Tutorial 20/11/2014

Order Statistics: Monte Carlo study

Recall the following familiar result: consider an arbitrary absolutely continuous cdf F_X for random variable X, and the transformation to random variable U, where

$$U = F_X(X).$$

The transform is 1-1, from $\mathbb{R}$ onto $(0, 1)$. For $0 < u < 1$, we have

$$F_U(u) = \mathrm{P}[U \leq u] = \mathrm{P}[F_X(X) \leq u] = \mathrm{P}[X \leq F_X^{-1}(u)]$$

where the inverse mapping $F_X^{-1}(u)$ is a well-defined mapping from $(0, 1)$ to $\mathbb{R}$.

Order Statistics: Monte Carlo study (cont.)

It follows that

$$F_U(u) = F_X(F_X^{-1}(u)) = u \qquad 0 < u < 1$$

with the usual cdf behaviour for other values of u.

Hence

$$U \sim \text{Uniform}(0, 1).$$

Furthermore, if $X_1, \ldots, X_n$ are iid from F_X, then rvs $U_1, \ldots, U_n$ formed by taking the mapping

$$U_i = F_X(X_i) \qquad i = 1, \ldots, n.$$

are also iid Uniform$(0, 1)$ rvs.

Order Statistics: Monte Carlo study (cont.)

Finally, consider the order statistics derived from $X_1, \ldots, X_n$, that is

$$Y_i = X_{(i)} \qquad i = 1, \ldots, n.$$

Then, after transformation, we have the order statistics of $U_1, \ldots, U_n$

$$Z_i = U_{(i)} \qquad i = 1, \ldots, n.$$

However, because the function $F_X(x)$ is monotonic increasing, we must have that

$$Z_i = F_X(Y_i) \qquad i = 1, \ldots, n$$

that is, the **ranking** of the Xs is the same as the ranking of the Us.

Order Statistics: Monte Carlo study (cont.)

These results mean that we can study the properties of the order statistics from an arbitrary distribution F_X by first considering the mapping of $X_{(1)}, \ldots, X_{(n)}$ to $U_{(1)}, \ldots, U_{(n)}$ via the transform

$$U_{(i)} = F_X^{-1}(X_{(i)}) \qquad i = 1, \ldots, n.$$

then using results for the uniform, then mapping back through F_X to the X scale.

Order Statistics: Monte Carlo study (cont.)

We now consider for a fixed n the order statistics

$$U_{(n,1)}, \ldots, U_{(n,j)}, \ldots, U_{(n,n)}$$

where

- the first index n indicates the sample size
- the second index j indicates the order statistic considered, for $j = 1, \ldots, n$.

For example

$$U_{n,1} \equiv \min\{U_1, \ldots, U_n\} \qquad U_{n,n} \equiv \max\{U_1, \ldots, U_n\}$$

etc.

Order Statistics: Monte Carlo study (cont.)

We have from the results from lectures, for $0 < u < 1$,

$$F_{U_{n,1}}(u) = 1 - (1-u)^n \qquad\qquad F_{U_{n,n}}(u) = u^n$$

with standard cdf behaviour elsewhere, so that

$$f_{U_{n,1}}(u) = n(1-u)^{n-1} \qquad\qquad f_{U_{n,n}}(u) = nu^{n-1}$$

and zero elsewhere. Also, for any $1 < j < n$

$$\begin{aligned} f_{U_{n,j}}(u) &= \frac{n!}{(j-1)!(n-j)!}\{F_U(u)\}^{j-1} f_U(u)\{1-F_U(u)\}^{n-j} \\ &= \frac{n!}{(j-1)!(n-j)!} u^{j-1}(1-u)^{n-j} \\ &\equiv \text{Beta}(j, n-j+1). \end{aligned}$$

Order Statistics: Monte Carlo study (cont.)

In this case

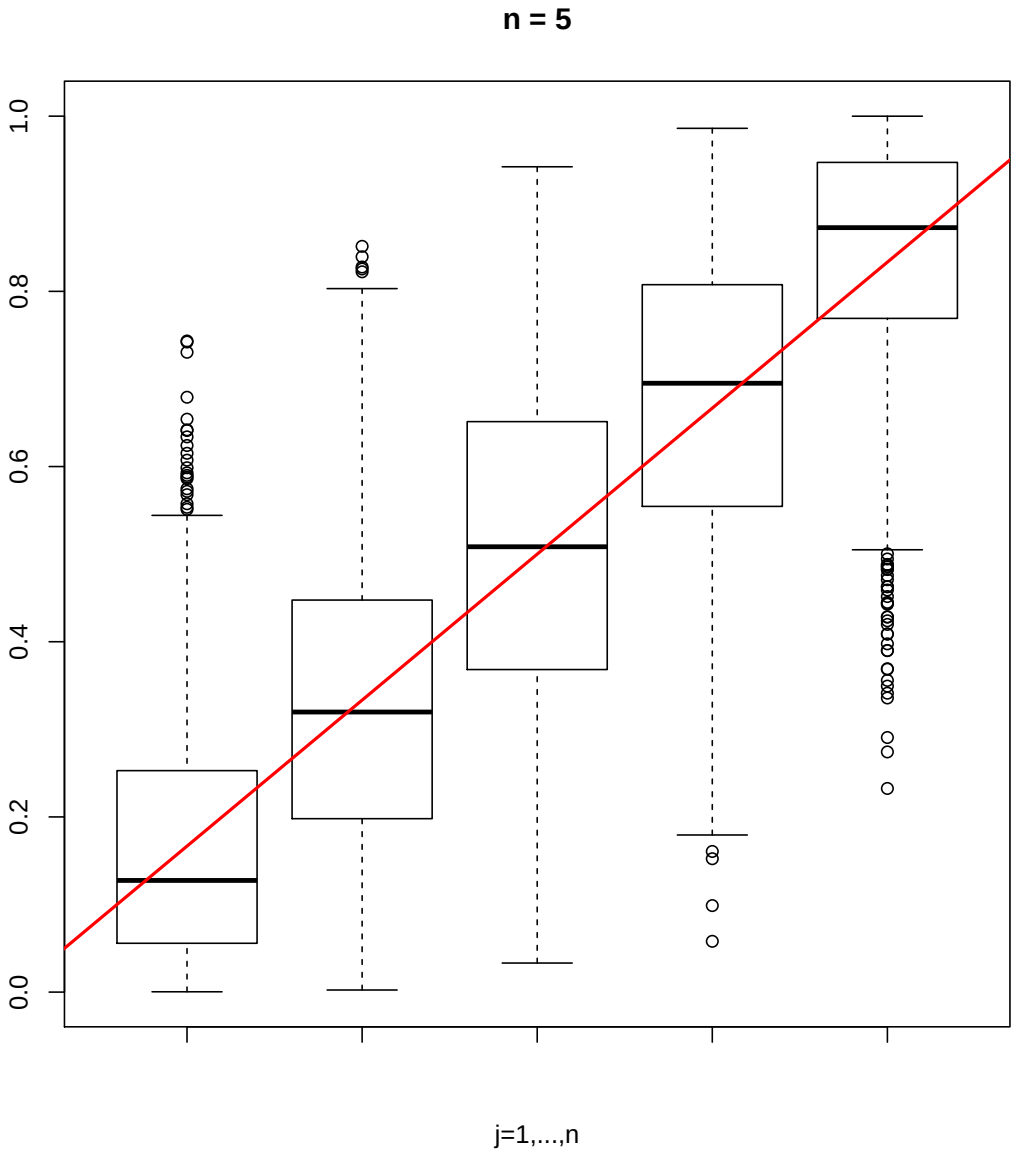
$$\mathbb{E}_{U_{n,j}}[U_{n,j}] = \frac{j}{n+1} \qquad j = 1, \ldots, n.$$

Thus, the expected value of $U_{n,j}$ increases linearly with j.

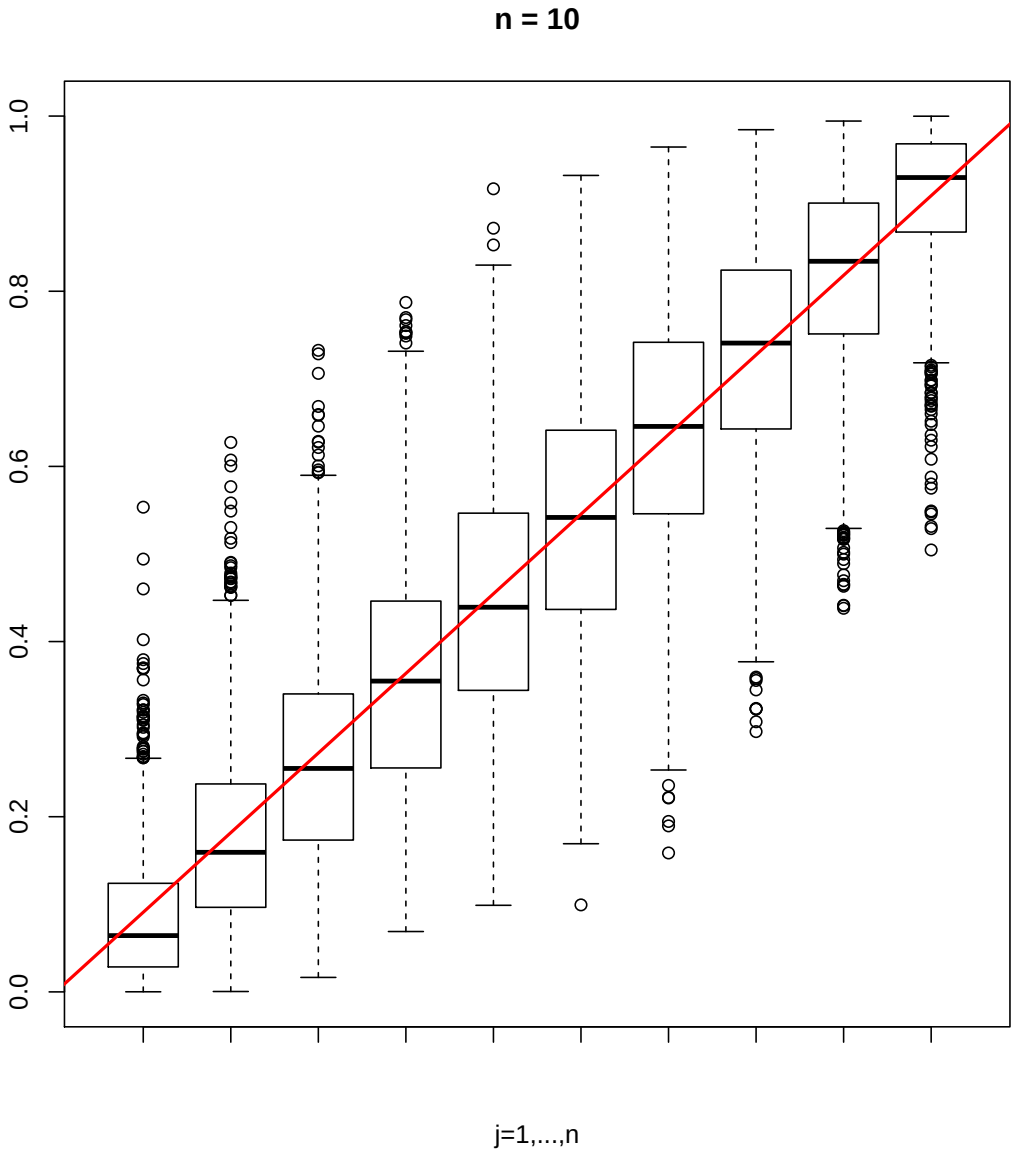
In the following Monte Carlo study, we consider the simulation of 1000 replicate data sets of different sample sizes in order to study the sampling distribution of the order statistics.

The following figures depict boxplots of $U_{(n,j)}$ for $j = 1, \ldots, n$, and $n = 5, 10, 20, 50, 100, 500$.

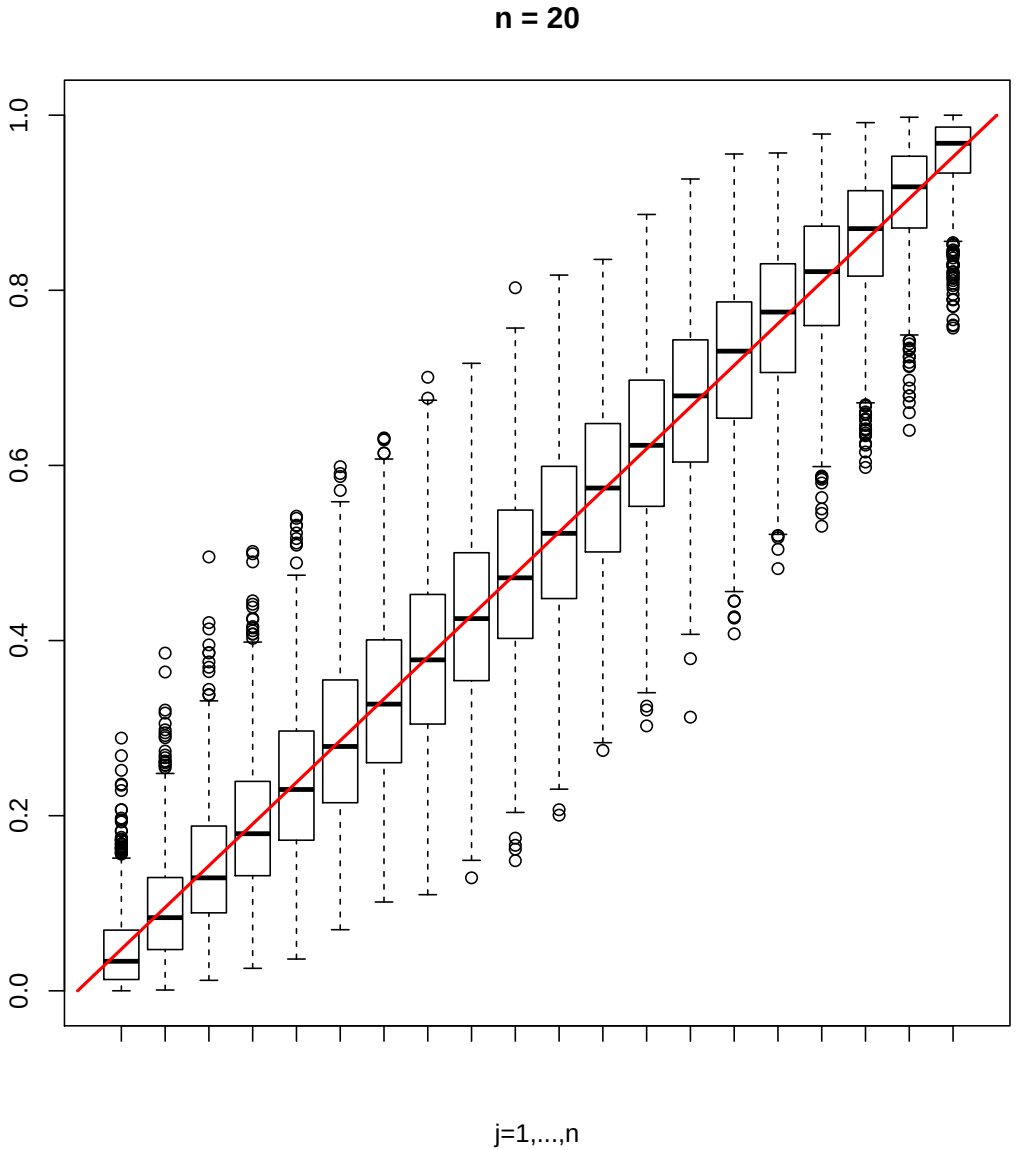
Monte Carlo study



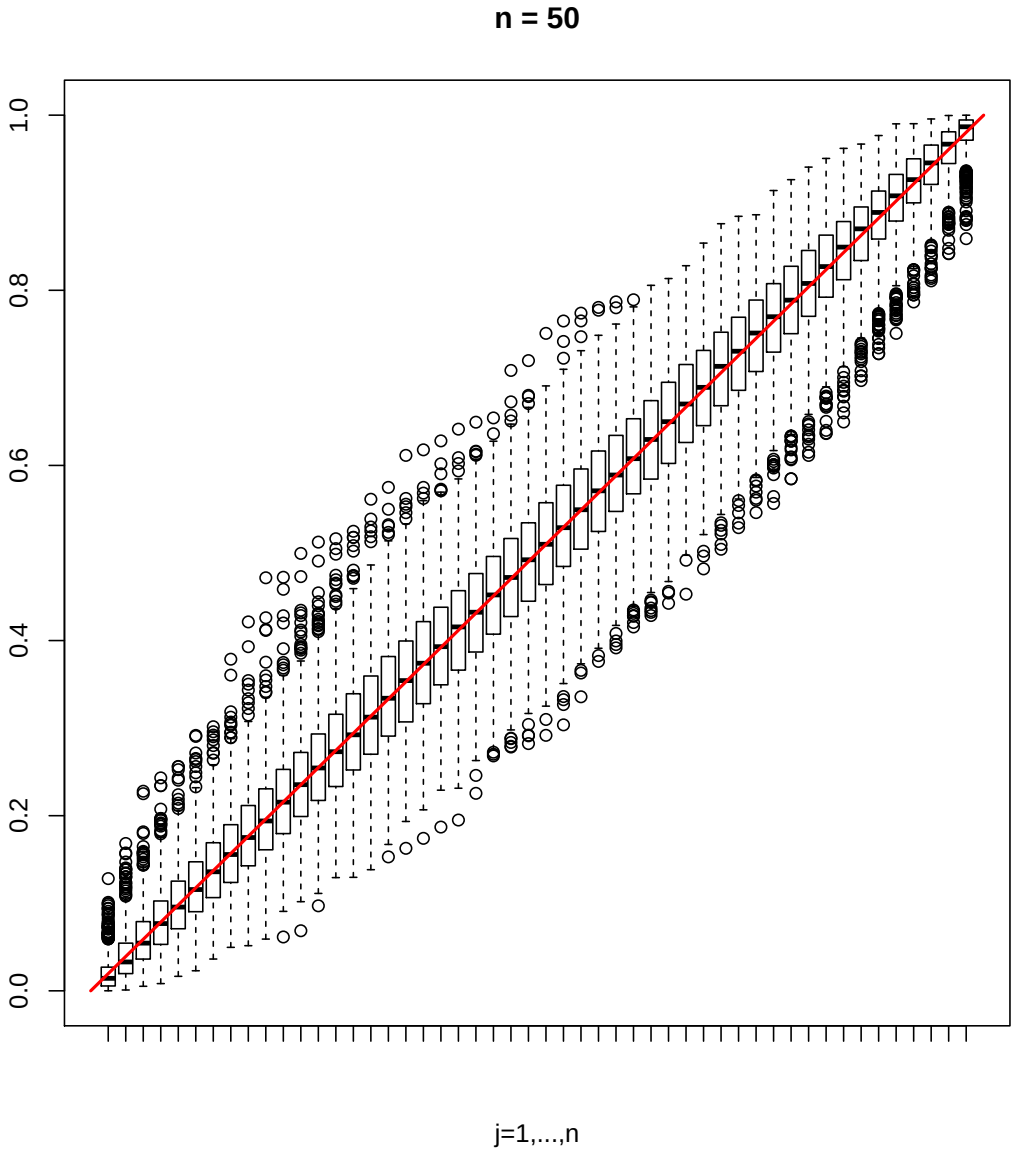
Monte Carlo study



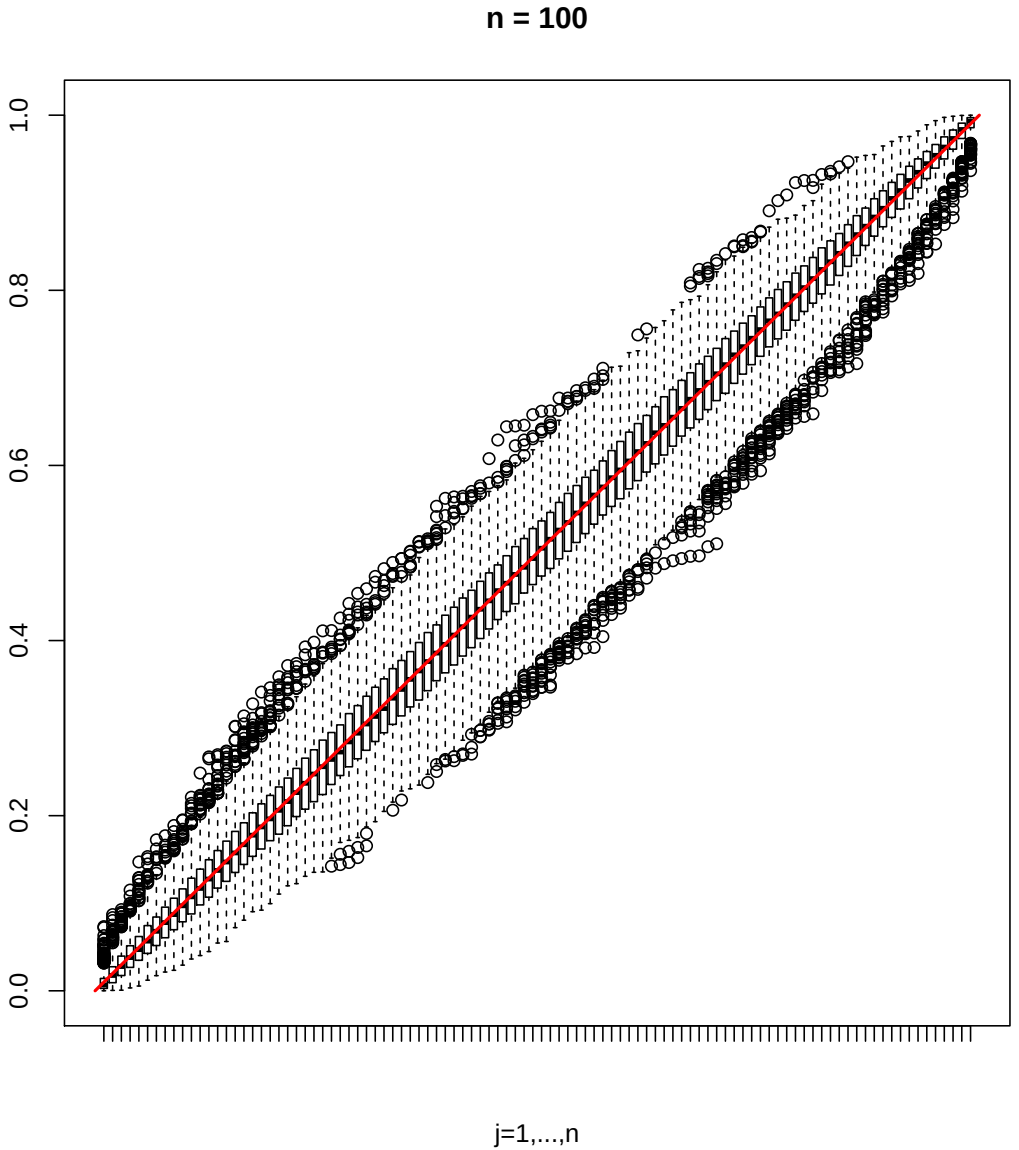
Monte Carlo study



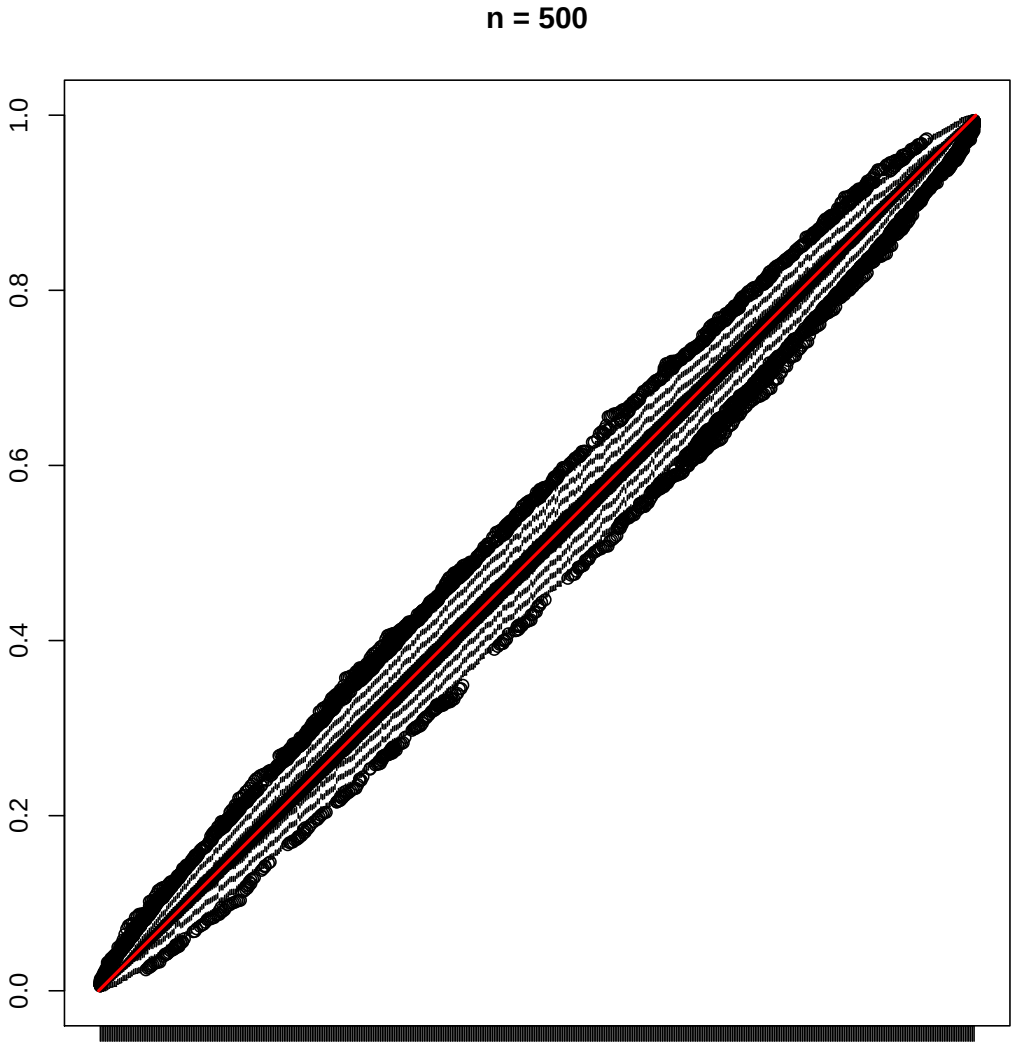
Monte Carlo study



Monte Carlo study



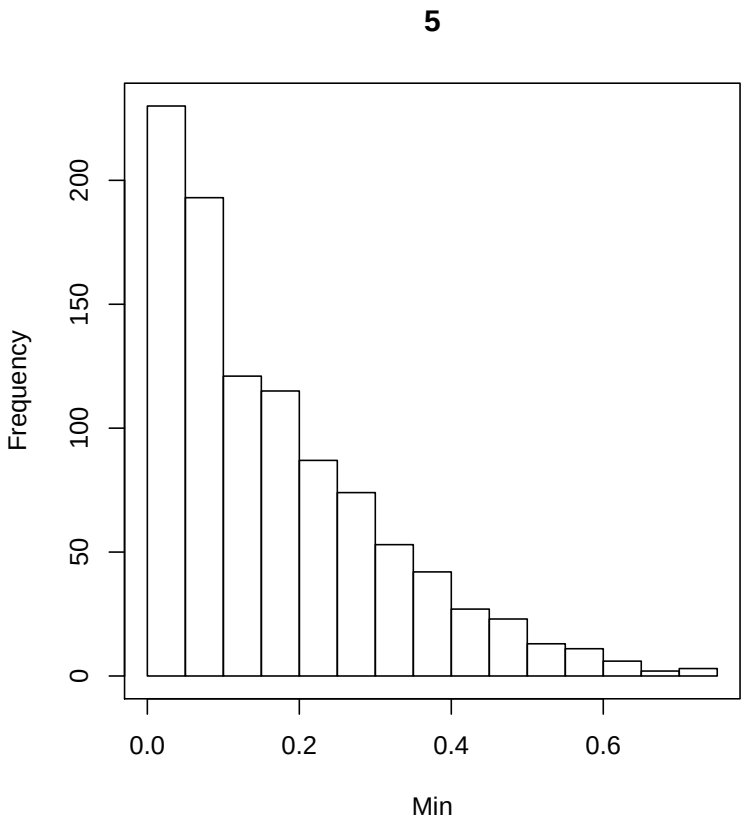
Monte Carlo study

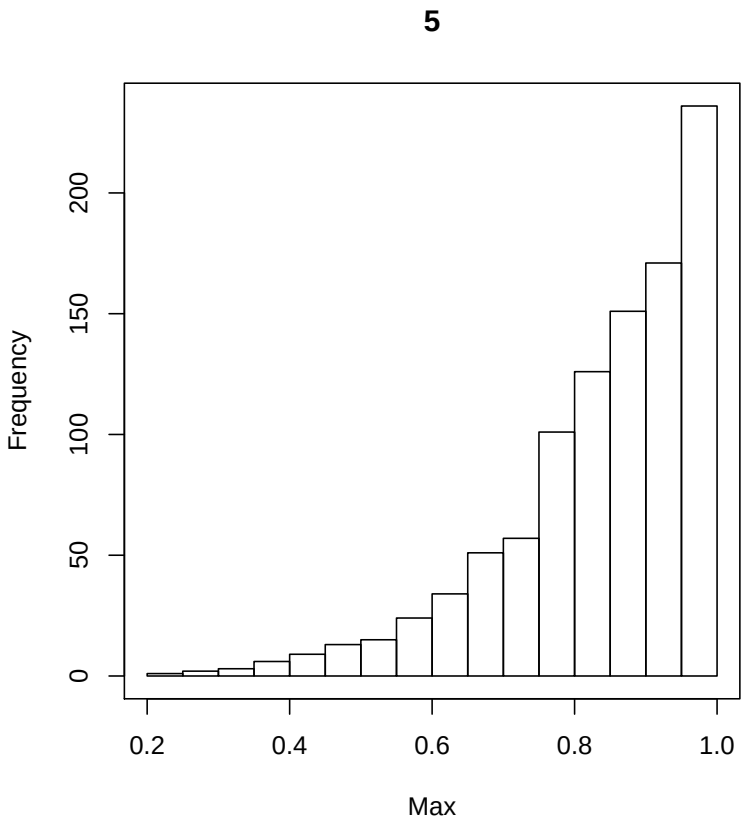


Monte Carlo study: Extreme order statistics

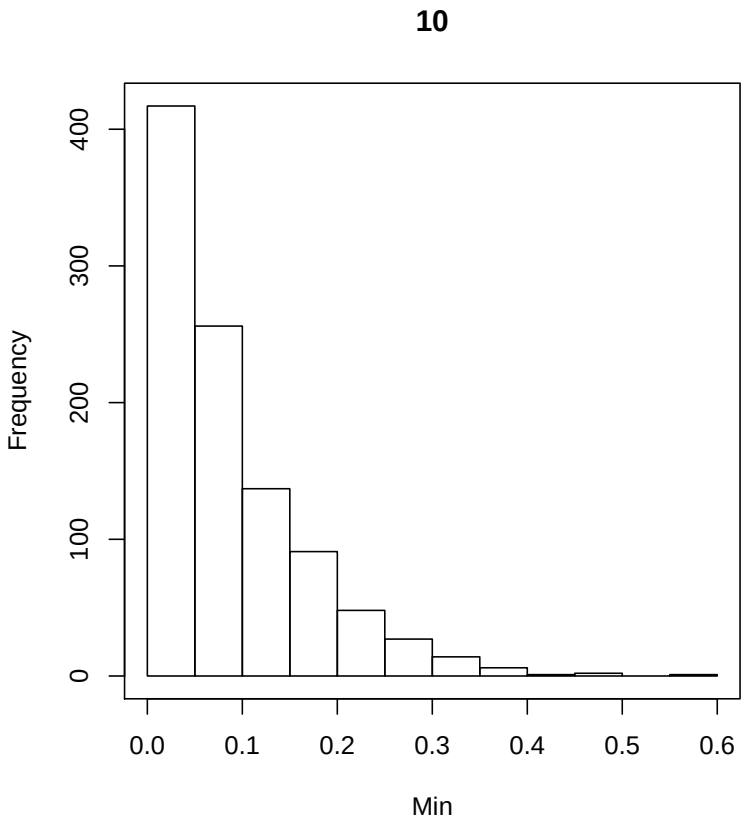
The following figures depict histograms of $U_{(n,1)}$ and $U_{(n,n)}$ for $n = 5, 10, 20, 50, 100, 500$.

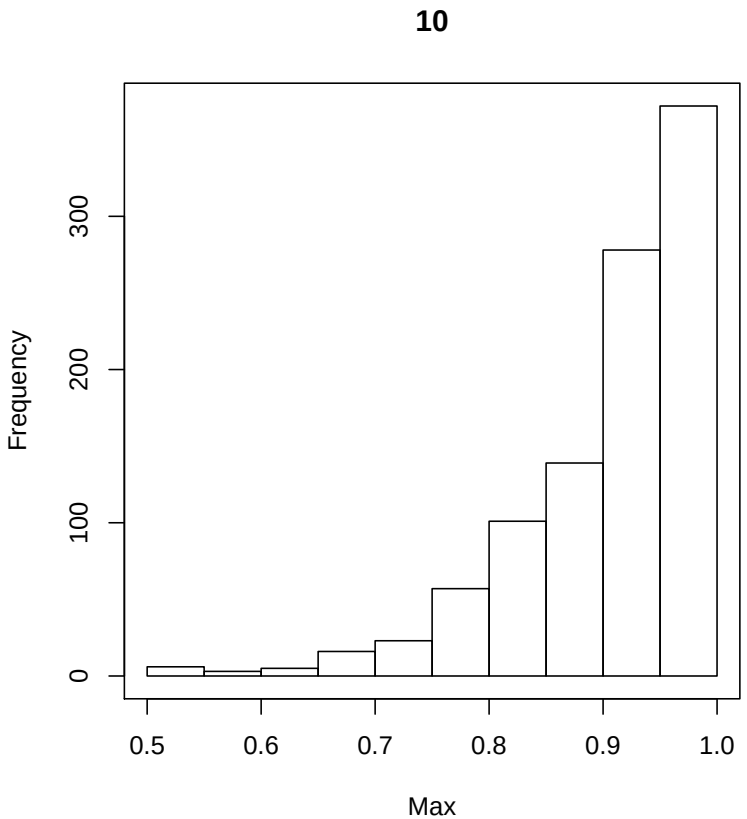
Monte Carlo study: Extreme order statistics



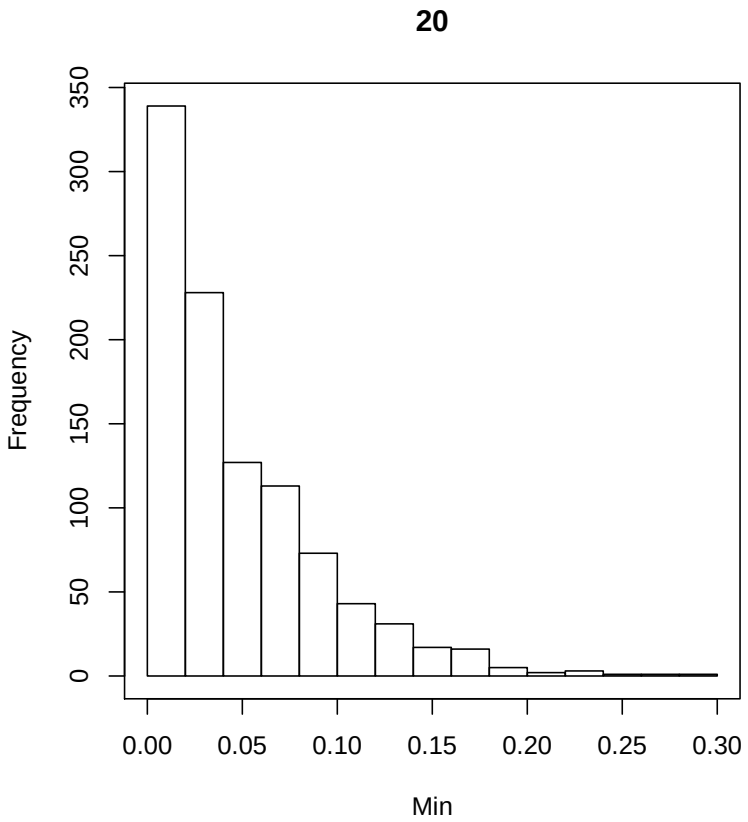


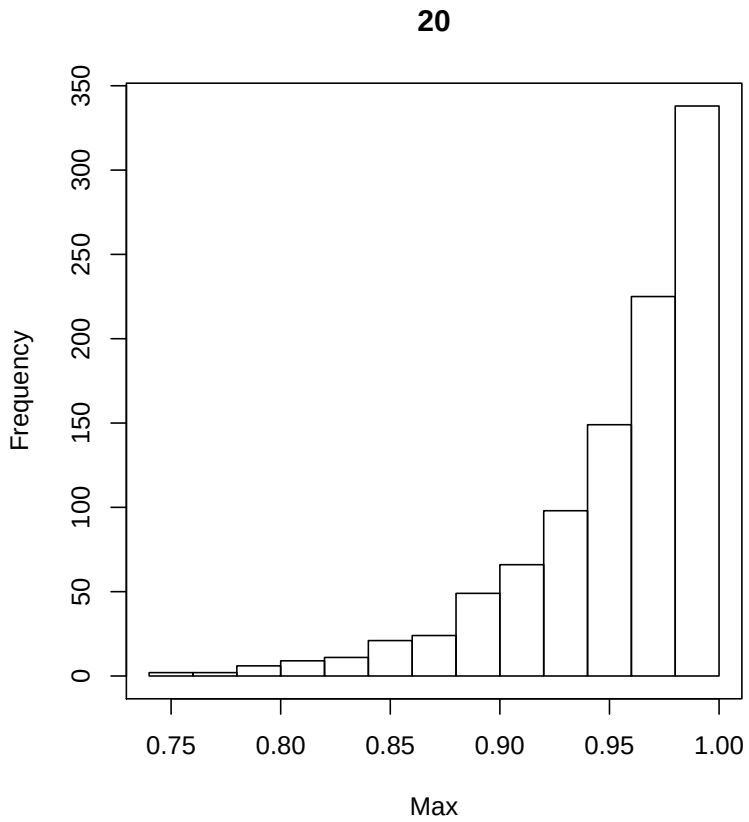
Monte Carlo study: Extreme order statistics



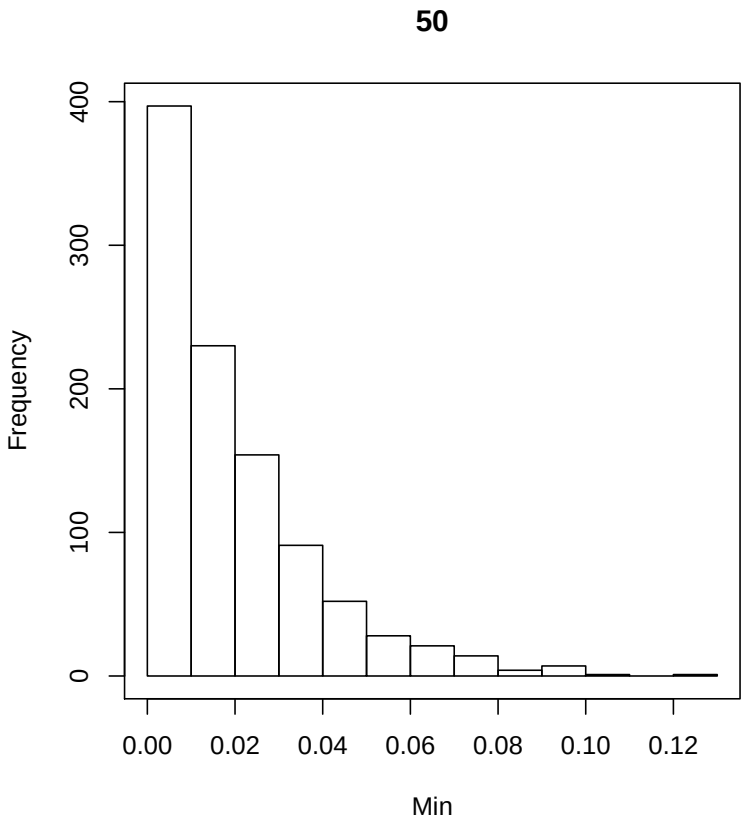


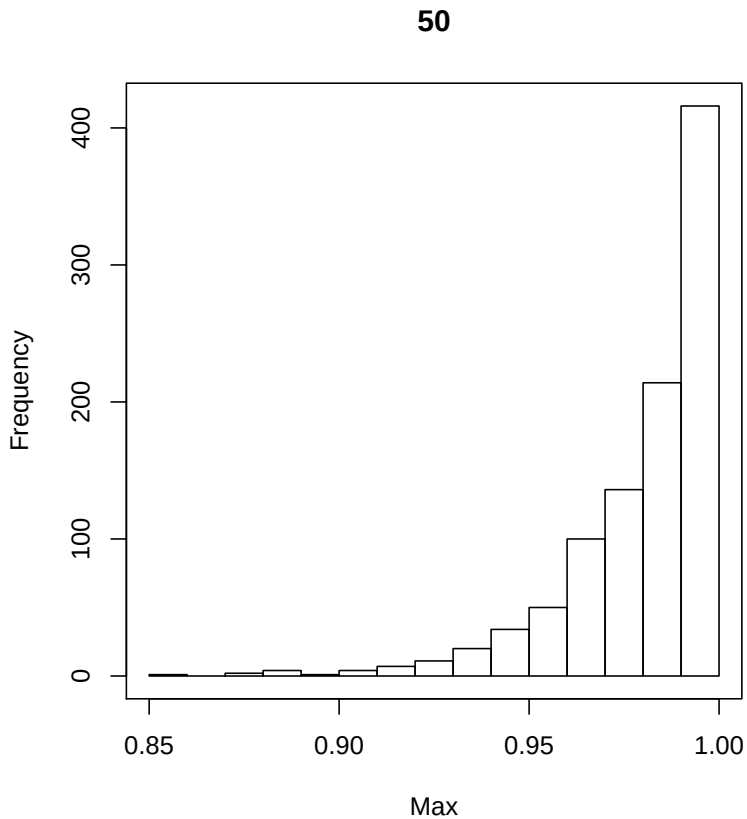
Monte Carlo study: Extreme order statistics



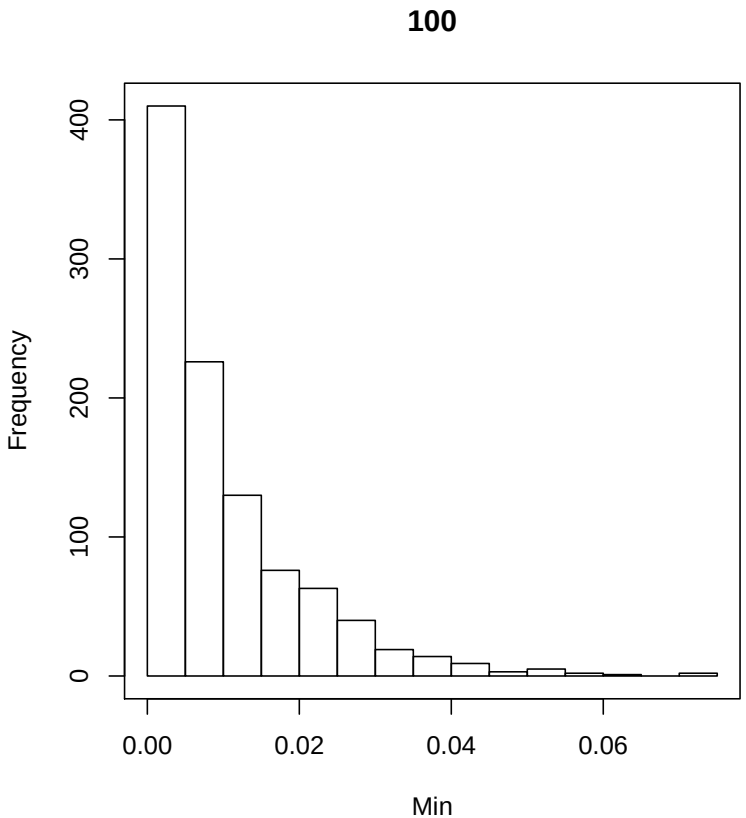


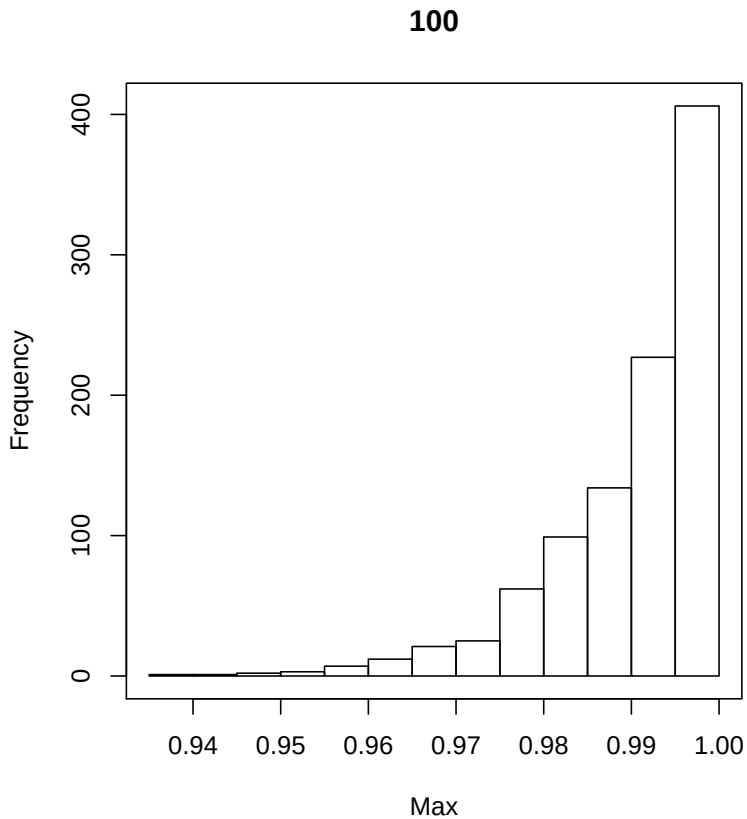
Monte Carlo study: Extreme order statistics



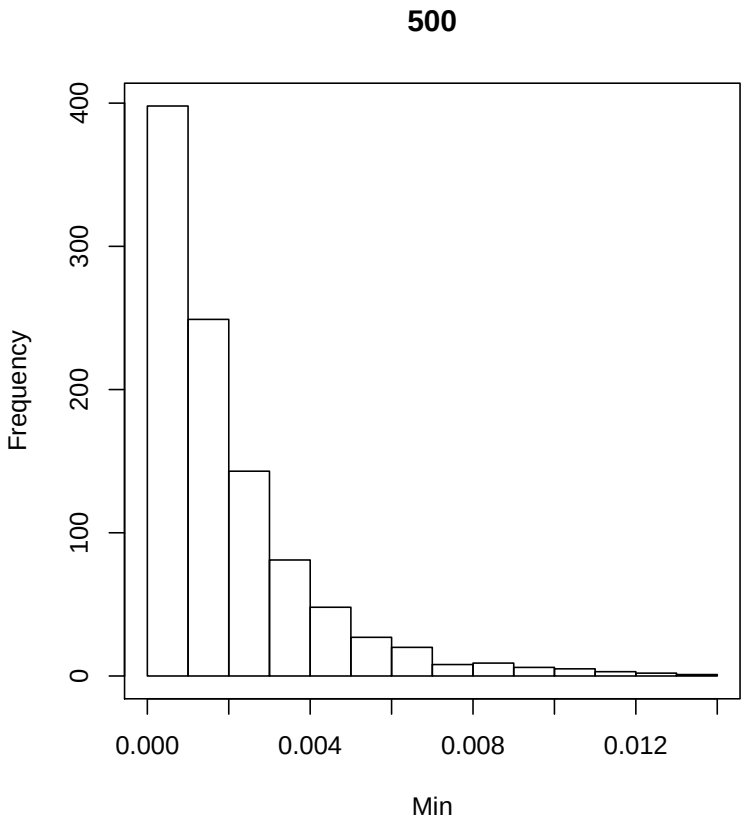


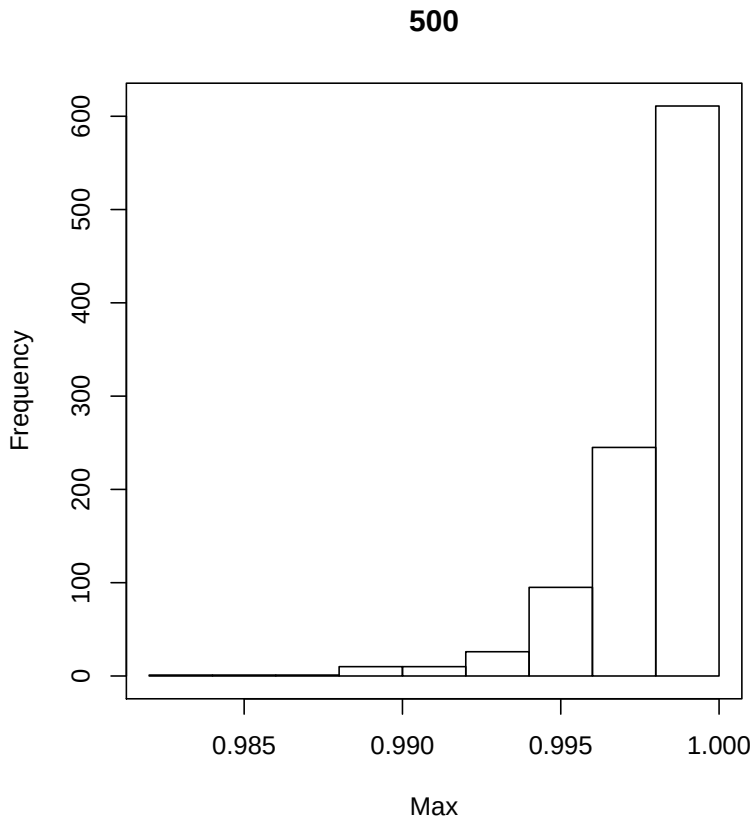
Monte Carlo study: Extreme order statistics





Monte Carlo study: Extreme order statistics





Monte Carlo study: Extreme order statistics

We see that the minimum and the maximum order statistics become increasingly concentrated near the ends of the range $(0, 1)$.

More generally, as j varies for fixed n, the variance of $U_{n,j}$ is

$$\frac{j(n-j+1)}{(n+1)^2(n+2)}$$

and is therefore a quadratic function in j, largest when

$$j \approx n/2.$$

Extreme order statistics: Limiting Distributions

We can study formally what happens to the distributions of the extreme order statistics as n increases. For any fixed u, we have

$$F_{U_{(n,1)}}(u) = \begin{cases} 0 & u < 0 \\ 1 - (1-u)^n & 0 \leq u \leq 1 \\ 1 & u > 1 \end{cases}$$

We consider what happens to this function at each u, in the limit as $n \longrightarrow \infty$.

Extreme order statistics: Limiting Distributions

We have

$$F_{U_{(n,1)}}(u) \longrightarrow \begin{cases} 0 & u \leq 0 \\ 1 & u > 1 \end{cases}$$

that is, the limiting function is a step function, with a single step at 0. The limiting function is not right continuous. However, it is clear that for any $\epsilon > 0$,

$$\mathrm{P}_{U_{(n,1)}}[U_{n,1} < \epsilon] = 1 - (1-\epsilon)^n \longrightarrow 1$$

as $n \longrightarrow \infty$. Thus, in the limit, we do have

$$U_{(n,1)} \xrightarrow{d} Z$$

where $\mathrm{P}_Z[Z = 0] = 1$, that is, the limiting distribution is degenerate at 0.

Extreme order statistics: Limiting Distributions

Similarly, for any fixed u, we have

$$F_{U_{(n,n)}}(u) = \begin{cases} 0 & u < 0 \\ u^n & 0 \leq u \leq 1 \\ 1 & u > 1 \end{cases}$$

Extreme order statistics: Limiting Distributions

We have in the limit

$$F_{U_{(n,n)}}(u) \longrightarrow \begin{cases} 0 & u < 1 \\ 1 & u \geq 1 \end{cases}$$

that is, the limiting function is a step function, with a single step at 1. The limiting function is right continuous. and we have

$$U_{(n,n)} \xrightarrow{d} Z$$

where now $\mathrm{P}_Z[Z = 1] = 1$, that is, the limiting distribution is degenerate at 1.

Other order statistics: Limiting Distributions

For $1 < j < n$, we have from first principles that for $0 < u < 1$

$$F_{U_{(n,j)}}(u) = \sum_{k=j}^{n} \binom{n}{k} \{F_U(u)\}^k \{1 - F_U(u)\}^{n-k}$$

$$= \sum_{k=j}^{n} \binom{n}{k} u^k (1-u)^{n-k}$$

Other order statistics: Limiting Distributions

Note that

$$F_{U_{(n,j)}}(u) = \sum_{k=j}^{n} \binom{n}{k} \{F_U(u)\}^k \{1 - F_U(u)\}^{n-k} = \mathrm{P}[S_n(u) \geq j]$$

where $S_n(u) \sim \text{Binomial}(n, u)$; this result is heuristically reasonable, as

$$\begin{aligned} F_{U_{(n,j)}}(u) &= \mathrm{P}[\, j\text{th order statistic is} \leq u \,] \\ &= \mathrm{P}[\, \text{Fewer than } n - j \text{ observations} > u \,] \\ &= \sum_{k=j}^{n} \mathrm{P}[\, \text{Precisely } n - k \text{ observations} > u \,] \end{aligned}$$

Other order statistics: Limiting Distributions

It is clear that for any **fixed** j,

$$\mathbb{E}_{U_{n,j}}[U_{n,j}] = \frac{j}{n+1} \longrightarrow 0$$

as $n \longrightarrow \infty$. However, considering a fixed j does not really make sense as n grows.

For any real x, let $\lfloor x \rfloor$ denote the largest integer no greater than x. If we instead consider the sequence of $\{q_n = \lfloor nq \rfloor\}$th order statistics for some fixed q, $0 < q < 1$, we have

$$\mathbb{E}_{U_{(n,q_n)}}[U_{(n,q_n)}] = \frac{\lfloor nq \rfloor}{n+1} \longrightarrow q$$

that is, this sequence of 'proportionate' order statistics converges in expectation to q.

Other order statistics: Limiting Distributions

Similarly the variance of $U_{(n,q_n)}$ is

$$\frac{\lfloor nq \rfloor (n - \lfloor nq \rfloor + 1)}{(n+1)^2(n+2)} = \mathrm{O}(n^{-1})$$

and the variance converges to zero as $n \longrightarrow \infty$. Thus

$$U_{(n,q_n)} \xrightarrow{p} q$$

as $n \longrightarrow \infty$.

The quantity $U_{(n,q_n)}$ is the qth *sample quantile*.

Example: Gamma distribution

In the following Monte Carlo study, we consider the simulation of 1000 replicate data sets of different sample sizes from the

$$\text{Gamma}(2, 1)$$

distribution to study the sampling distribution of the order statistics. We have

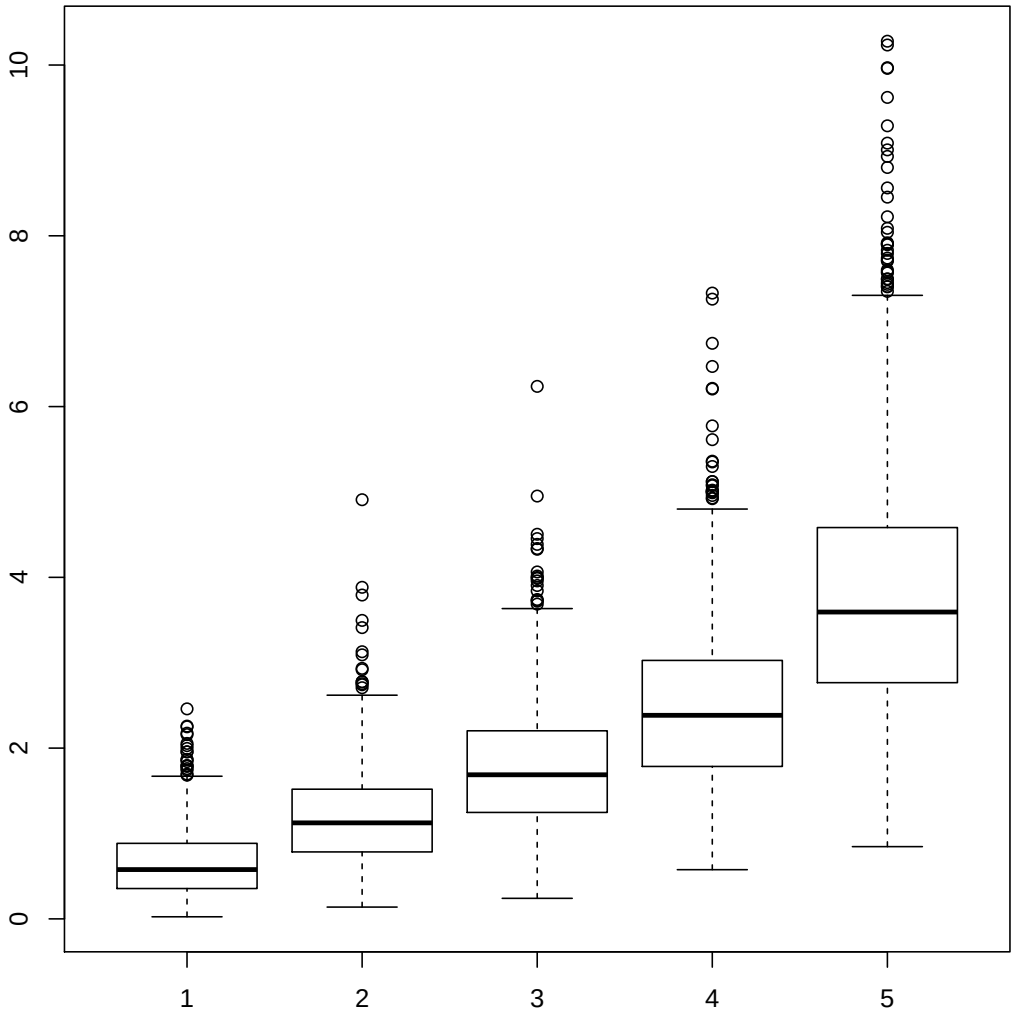
$$X_1, \ldots, X_n \sim \text{Gamma}(2, 1)$$

and

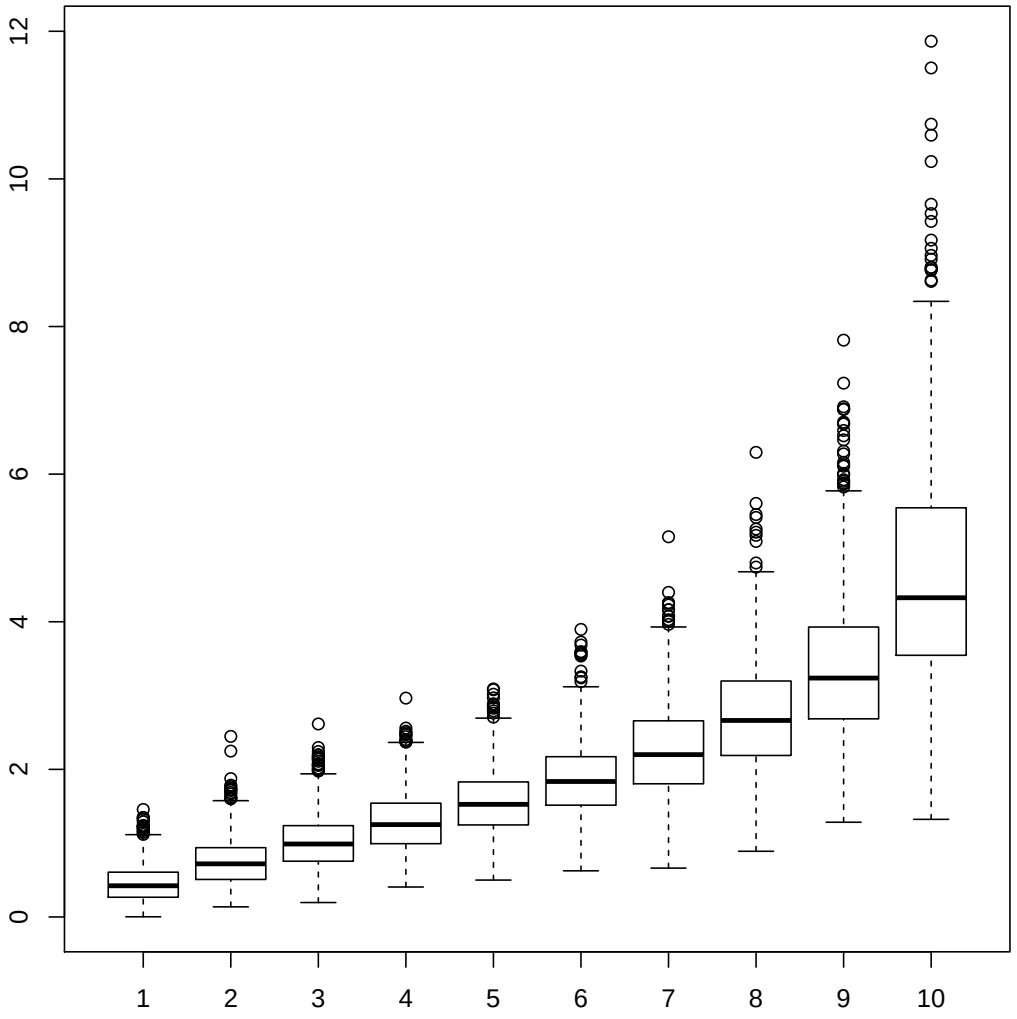
$$Y_{n,j} = j\text{th order statistic derived from } X_1, \ldots, X_n$$

The following figures depict boxplots of $Y_{(n,j)}$ for $j = 1, \ldots, n$, and $n = 5, 10, 20, 50, 100, 500$.

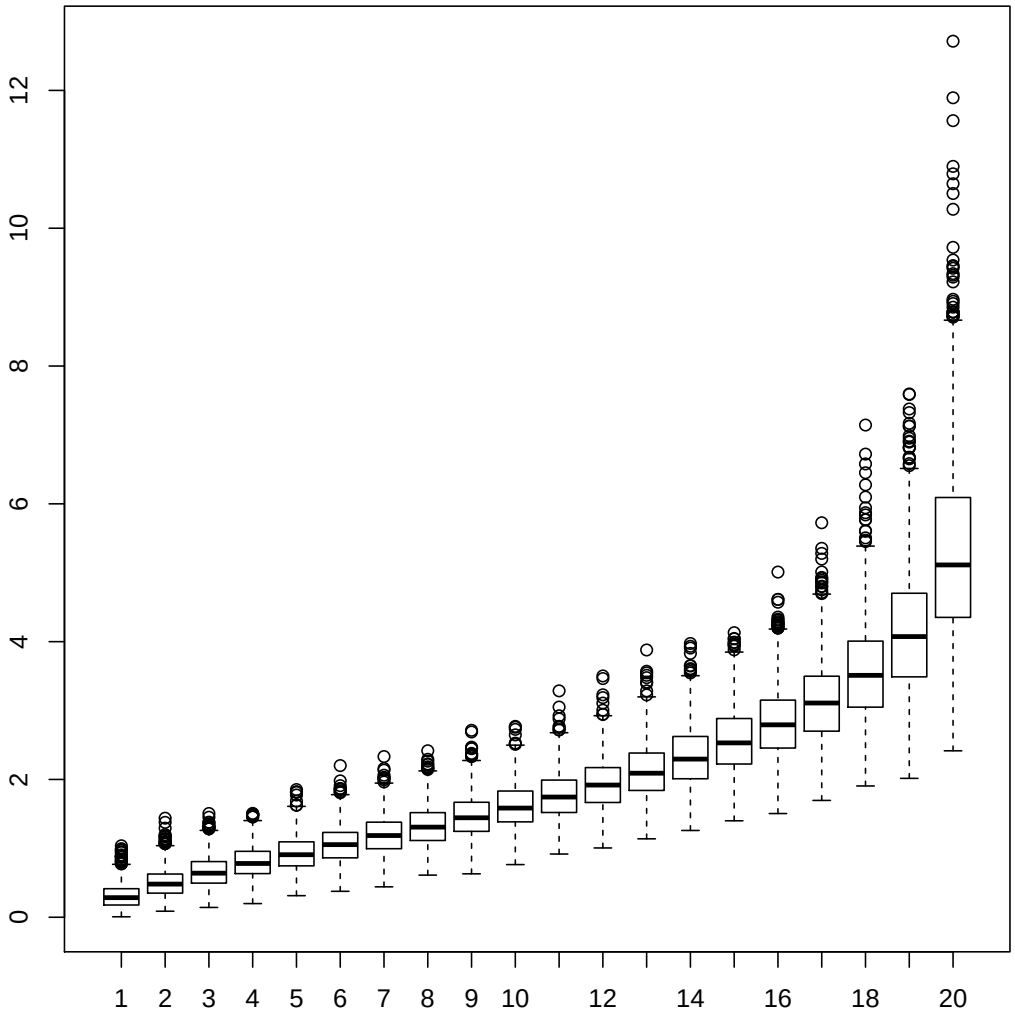
Monte Carlo study



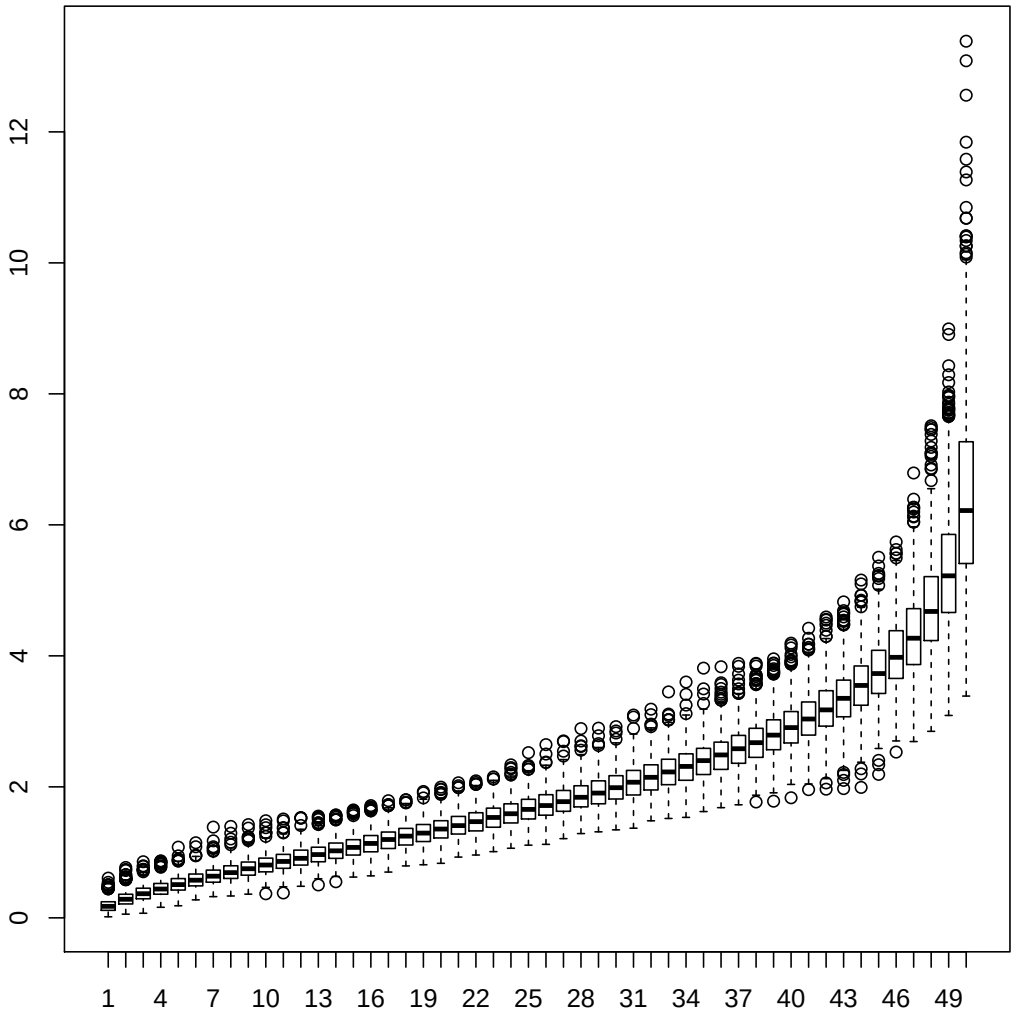
Monte Carlo study



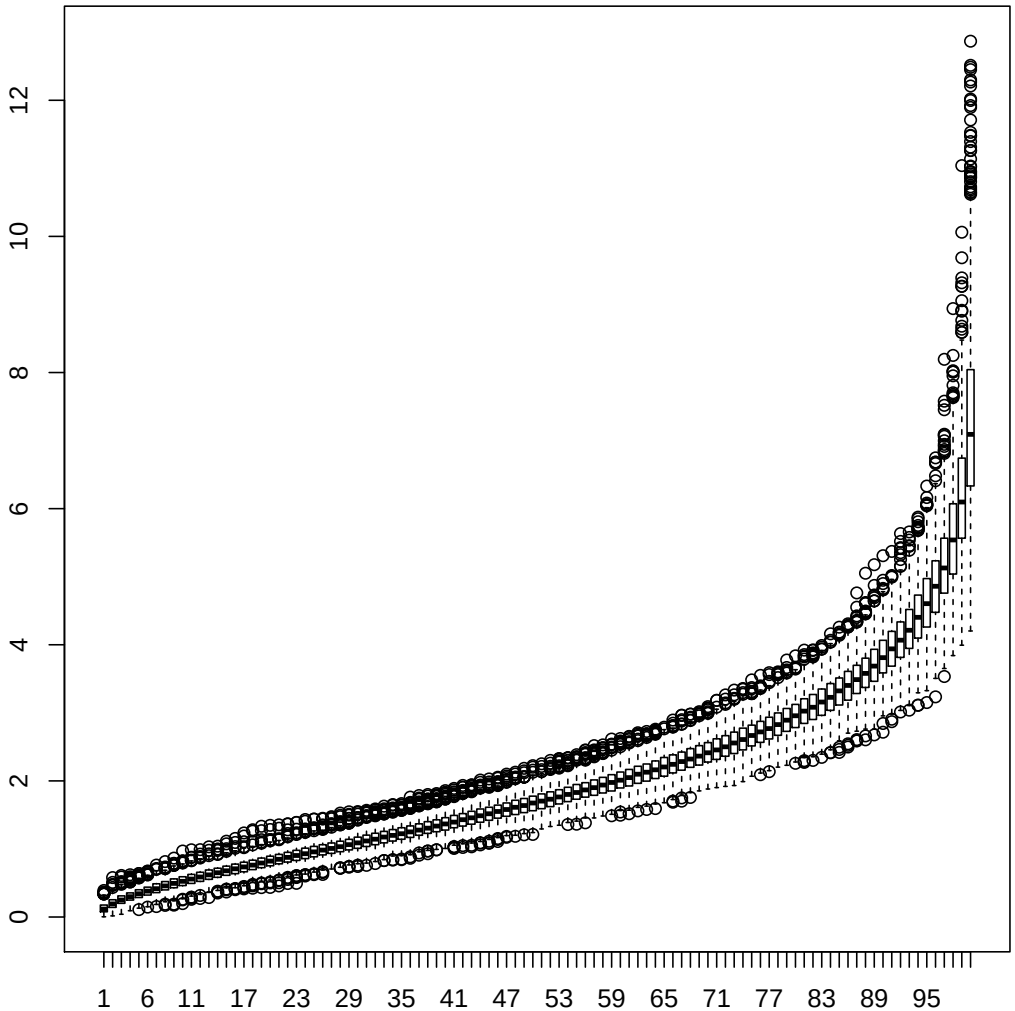
Monte Carlo study



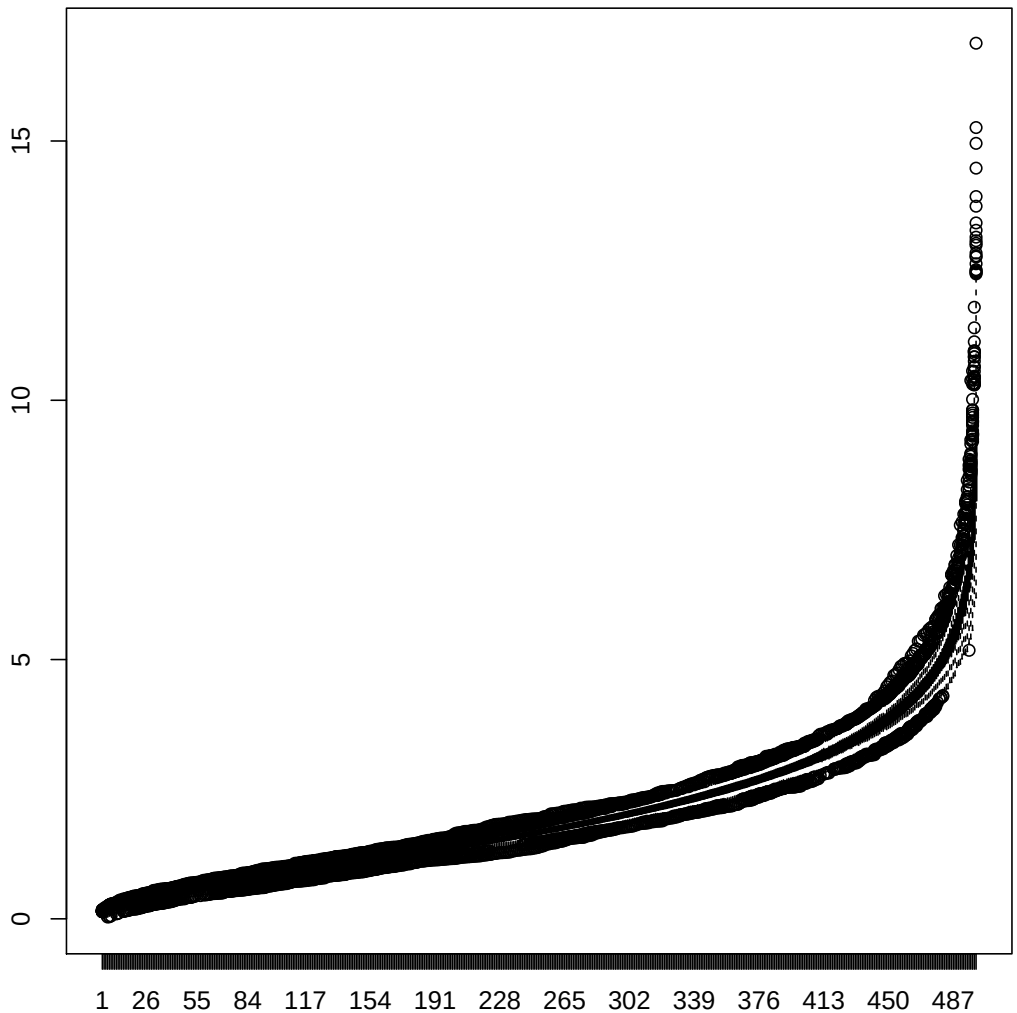
Monte Carlo study



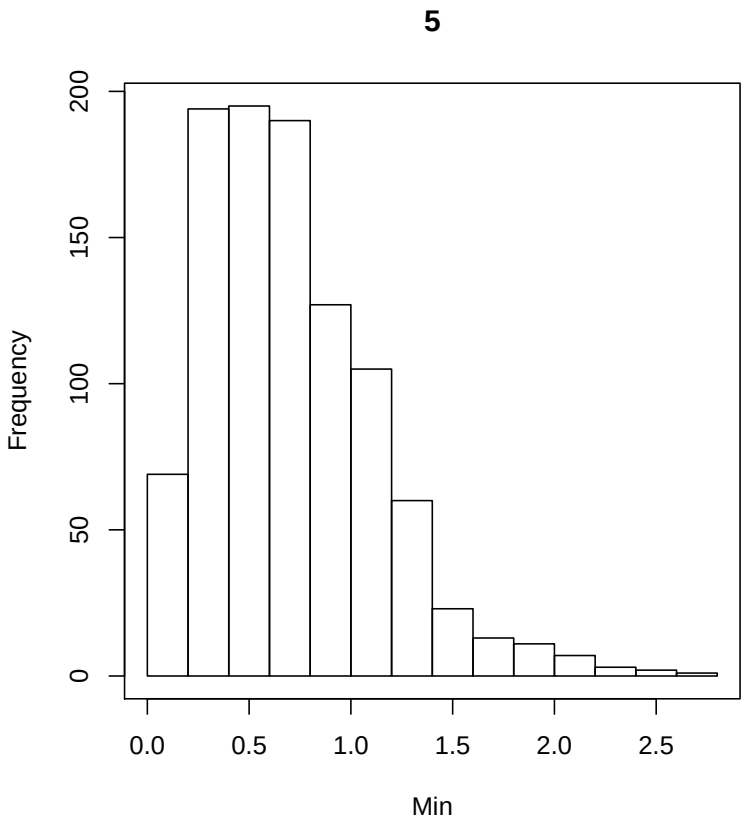
Monte Carlo study

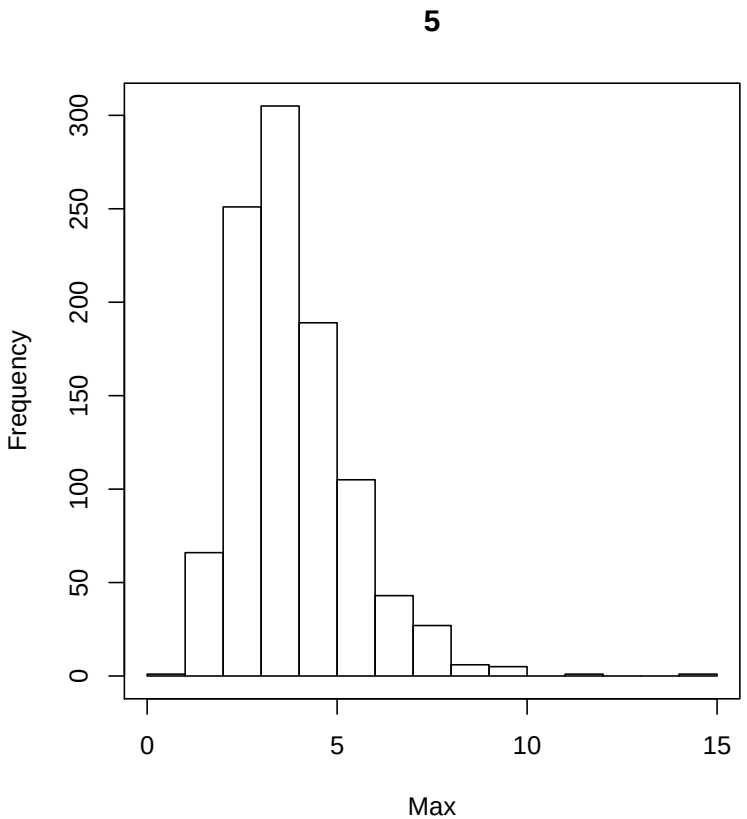


Monte Carlo study

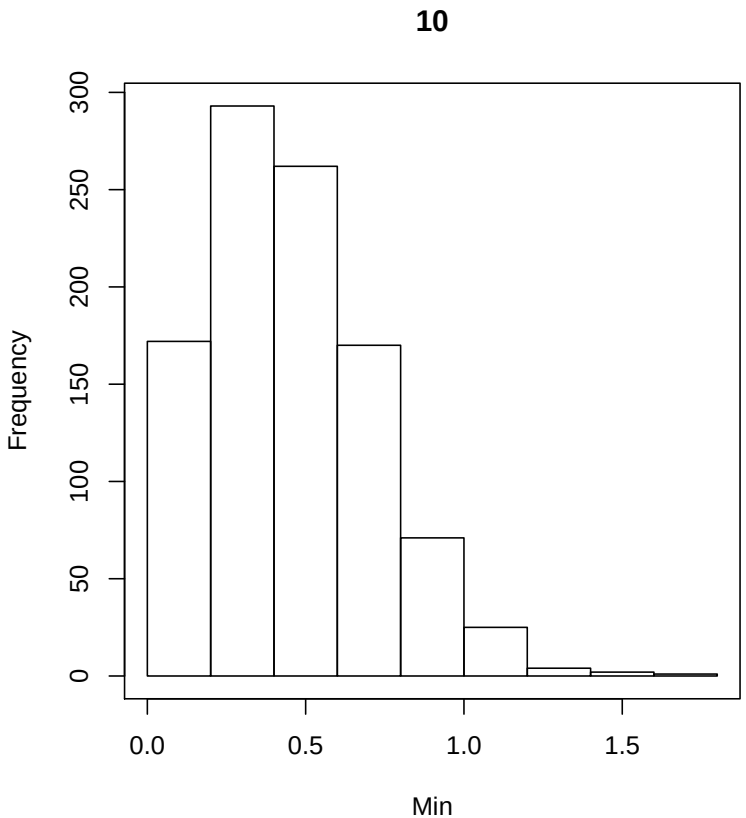


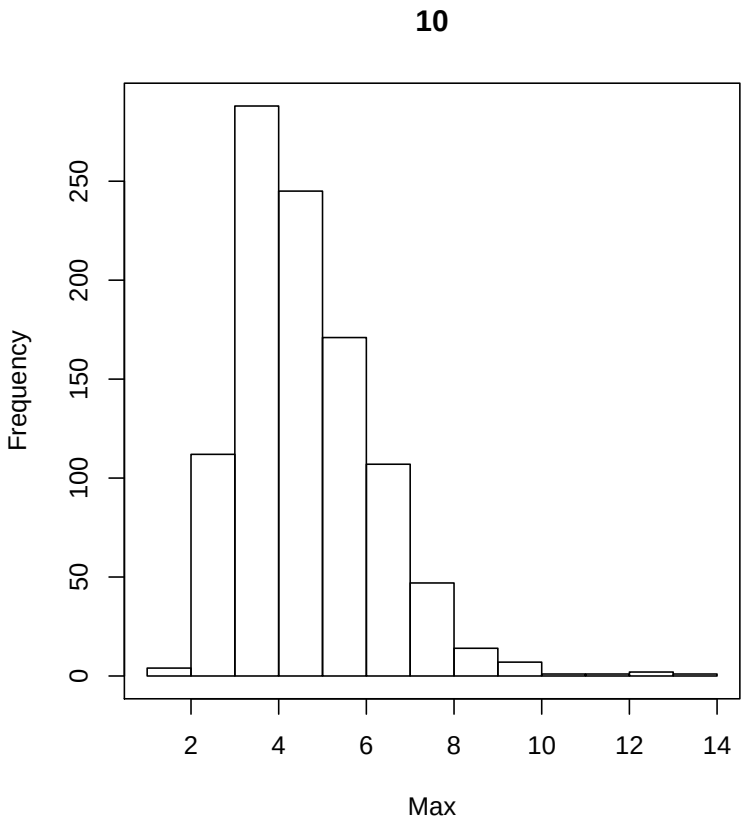
Monte Carlo study: Extreme order statistics



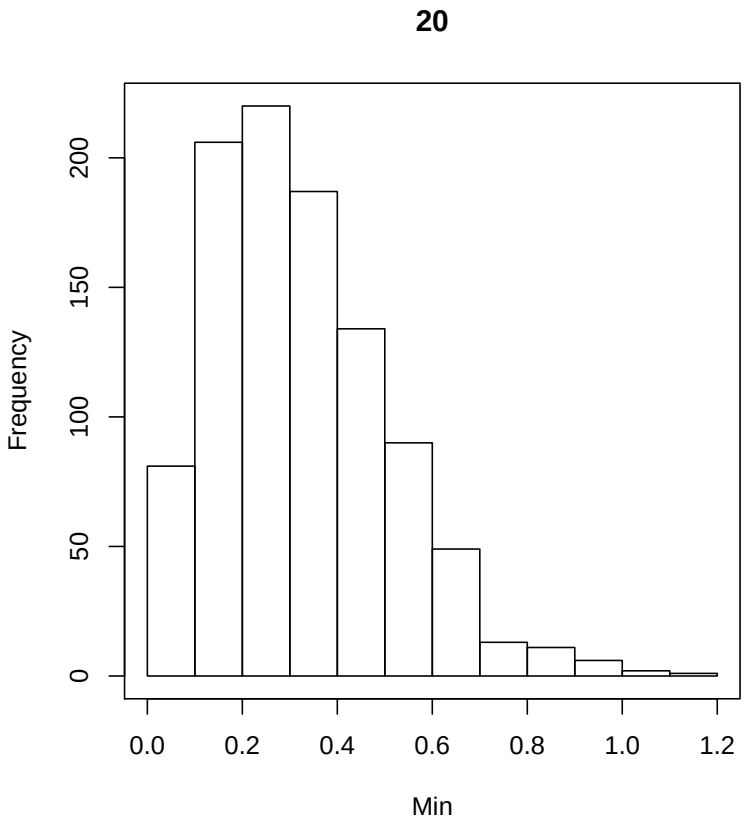


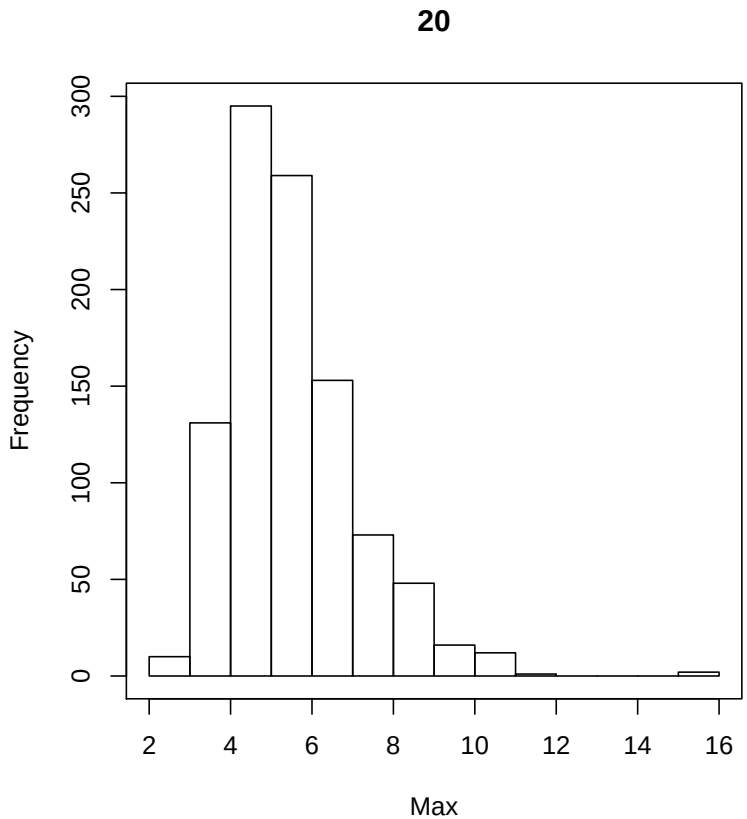
Monte Carlo study: Extreme order statistics



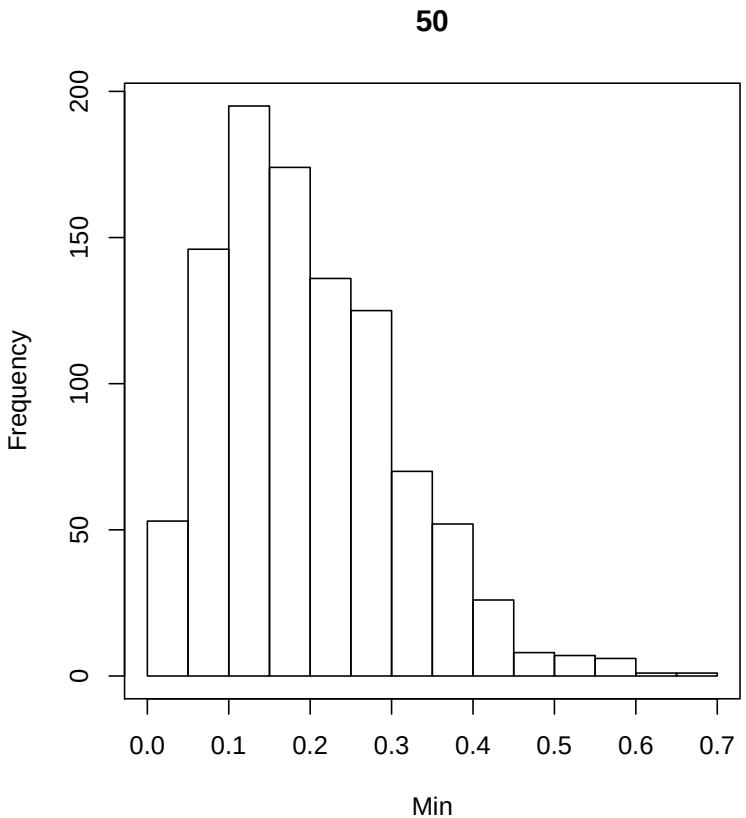


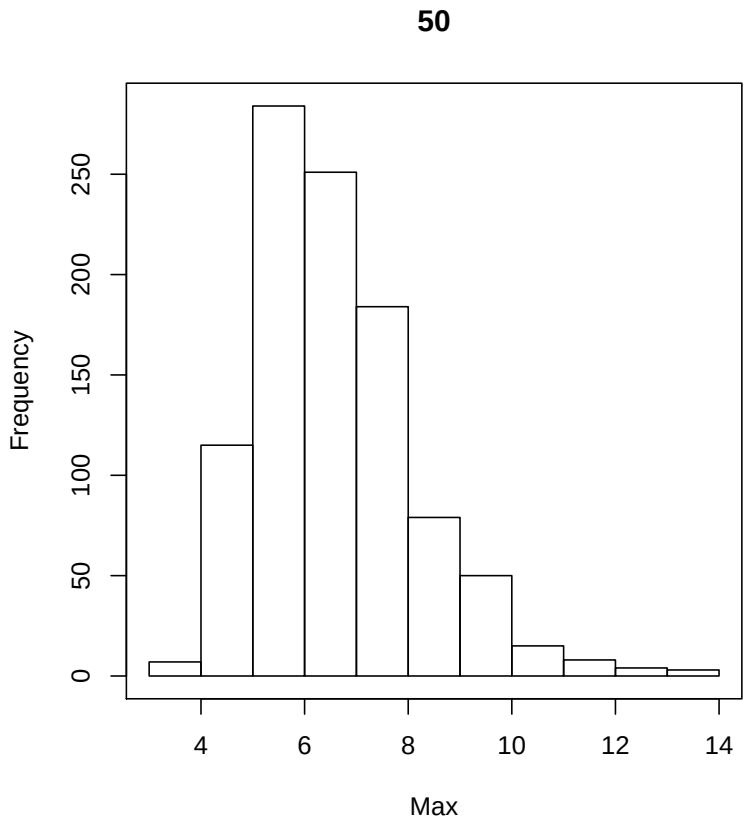
Monte Carlo study: Extreme order statistics



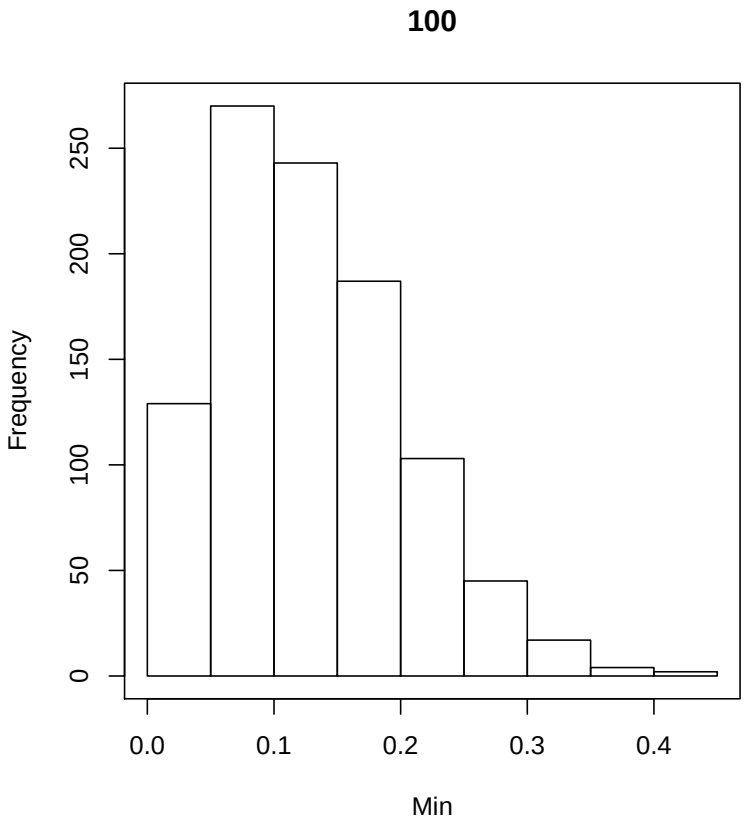


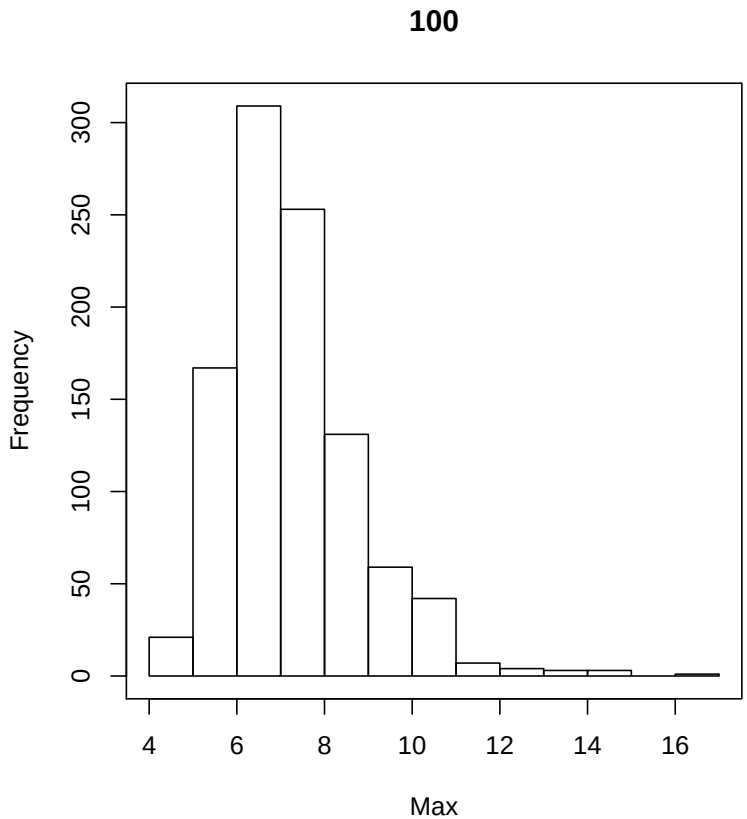
Monte Carlo study: Extreme order statistics



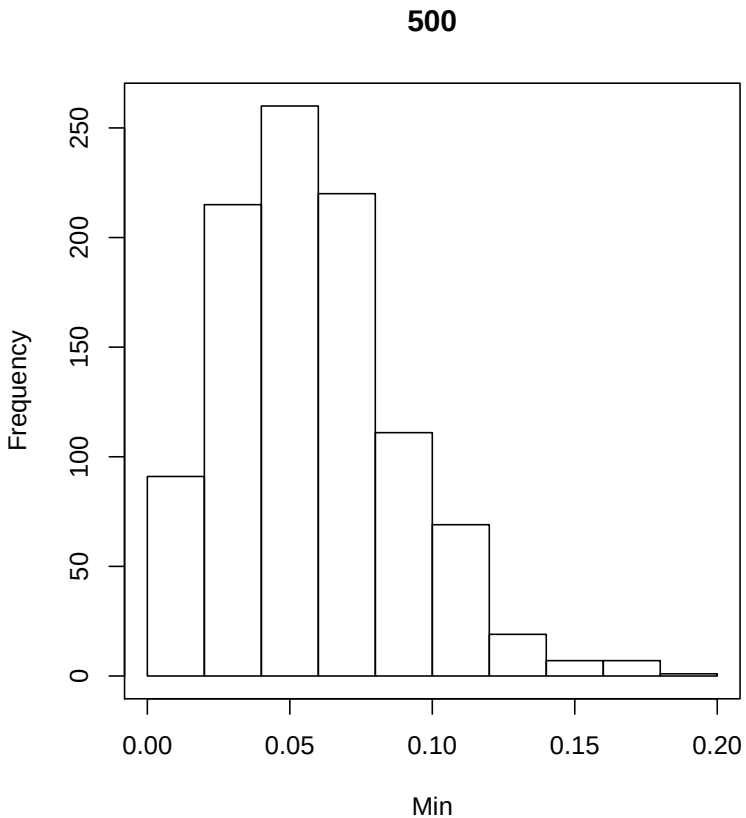


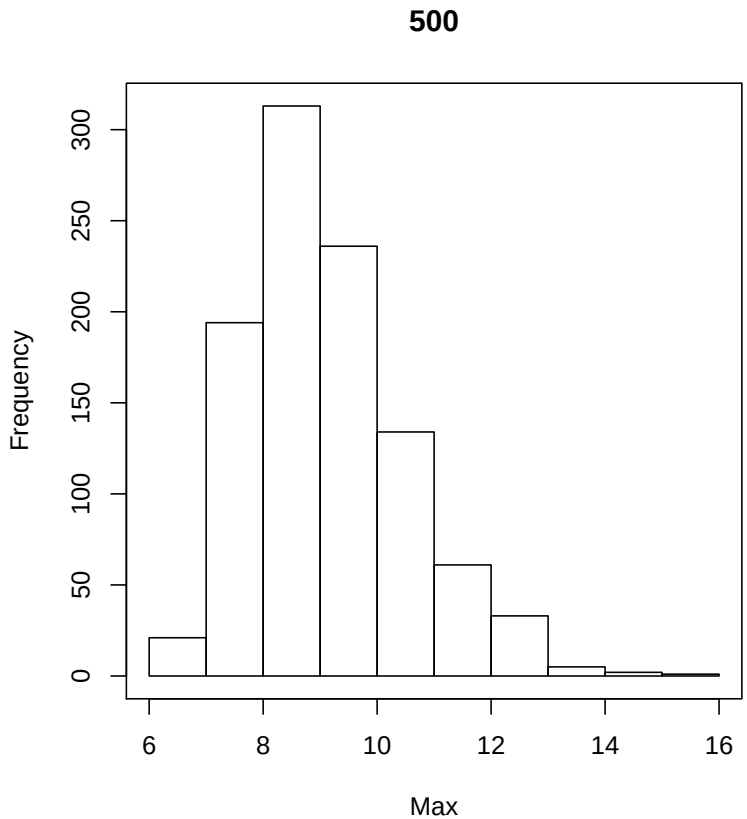
Monte Carlo study: Extreme order statistics





Monte Carlo study: Extreme order statistics





Gamma case: Asymptotic Properties

It is evident that $Y_{n,1}$ converges in distribution to a random variable that is degenerate at zero.

For the maximum order statistic, we have for each $n \geq 1$, by the Theorem of Total Probability

$$\begin{aligned}
\mathrm{P}[Y_{n+1,n+1} > Y_{n,n}] &= \int_0^\infty \mathrm{P}[Y_{n+1,n+1} > Y_{n,n} | Y_{n,n} = t] f_{Y_{n,n}}(t)\, \mathrm{d}t \\
&= \int_0^\infty \mathrm{P}[X_{n+1} > t] f_{Y_{n,n}}(t)\, \mathrm{d}t \\
&= \int_0^\infty (1 - F_X(t)) \left\{ n f_X(t) \{F_X(t)\}^{n-1} \right\} \mathrm{d}t \\
&= 1 - \int_0^\infty n f_X(t) \{F_X(t)\}^n \, \mathrm{d}t = \frac{1}{n+1}
\end{aligned}$$

In fact this result holds generally.

Gamma case: Asymptotic Properties

However, for **any** $y > 0$,

$$\mathrm{P}[Y_{n,n} > y] = 1 - \{F_X(y)\}^n \longrightarrow 1$$

as $n \longrightarrow \infty$. Therefore, there is **no limiting distribution** of the maximum order statistic, as the maximum order statistic grows without bound. Another way to write this is

$$Y_{n,n} \xrightarrow{p} +\infty$$

as $n \longrightarrow \infty$.

Gamma case: Asymptotic Properties

From the previous study of the Uniform$(0,1)$ order statistics, we know that

$$U_{(n,n)} \xrightarrow{p} 1$$

as $n \longrightarrow \infty$; we may then consider the inverse mapping

$$Y_{n,n} = F_X^{-1}(U_{(n,n)})$$

and from this deduce that

$$Y_{n,n} \xrightarrow{p} +\infty$$

as the mapping F_X^{-1} is continuous, and

$$F_X^{-1}(1) = +\infty.$$