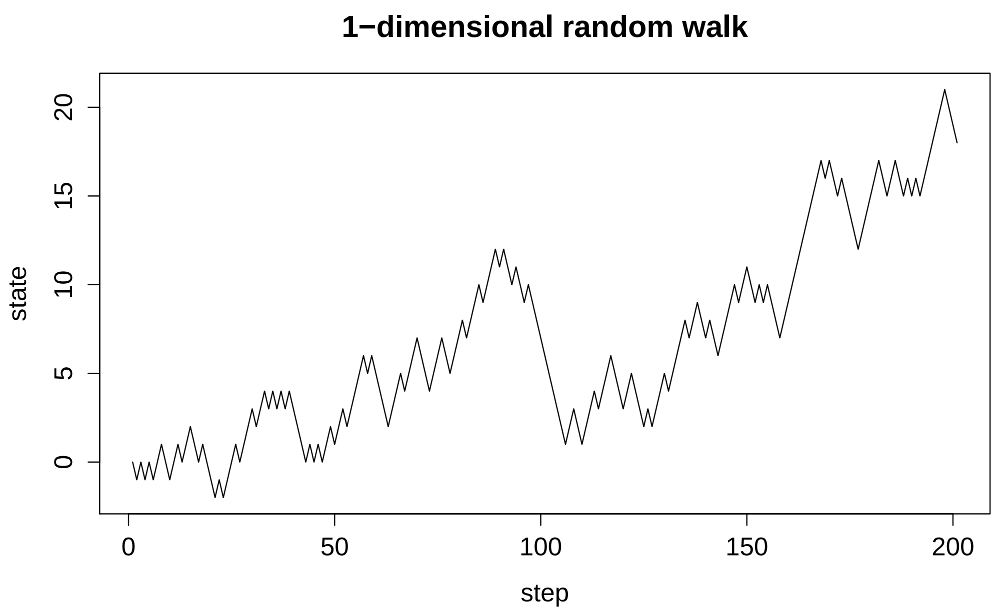
The obtained result seems to be consistent with a resolution of 10^{16} or higher.

Q2

The following code is executed in `R`.

```
> n <- 200; p <- 0.52
> x <- c(0,cumsum(2*rbinom(n,1,p)-1))
> plot(x, type="l", lwd=1, ylab="state", xlab="step", main="1-dimensional random walk")
```

This gives the following biased random walk:



Q3

Given that $X, Y \sim N(0,1)$, we want to transform (X,Y) into (X,Y^*), with $Y^* = aX + bY$, with a, b chosen in such a way that $\mathrm{Var}[Y^*] = 1$ and $r(X,Y^*) = 0.8$.

$$\begin{aligned} r(X,Y^*) &= \frac{\mathbb{E}[XY^*] - \mathbb{E}[X]\,\mathbb{E}[Y*]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y^*]}} = \frac{\mathbb{E}[aX^2 + bXY] - 0 \cdot \mathbb{E}[Y^*]}{\sqrt{1*1}} \\ &= a * \mathbb{E}[X^2] + b * \mathbb{E}[XY] = a(\mathrm{Var}[X] - \mathbb{E}[X]^2) = a \end{aligned}$$

Here we use that $\mathbb{E}[X] = 0$, $\mathrm{Var}[X] = \mathrm{Var}[Y] = \mathrm{Var}[Y^*] = 1$. Also $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[]Y] = 0$, because X and Y are independent. Next we use the condition that the variance of Y^* must also be 1.

$$\mathrm{Var}[Y^*] = \mathrm{Var}[aX + bY] = a^2\,\mathrm{Var}[X] + b^2\,\mathrm{Var}[Y] + 2ab\,\mathrm{Cov}[X,Y] = a^2 + b^2 = 1$$

We can conclude from this that $a = 0.8$ and $b = \sqrt{1 - a^2} = 0.6$.

In `R` code:

```
set.seed(2004); options(digits=2)
X <- rnorm(1000); Y <- rnorm(1000)
a <- .8; b <- sqrt(1 - a^2); Y <- a*X + b*Y
```

Q4

The variance-covariance matrix Σ of the random vector (X,Y^*) is equal to the correlation matrix because $\mathrm{Var}[X] = \mathrm{Var}[Y^*] = 1$. It is given by the following expression.

$$\Sigma = \begin{pmatrix} \mathrm{Cov}[X,X] & \mathrm{Cov}[X,Y^*] \\ \mathrm{Cov}[Y^*,X] & \mathrm{Cov}[Y^*,Y^*] \end{pmatrix} = \begin{pmatrix} \mathrm{Var}[X] & r(X,Y^*) \\ r(X,Y^*) & \mathrm{Var}[Y^*] \end{pmatrix} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$$

Q5

If A is to be the Cholesky decomposition of Σ, it should be a lower triangular matrix with real and positive entries and $AA^* = \Sigma$, where A^* is the conjugate transpose of A. Checking this gives

$$\begin{pmatrix} 1 & 0 \\ a & b \end{pmatrix} \begin{pmatrix} 1 & a \\ 0 & b \end{pmatrix} = \begin{pmatrix} 1 & a \\ a & a^2 + b^2 \end{pmatrix} = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix} = \Sigma$$

Q6

We execute the following `R` code.

```
> c(mean(X), var(X), mean(Y), var(Y), cor(X,Y))
[1] 0.051 0.983 0.070 0.994 0.796
```

The means of X and Y^* are close to 0. The variances close to 1 and the correlation is close to 0.8. This resembles the theoretical values quite close.

Q7

Let (X, Y) be bivariate Normal with $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, $\mathrm{Var}[X] = \mathrm{Var}[Y] = 1$ and $r(X, Y) = r$. W is independent of (X, Y). Then

$$\begin{aligned}
r(XW, YW) &= \frac{\mathbb{E}[XWYW] - \mathbb{E}[XW]\,\mathbb{E}[YW]}{\sqrt{\mathrm{Var}[XW]\,\mathrm{Var}[YW]}} \\
&= \frac{\mathbb{E}[W^2]\,\mathbb{E}[XY] - \mathbb{E}[W]^2\,\mathbb{E}[X]\,\mathbb{E}[Y]}{\sqrt{(\mathbb{E}[W^2]\,\mathbb{E}[X^2] - \mathbb{E}[W]^2\,\mathbb{E}[X]^2)\,(\mathbb{E}[W^2]\,\mathbb{E}[Y^2] - \mathbb{E}[W]^2\,\mathbb{E}[Y]^2)}} \\
&= \frac{\mathbb{E}[W]^2\,\mathbb{E}[XY] - 0}{\sqrt{(\mathbb{E}[W^2]\,\mathbb{E}[X^2] - 0)\,(\mathbb{E}[W^2]\,\mathbb{E}[Y^2] - 0)}} \\
&= \frac{\mathbb{E}[W^2]}{\sqrt{\mathbb{E}[W^2]^2}}\frac{\mathbb{E}[XY]}{\mathrm{Var}[X]\,\mathrm{Var}[Y]} = r \cdot \frac{\mathbb{E}[W^2]}{\sqrt{\mathbb{E}[W^2]^2}} = r
\end{aligned}$$

When $\mathbb{E}[W^2]$ is finite, the final step in the derivation is allowed. This is equivalent with demanding $\mathrm{Var}[W]$ to be finite.

Q8

Next, we execute the following `R` code.

```
> chi5 <- sqrt(rchisq(1000, df=5)/5)
> X <- X/chi5; Y <- Y/chi5
> c(mean(X), var(X), mean(Y), var(Y), cor(X,Y))
[1] 0.038 1.525 0.068 1.528 0.786
```

We take X and $Y*$ as defined earlier. $V \sim \chi^2_k$ and $W = \sqrt{k/V}$ with $k = 5$. The population mean of XW is 0 because $\mathbb{E}[XW] = \mathbb{E}[X]\,\mathbb{E}[W] = 0$. The same holds for Y^*W. $r(XW, Y^*W) = 0.8$ has been proven in question 7. Next we determine the variance of XW.

$$\begin{aligned}
\mathrm{Var}[XW] &= \mathbb{E}[X]^2\,\mathrm{Var}[W] + \mathrm{Var}[X]\,\mathbb{E}[W]^2 + \mathrm{Var}[X]\,\mathrm{Var}[W] \\
&= 0 + \mathrm{Var}[X](\mathbb{E}[W]^2 + \mathrm{Var}[W]) = \mathrm{Var}[X]\,\mathbb{E}[W^2] \\
&= 1 \cdot \mathbb{E}[k/V] = k\,\mathbb{E}[1/V] = \frac{k}{k-2}
\end{aligned}$$

Here we use that $\mathbb{E}[1/V] = 1/(k-2)$, which we will derive below, $f_{\chi^2}(x; k)$ is the probability density function of the chi-squared distribution with k degrees of freedom.

$$\begin{aligned}
\mathbb{E}[1/V] &= \int_0^\infty \frac{1}{x} f_{\chi^2}(x;k)dx = \int_0^\infty \frac{1}{x}\frac{x^{(k/2-1)}e^{-x/2}}{2^{k/2}\Gamma(k/2)}dx = \int_0^\infty \frac{x^{(k/2-2)}e^{-x/2}}{2^{k/2}\Gamma(k/2)}dx \\
&= \int_0^\infty \frac{x^{((k-2)/2-1)}e^{-x/2}}{2^{k/2}\Gamma(k/2)}dx = \int_0^\infty \frac{x^{((k-2)/2-1)}e^{-x/2}}{2 \cdot 2^{(k-2)/2}\frac{k-2}{2}\Gamma(\frac{k-2}{2})}dx \\
&= \frac{1}{k-2}\int_0^\infty f_{\chi^2}(x; k-2)dx = \frac{1}{k-2}
\end{aligned}$$

In the derivation, we use $\Gamma(k/2) = \Gamma(\frac{k-2}{2} + 1) = \frac{k-2}{2}\Gamma(\frac{k-2}{2})$.