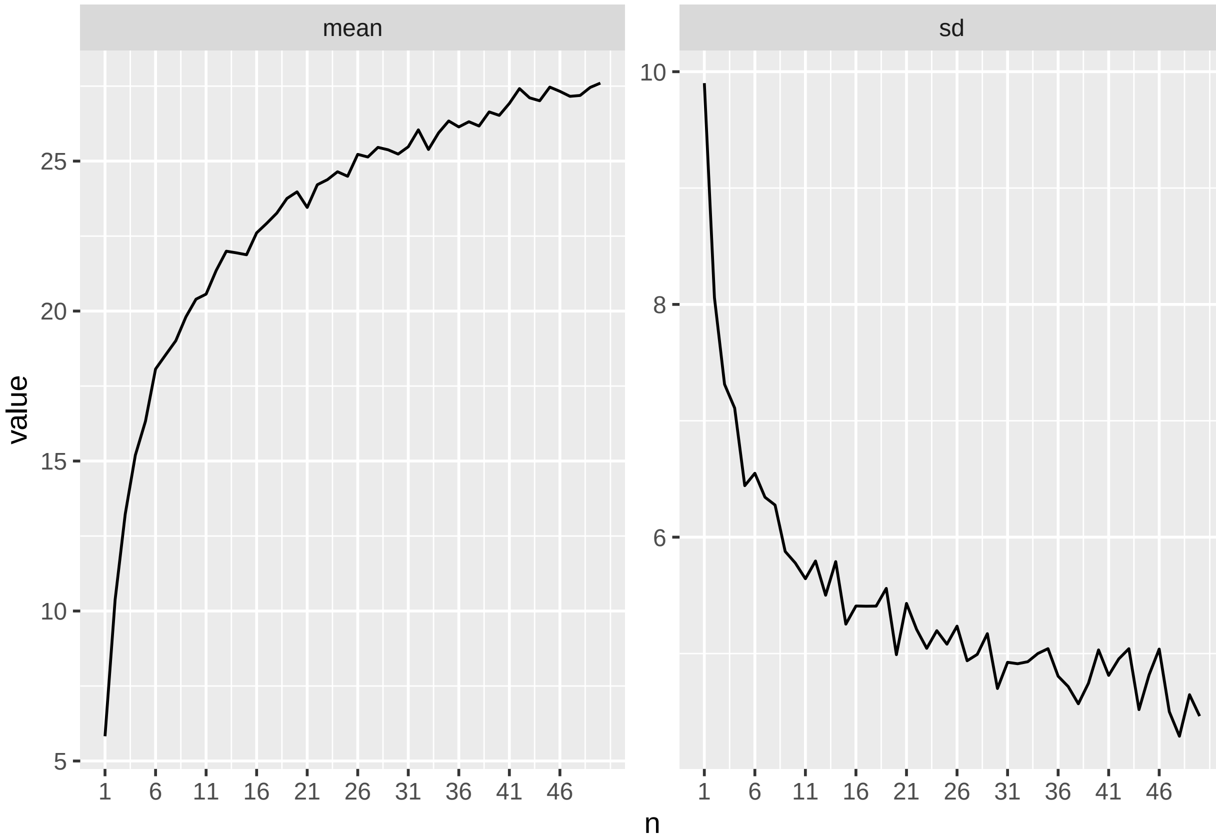
Mean and sd of null distribution

Sayani Gupta

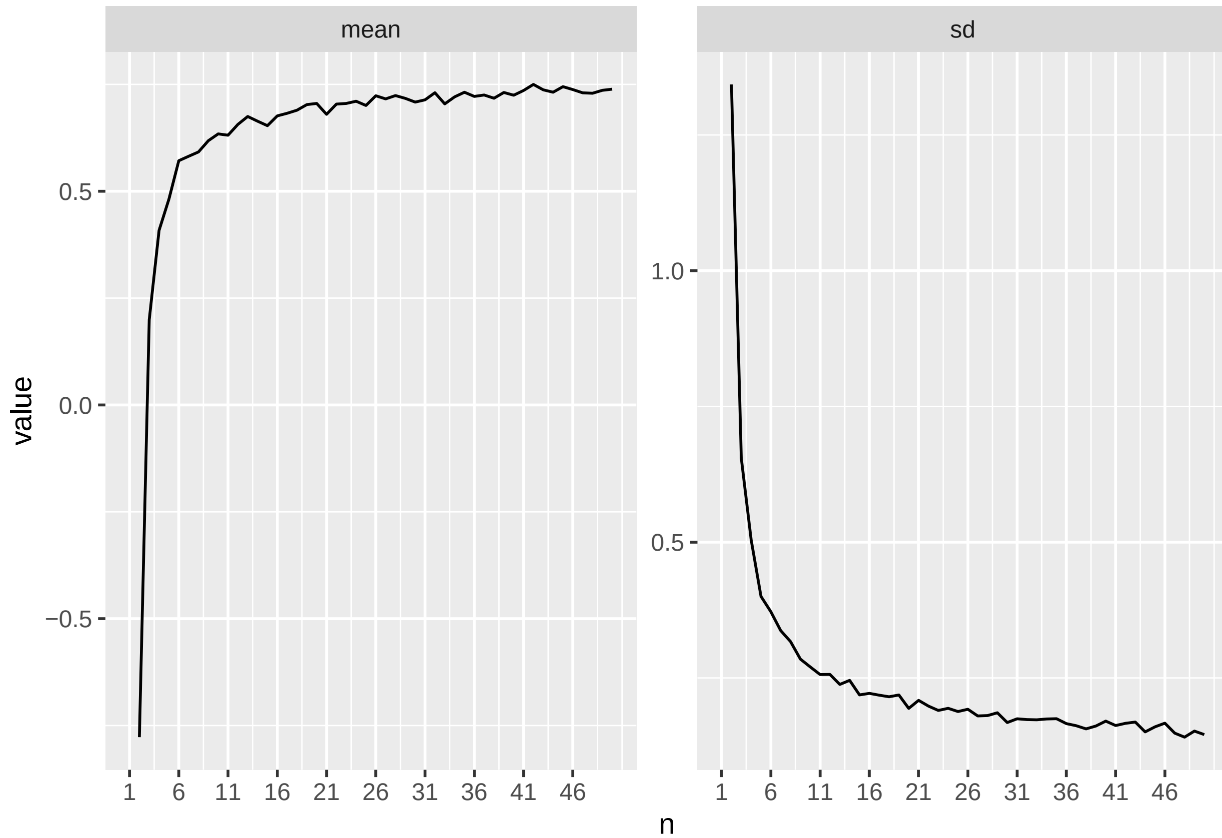
Suppose that X_1, X_2 , ... , X_n are i.i.d. random variables with expected values $E(X_i) = \mu < \infty$ and variance $Var(X_i) = \sigma^2 < \infty$. Let $Y = max(X_1, X_2, \ldots, X_n)$.

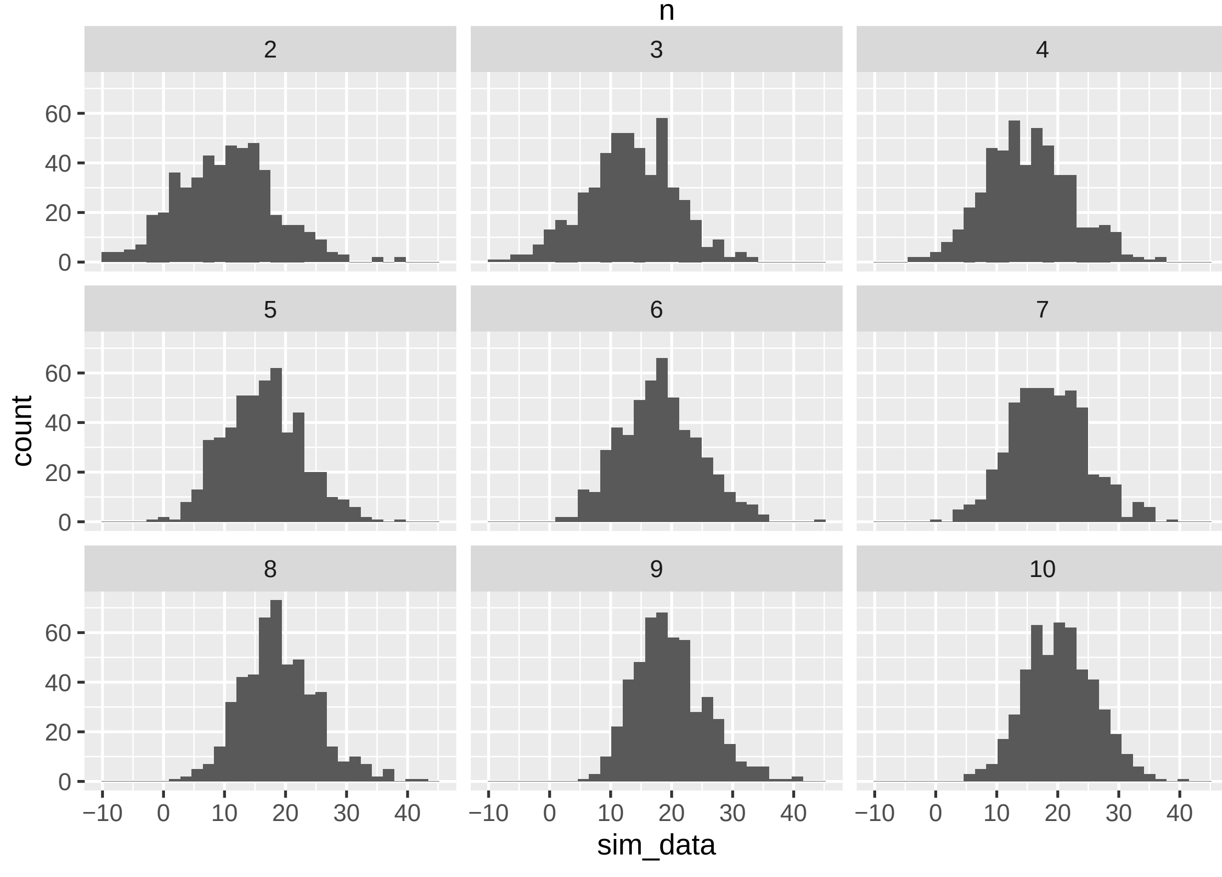
Let $F_X(x)$ be the common distribution of the variables X_i and let $F_Y(y)$ be the corresponding distribution of Y. $F_Y(y)$ could be obtained from $F_X(x)$ simply by using: $F_Y(y) = P[(X_1 \leq y) \cap (X_2 \leq y) \cap ... \cap (X_n \leq y)] = F_X(y)^n$. For large n, the distribution of Y approaches a standard shape, which does not depend on F_X. But what about the case when n is not large enough? The distribution of maximum in that case will indeed depend on n and the underlying distribution of X. If $F_X(x)$ is the CDF of X, then $F_Y(y) = F_x(y)^n$. Suppose Φ nd ϕ are the cdf and pdf of a standard normal distribution, then $f_Y(y) = n\Phi(y)^{n-1}\phi(y)$, which depends on n. Hence, we are trying to normalise for n. Also, it depends on the underlying distribution of X, which we have assumed as normal in our case. As n grows, we can see the right tail growing, which implies that the probability that we will get a higher maximum is more. Now, for large n, we used EVT to normalise for n, that is, we brought them to the same scale without distorting the range of the distribution. But in our case, we will mostly have small n. It is important to ensure that they have the same mean and variation, for being able to compare the maximum value across n. We observe from the following graphs that our normalisation works after $n = 6$, after which the difference in mean and standard deviation flattens out a lot.

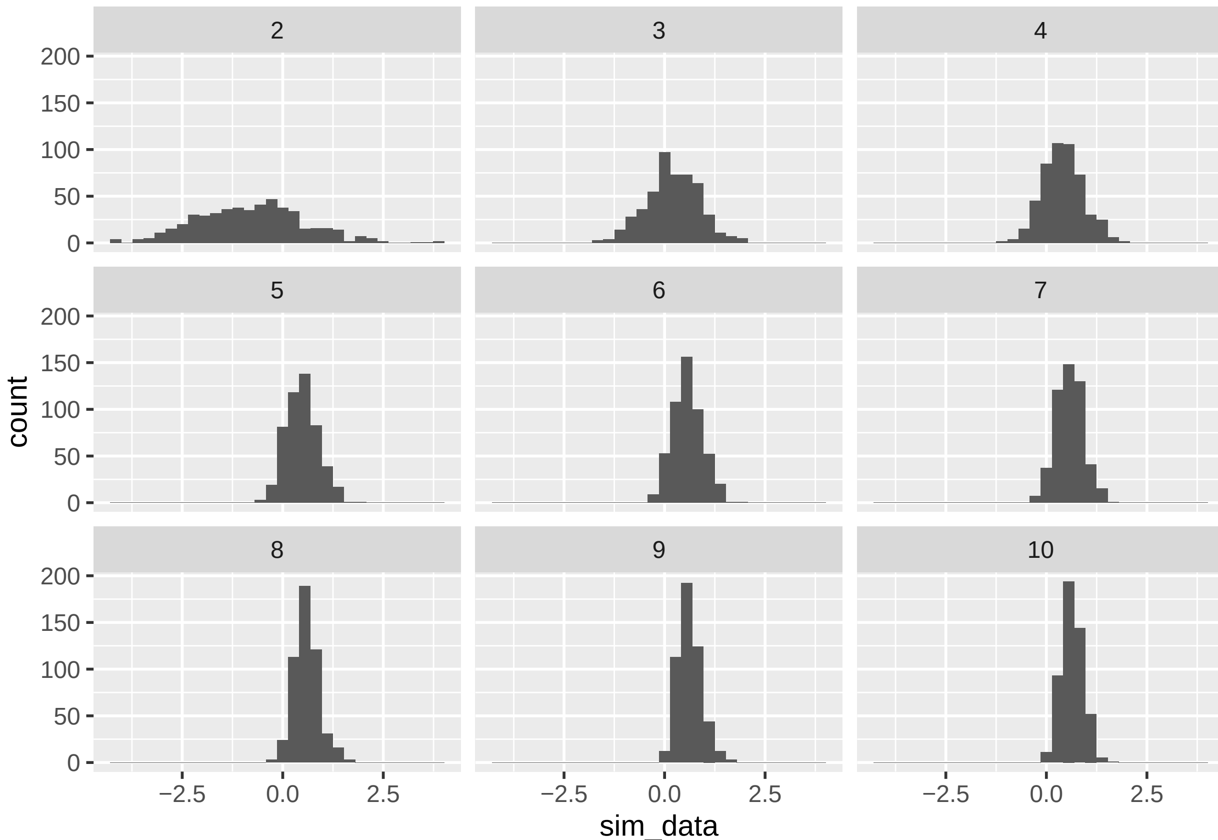
Mean and standard deviation of the distribution of maximum



Mean and standard deviation of the distribution of normalised maximum







Looking at the smaller values of n, what we already saw last week.

Distribution of max for smaller n

Distribution of normalised max for smaller n