Distribution of max, min and ranges for a sequence of uniform rv's

João Neto

October, 2014

Refs:

- (http://observations.rene-grothmann.de/distribution-of-minima-and-maxima-and-spreads/)http://observations.rene-grothmann.de/distribution-of-minima-and-maxima-and-spreads/ (http://observations.rene-grothmann.de/distribution-of-minima-and-maxima-and-spreads/)
- (http://www.johndcook.com/blog/2014/10/24/sample-range/)http://www.johndcook.com/blog/2014/10/24/sample-range/ (http://www.johndcook.com/blog/2014/10/24/sample-range/)

Say we have n iid uniform rvs

$$X_i \sim U(0,1), i = 1 \ldots n$$

The cdf of their minimum $Y = \min(X_1, \ldots, X_n)$ is:

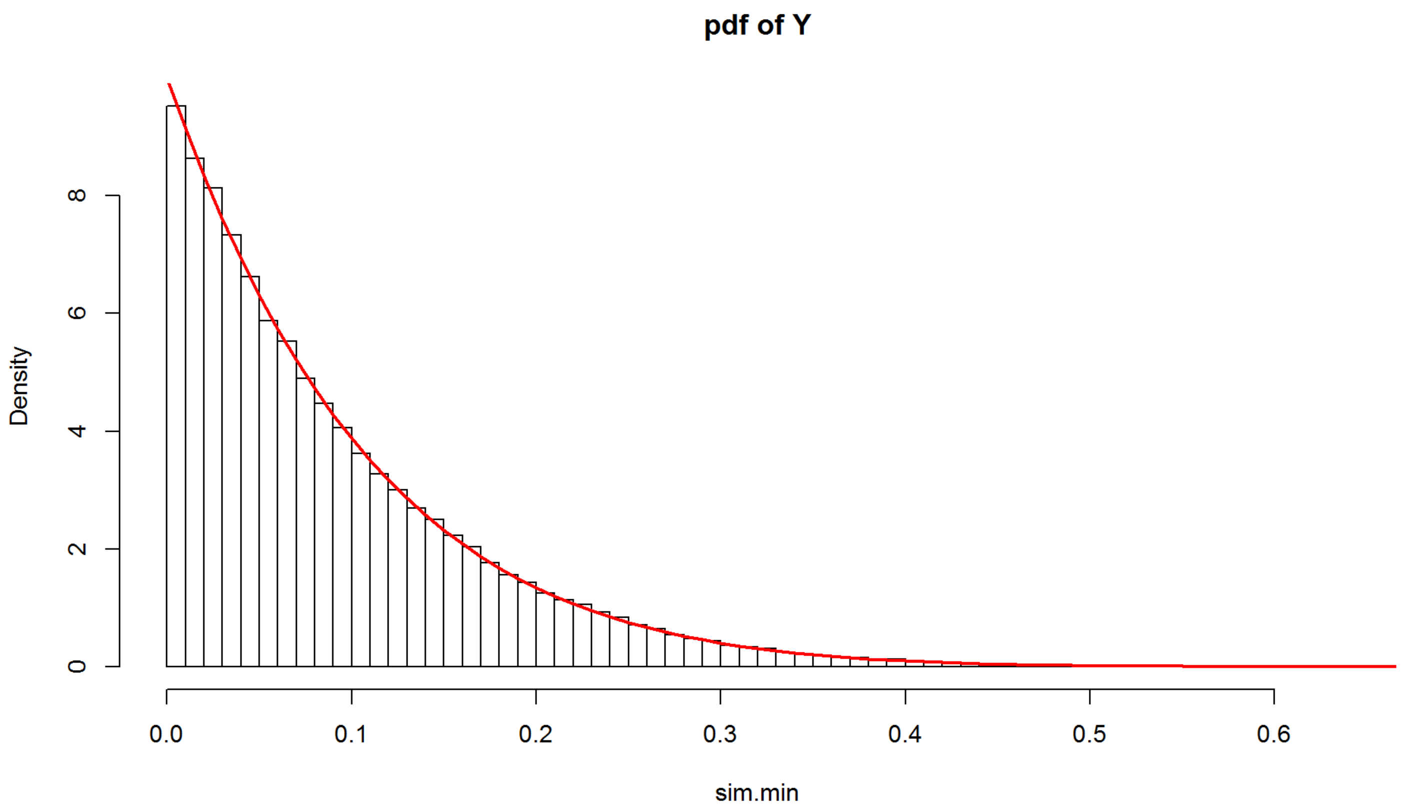
$$\begin{array}{rcll} p(Y \leq x) & = & 1 - p(Y \geq x) & \\ & = & 1 - \prod_{i=1}^n p(X_i \geq x) & Y = \min(X_i) \\ & = & 1 - p(X \geq x)^n & X_i \text{ iid } X \sim U(0,1) \\ & = & 1 - (1 - p(X \leq x))^n & \\ & = & 1 - (1-x)^n & p(X \leq x) = x \end{array}$$

Thus the pdf for Y is

$$f_Y(x) = \frac{d}{dx} P(Y \leq x) = n(1-x)^{n-1}$$

We can make a simulation to confirm this result:

```
n <- 10

pdf.min <- function(x) {    # pdf function for the minimum
  n*(1-x)^(n-1)
}

sample.min <-  function() { # miminum of sample with n U(0,1) rvs
  min(runif(n))
}

sim.min <- replicate(1e5, sample.min()) # simulation

hist(sim.min, breaks=50, prob=T, main="pdf of Y")
curve(pdf.min, 0, 1, col="red", lwd=2, add=T)
```



The maximum $Z = \max(X_1, \ldots, X_n)$ has similar development:

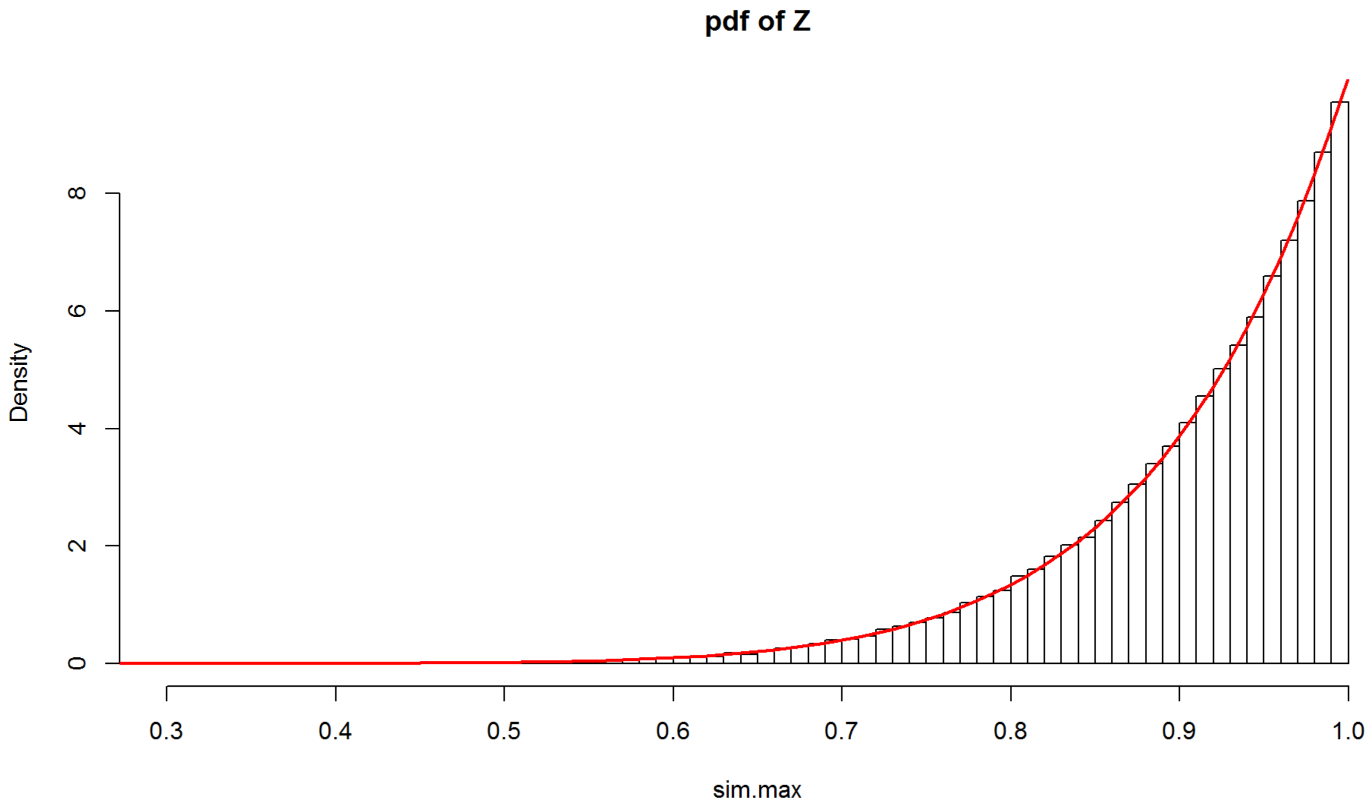
$$\begin{aligned} p(Z \leq x) &= \prod_{i=1}^n p(X_i \geq x) \\ &= x^n \qquad\qquad p(X \leq x) = x \end{aligned}$$

so, the pdf of Z is

$$f_Z(x) = nx^{n-1}$$

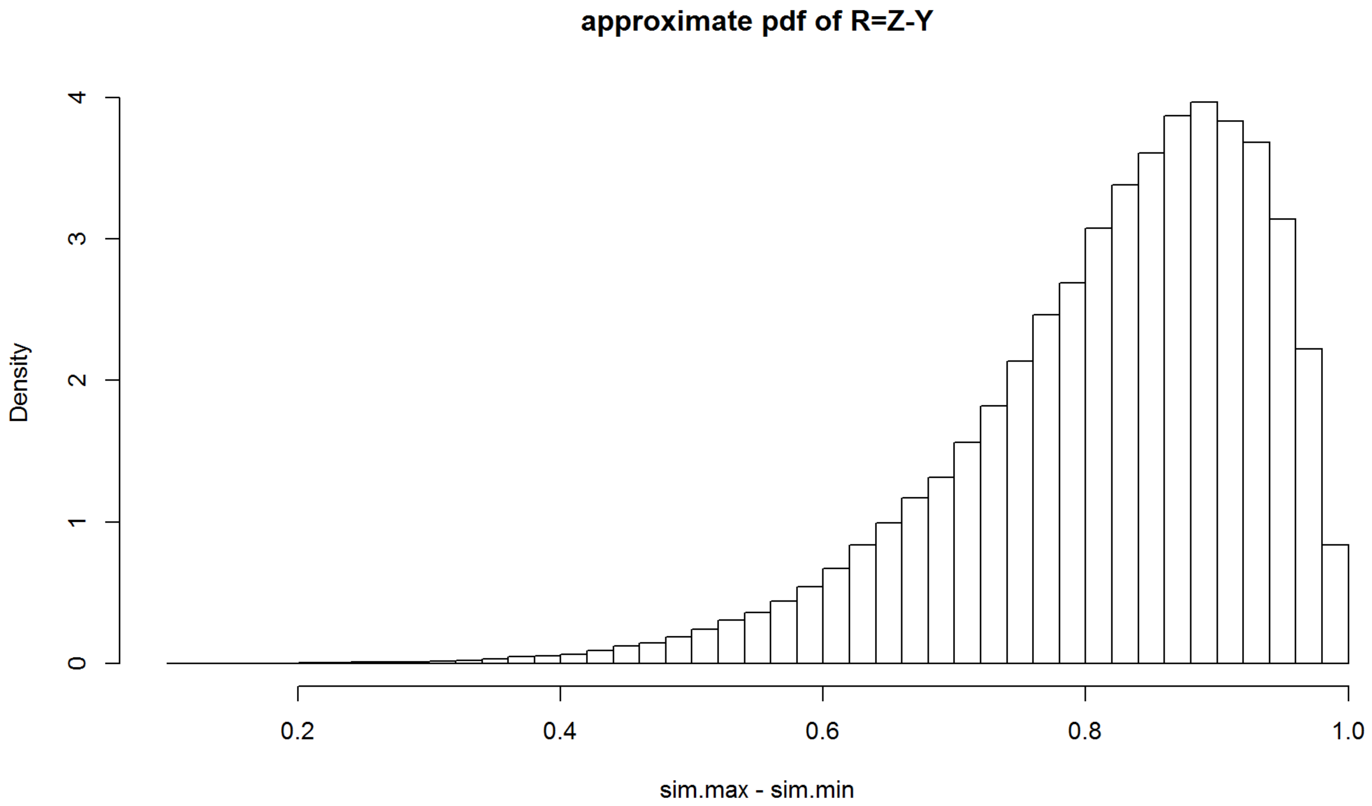
Again:

```
n <- 10

pdf.max <- function(x) {    # pdf function for the minimum
  n*x^(n-1)
}

sample.max <-  function() { # miminum of sample with n U(0,1) rvs
  max(runif(n))
}

sim.max <- replicate(1e5, sample.max()) # simulation

hist(sim.max, breaks=50, prob=T, main="pdf of Z")
curve(pdf.max, 0, 1, col="red", lwd=2, add=T)
```



The distribution of the range $R = Z - Y$ of these n values should be something like this:

```
hist(sim.max-sim.min, breaks=50, prob=T, main="approximate pdf of R=Z-Y")
```



which resembles a beta distribution. But is it? Notice that the true pdf for R is not the difference $Z - Y$ because they are not independent. To compute R's cdf we assume that x is the minimum value and the range is d.

There are two mutually exclusive events:

- $x < 1 - d$ so that we have a range $[x, x+d]$. This means two events happening, the minimum $Y = x$ and all the remaining $n-1$ points are within the interval which has length $d/(1-x)$, let's call this event W.
- $x > 1 - d$ so that we have range $[x, 1]$, ie, the minimum $Y \geq 1 - d$, ie, all n points are within a range d.

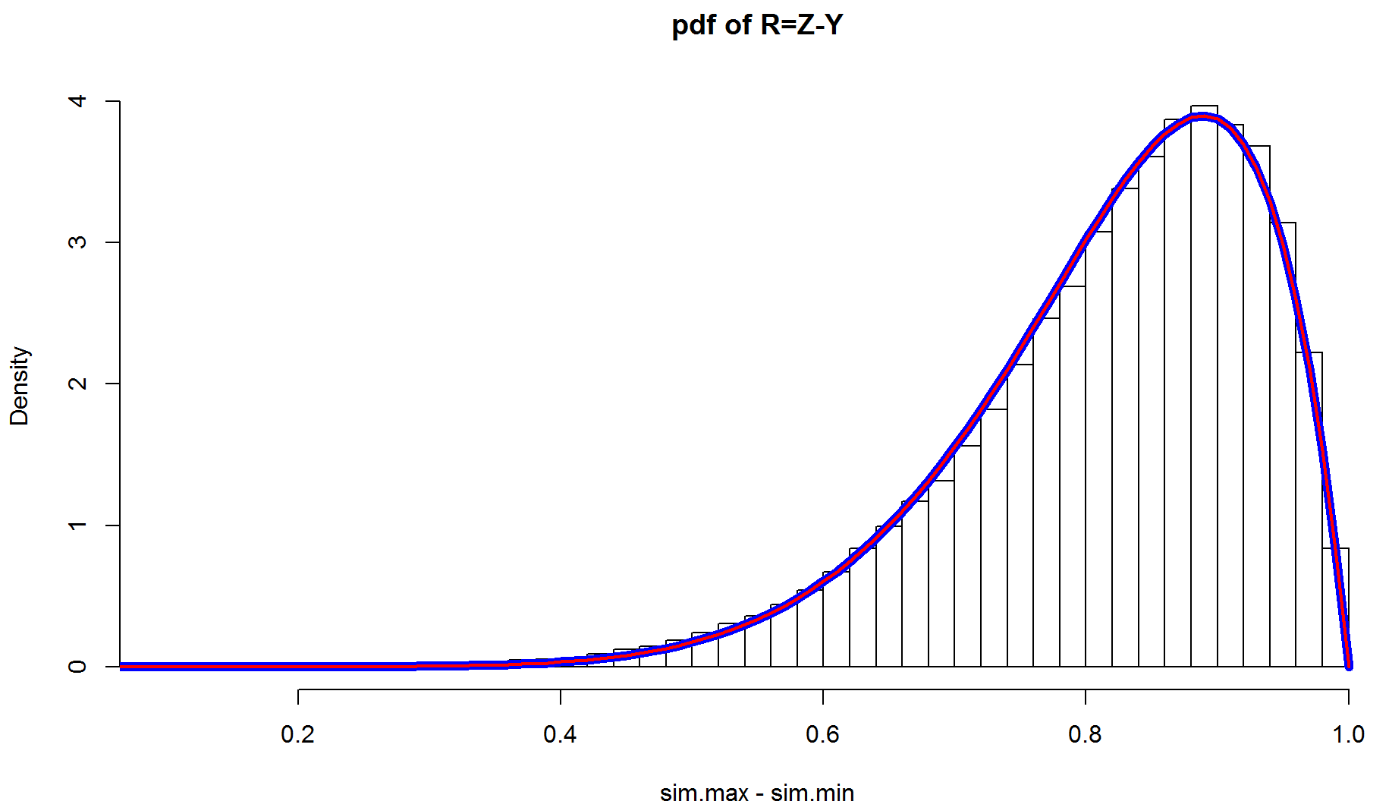
$$
\begin{aligned}
p(R \leq d) &= \int_0^{1-d} f_Y(x) p(W) dx + p(Y \geq 1-d) \\
&= \int_0^{1-d} n(1-x)^{n-1} \left(\frac{d}{1-x}\right)^{n-1} dx + d^n \\
&= \int_0^{1-d} n d^{n-1} dx + d^n \\
&= n d^{n-1}(1-d) + d^n
\end{aligned}
$$

To find the pdf:

$$f_R(x) = \frac{d}{dx} nx^{n-1}(1-x) + x^n = (1-x)x^{n-2}(n-1)n$$

We see that $R \sim \text{Beta}(n-1, 2)$

```
pdf.range <- function(x) {
  (1-x)*x^(n-2)*(n-1)*n
}

pdf.beta <- function(x) dbeta(x,n-1,2)

hist(sim.max-sim.min, breaks=50, prob=T, main="pdf of R=Z-Y")
curve(pdf.range, 0, 1, col="blue", lwd=6, add=T)
curve(pdf.beta,  0, 1, col="red",  lwd=2, add=T)
```

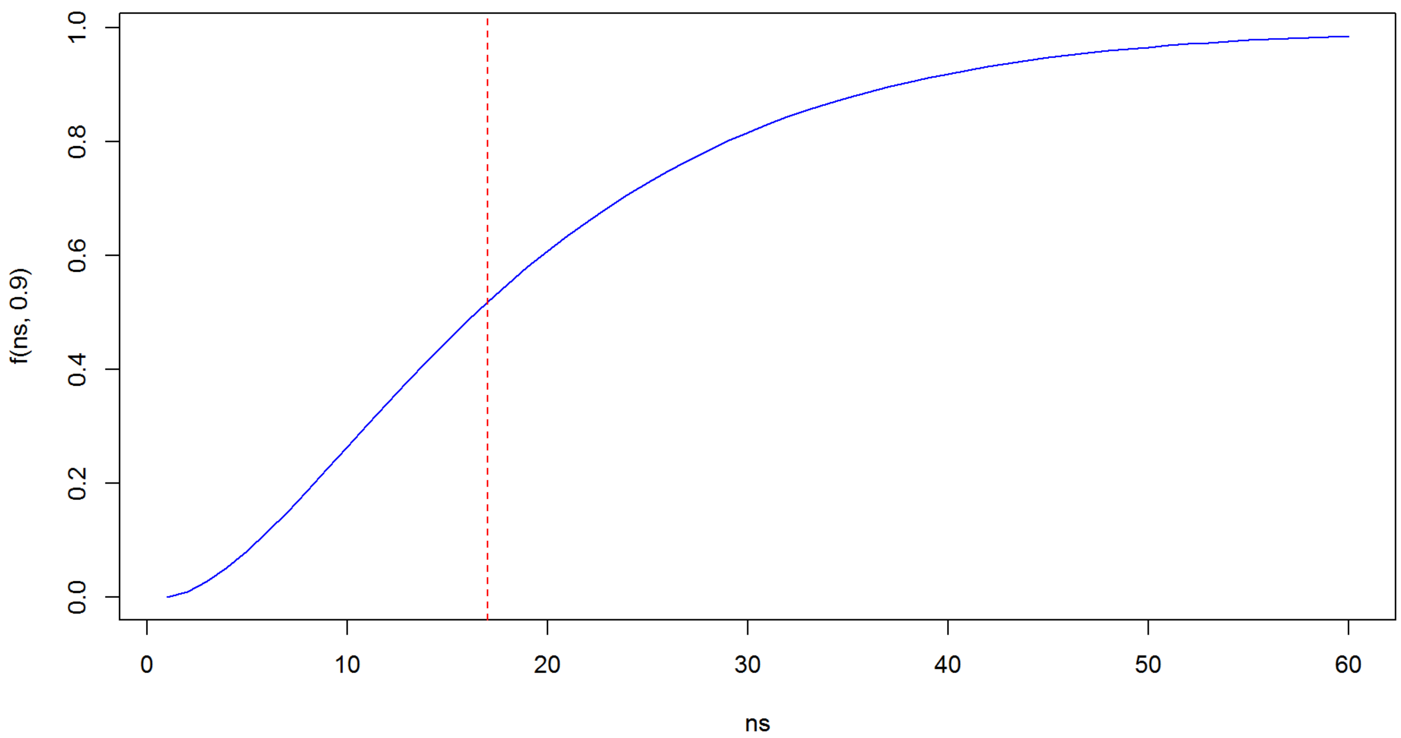


If we ask what is the probability for a sample range to be greater than a value c, we need to compute $p(R \geq c)$

$$\int_c^1 n(n-1)x^{n-2}(1-x)dx = 1 - c^{n-1}(n - c(n-1))$$

We can ask now what should the minimum n be so that the probability is greater than 0.5 for the sample range to be 90% of total range, ie, $c = 0.9$.

```
f <- function(n,c) {
  1 - c^(n-1)*(n-c*(n-1))
}

ns <- 1:60
plot(ns,f(ns,.9), type="l", col="blue")
n <- which(f(ns,.9)>0.5)[1]
abline(v=n, lty=2, col="red")
```



We need n=17 samples.