Computational Statistics
The Bootstrap

georges.tod@outlook.com

March 2020

1 Introduction

In this report, numerical experiments illustrate the use of bootstrapping. As an application, the estimation of $\theta = E(X)/E(Y)$ is proposed, where X and Y are two independent gamma-distributed random variables. As an estimator of θ, we will investigate $T = \overline{X}/\overline{Y}$.
As a remainder, the Gamma distribution probability density function with parameters α and λ is defined by,

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \tag{1}$$

where $x > 0$, $\alpha > 0$, $\lambda > 0$ and $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$. For X and Y the following notations are respectively adopted, $\mu_X = \alpha_X/\lambda_X$ and $\mu_Y = \alpha_Y/\lambda_Y$. So the exact expression of θ is,

$$\theta = \frac{\alpha_X \lambda_Y}{\alpha_Y \lambda_X} \tag{2}$$

2 Empirical distributions

Let us assume we are given a sample $X_1, X_2, ..., X_n$ from X and $Y_1, Y_2, ..., Y_n$ from Y and that we know that X_i's are independent and identically gamma-distributed . Same for Y_i's. Assuming we know the distribution's parameters, for a fixed very large value of n, we simulate the true distribution of T that will be our reference, see Figure 1.

2.1 Parametric bootstrap

Let us assume we are given a sample $X_1, X_2, ..., X_n$ from X and $Y_1, Y_2, ..., Y_n$ from Y and that we know that X_i's are independent and identically gamma-distributed . Same for Y_i's. But in this case, we do not know the parameters of the distributions and n is fixed and not as large we might want it to be. In order to estimate the distribution parameters we use the method of moments. Below for X_1,

$$m_1 = E(X_1) = \hat{\mu}_X = \hat{\alpha}_X / \hat{\lambda}_X = \overline{X} \tag{3}$$

$$m_2 = E(X_1^2) = V(X_1) + E(X_1)^2 = \hat{\alpha}_X / \hat{\lambda}_X^2 + \overline{X}^2 = \overline{X^2} \tag{4}$$

The two previous equations give us the estimates for the Gamma distribution parameters from our sample,

$$\hat{\alpha}_X = \frac{\overline{X}^2}{\overline{X^2} - \overline{X}^2}, \hat{\lambda}_X = \frac{\overline{X}}{\overline{X^2} - \overline{X}^2} \tag{5}$$

Similar expressions can be obtained for the sample picked from Y.
With the estimated parameters, the distribution function of X can be estimated by,

$$\hat{F}_X = Gamma(\hat{\alpha}_X, \hat{\lambda}_X)$$

The parametric bootstrap method then consists in drawing B times n samples with replacement such that $X_1^*, X_2^*, ..., X_n^* \sim \hat{F}_X$. Similarly, the same can be done for Y, which allows us to obtain the empirical distribution of T^* from the observations,

$$T_b^* = \frac{\hat{\mu}^*_{X_b^*}}{\hat{\mu}^*_{Y_b^*}} = \frac{1/n \sum_{i=1}^n X_{i,b}^*}{1/n \sum_{i=1}^n Y_{i,b}^*} \tag{6}$$

with $b \in 1, .., B$, see Figure 1.

2.2 Nonparametric bootstrap

The nonparametric bootstrap is more general: it is assumed that the distribution from which the samples is taken is unknown. In this case, samples we draw B times n samples with replacement *directly* from $X_1, ..., X_n$ and $Y_1, ..., Y_n$. Using the expression of equation 6, the empirical distribution of T^* can be obtained, see Figure 1.

3 Bootstrap estimator quality

3.1 Exact bias and variance

The exact values of bias (given) and variance (see Appendices A and B) are,

$$\beta = E(T) - \theta = \theta \cdot \frac{1}{n_Y \alpha_Y - 1}$$

$$V(T) = \frac{\lambda_Y^2}{\lambda_X^2} \cdot \left(\frac{\alpha_X / n_X + \alpha_X^2}{(\alpha_Y - 1/n_Y)(\alpha_Y - 2/n_Y)} - \frac{\alpha_X^2}{(\alpha_Y - 1/n_Y)^2} \right)$$

If $n_X = n_Y = n$ then when $n \to +\infty$, both bias and variance converge to 0 at the same speed. Both of these quantities are plotted in black dashed lines on Figure 2. Estimations of these quantities are also plotted using the true distribution obtained earlier by simulation, see blue circles on Figure 2.

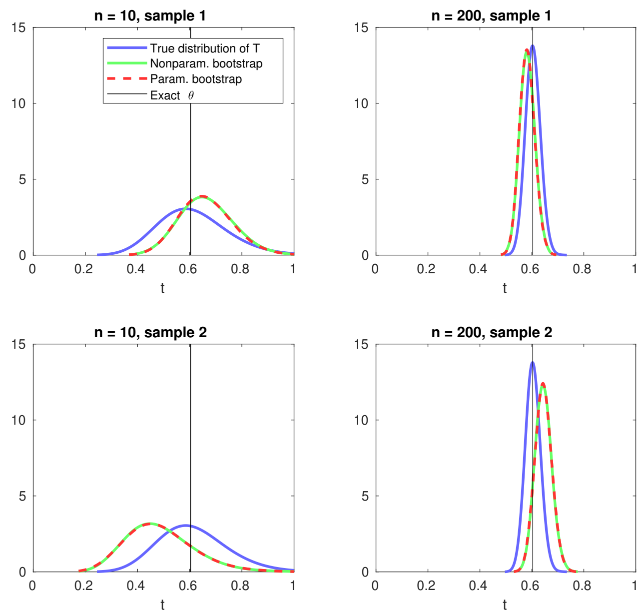

Figure 1: Nonparametric and parametric bootstraps compared to the true distribution for a couple of samples and sample sizes. The estimation of T is biased even at a larger sample size $n = 200$. From the fact we use less knowledge in the nonparametric bootstrap than in the parametric bootstrap, one would expect a poorer estimation of the true distribution using the nonparametric bootstrap. However, in this case, both give quite similar results. From a computational point of view, if B is chosen identical in the parametric and nonparametric bootstraps, then the nonparametric computation is faster as there are less operations that need to be performed: no need to sample from a distribution.

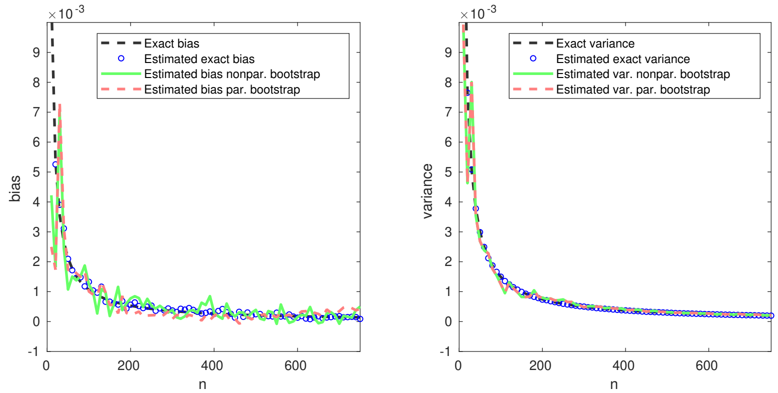

Figure 2: Bootstraps exhibit large bias and variance for small sample sizes, see here for $n < 100$. When the sample size increases, the bias converges to zero with growing n a bit more erratically than the variance. Both parametric and nonparametric bootstraps are equally good at capturing the variance. Remark: it should be kept in mind that we are comparing with θ^* only, which does not say much about how good of an estimator of θ, T^* is.

3.2 Parametric bootstrap bias and variance

In this case, $\theta^* = \frac{\hat{\alpha}_X \hat{\lambda}_Y}{\hat{\alpha}_Y \hat{\lambda}_X}$, therefore according to the definition of the bias β^* in the bootstrap world,

$$\beta^* = E(T^*) - \theta^* = \frac{1}{B}\sum_{b=1}^{B} T_b^* - \frac{\hat{\alpha}_X \hat{\lambda}_Y}{\hat{\alpha}_Y \hat{\lambda}_X} \tag{7}$$

And the variance,

$$V(T^*) = \frac{1}{B-1}\sum_{b=1}^{B}\left(T_b^* - \frac{1}{B}\sum_{b=1}^{B} T_b^*\right)^2 \tag{8}$$

3.3 Nonparametric bootstrap bias and variance

In this case, $\theta^* = \frac{\overline{X}}{\overline{Y}}$, therefore according to the definition of the bias β^* in the bootstrap world,

$$\beta^* = E(T^*) - \theta^* = \frac{1}{B}\sum_{b=1}^{B} T_b^* - \frac{\overline{X}}{\overline{Y}} \tag{9}$$

And the variance has the same expression than for the parametric bootstrap, see equation 8. Numerical results are presented on Figure 2.

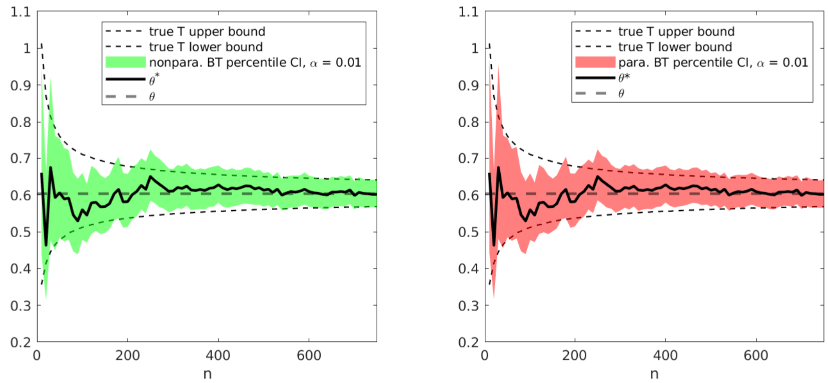

Figure 3: 'Truth' and bootstrap percentile confidence intervals at $\alpha = 0.01$. Similar results are found for both bootstraps. In both cases, the confidence intervals can under cover the true bounds. When the true bounds are over covered, this is less problematic as the result is just conservative. Furthermore, here we can grasp the potentially large deviations of θ^* w.r.t. θ and the fact bootstrapping should be used with care.

4 Confidence intervals

4.1 percentile intervals

The bootstrap approximation of the alternative confidence interval (CI) is given in theory for the median of the estimator, that we denote $\tau^* = median(T^*)$,

$$p(t^*_{1-\alpha/2} \leq \tau^* \leq t^*_{\alpha/2}) \approx 1 - \alpha$$

In any case, τ^* can be very easily computed for both parametric and nonparametric bootstraps. Now to obtain a confidence interval on θ^*, we introduce δ^* such that,

$$\delta^* = \tau^* - \theta^*$$

As a remainder, for the parametric bootstrap, $\theta^* = \frac{\hat{\alpha}_X \hat{\lambda}_Y}{\hat{\alpha}_Y \hat{\lambda}_X}$ is known. And for the nonparametric bootstrap, $\theta^* = T$ is also known. So finally, the percentile confidence intervals C_n for θ^* can be computed see Figure 3 by,

$$C_n = \left(t^*_{\alpha/2} - \delta^*, t^*_{1-\alpha/2} - \delta^* \right)$$

4.2 basic (or pivotal) interval

The basic bootstrap confidence interval for the nonparametric boostrap is given in theory by,

$$C_n = \left(2T - t^*_{1-\alpha/2}, 2T - t^*_{\alpha/2} \right)$$

See Figure 4 for an illustration.

4.3 t-like interval

A more beautiful confidence interval than the previous ones can be obtained using the bootstrap-t method.
First we define U^*, a studentized version of our bootstraped t statistics,

$$U^* = \frac{T^* - T}{S_T^*} \tag{10}$$

Where an approximation of S_T^* is given by,

$$S_T^* = {T^*}^2 \cdot \frac{A^*}{A^* + 1} \tag{11}$$

with,

$$A^* = \frac{S^2_{\overline{X^*}}}{\overline{X^*}^2 - S^2_{\overline{X^*}}} + \frac{S^2_{\overline{Y^*}}}{\overline{Y^*}^2 - S^2_{\overline{Y^*}}}$$

A confidence interval for θ^* is then given by,

$$C_n = \left(T - u^*_{1-\alpha/2} \cdot S_T, T - u^*_{\alpha/2} \cdot S_T\right) \tag{12}$$

where, S_T takes the same expression as S_T^* using directly the sample (no resampling), see equation 11. The drawback of this approach is it is limited by the knowledge of a good S_T candidate. See Figure 4 for an illustration.

A About the variance of the inverse of a Gamma random variable

Let us define, $Y \sim \Gamma(\alpha, \lambda)$. We know $E\left(\frac{1}{Y}\right) = \frac{\lambda}{\alpha - 1}$, then

$$\begin{aligned}
V\left(\frac{1}{Y}\right) =& E\left[\left(\frac{1}{Y} - E\left(\frac{1}{Y}\right)\right)^2\right] \\
=& \int_0^\infty \left(\frac{1}{y} - \frac{\lambda}{\alpha - 1}\right)^2 f(y)dy \\
=& \left(\frac{\lambda}{\alpha - 1}\right)^2 - 2\frac{\lambda}{\alpha - 1} \underbrace{\int_0^\infty \frac{1}{y} f(y)dy}_{=E(1/Y)} + \int_0^\infty \frac{1}{y^2} f(y)dy
\end{aligned}$$

For the last integral, we use $\Gamma(\alpha) = (\alpha - 1) \cdot (\alpha - 2) \cdot \Gamma(\alpha - 2)$

$$\begin{aligned}
\int_0^\infty \frac{1}{y^2} f(y)dy =& \frac{\lambda^2}{(\alpha - 1)(\alpha - 2)} \underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha - 2)} \cdot y^{\alpha - 3} \lambda^{\alpha - 2} e^{-\lambda y} dy}_{\text{integrand of the density of } \Gamma(\alpha - 2, \lambda) = 1} \\
=& \frac{\lambda^2}{(\alpha - 1)(\alpha - 2)}
\end{aligned}$$

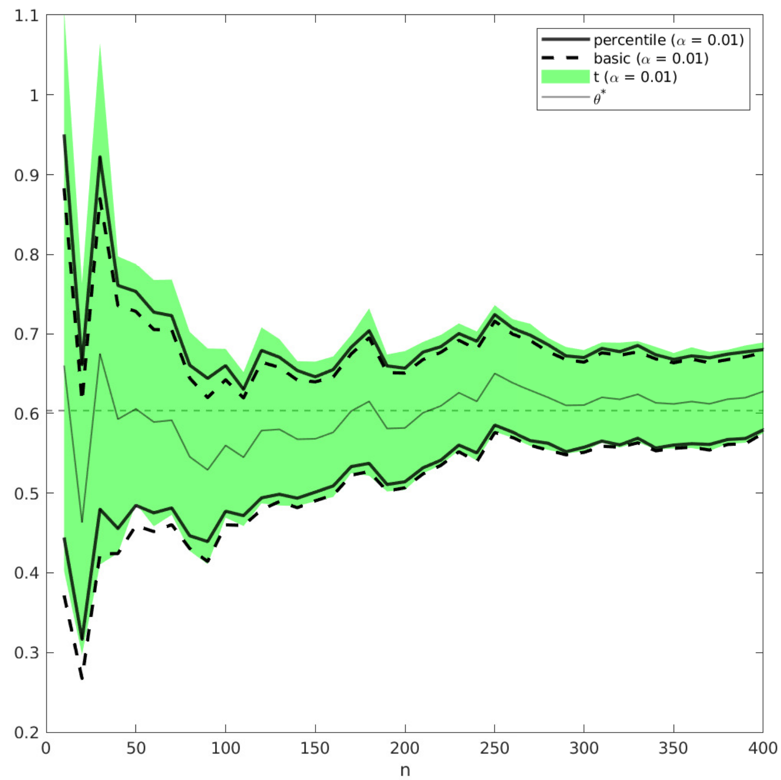

Figure 4: Comparing some nonparametric bootstrap confidence intervals at $\alpha = 0.01$. The main differences at noticeable for low sample sizes. The most conservative seems to be overall the t-bootstrap confidence interval.

Which leads finally to,

$$V\left(\frac{1}{Y}\right) = \frac{\lambda^2}{(\alpha-1)^2(\alpha-2)} \tag{13}$$

B About the exact value of the variance of T

By definition of T,

$$\begin{aligned} V(T) =& V\left(\frac{\overline{X}}{\overline{\overline{Y}}}\right) = E\left[\left(\frac{\overline{X}}{\overline{\overline{Y}}}\right)^2\right] - E(T)^2 \\ =& E(\overline{X}^2) \cdot E\left(\frac{1}{\overline{\overline{Y}}^2}\right) - E(T)^2 \end{aligned}$$

Where we know,

$$E(\overline{X}^2) = V(\overline{X}) + E(\overline{X})^2 = \frac{\alpha_X}{n_X \lambda_X^2} + \frac{\alpha_X^2}{\lambda_X^2}$$

And using known results and Appendix A,

$$\begin{aligned} E\left(\frac{1}{\overline{\overline{Y}}^2}\right) =& V\left(\frac{1}{\overline{\overline{Y}}}\right) + E\left(\frac{1}{\overline{\overline{Y}}}\right)^2 = \frac{n_Y^2 \lambda_Y^2}{(n_Y\alpha_Y - 1)^2(n_Y\alpha_Y - 2)} + \frac{n_Y^2\lambda_Y^2}{(n_Y\alpha_Y - 1)^2} \\ =& \frac{n_Y^2\lambda_Y^2}{(n_Y\alpha_Y - 1)(n_Y\alpha_Y - 2)} \end{aligned}$$

Which leads to,

$$V(T) = \left(\frac{\alpha_X}{n_X\lambda_X^2} + \frac{\alpha_X^2}{\lambda_X^2}\right) \cdot \frac{n_Y^2\lambda_Y^2}{(n_Y\alpha_Y - 1)(n_Y\alpha_Y - 2)} - \left(\frac{\alpha_X}{\lambda_X} \cdot \frac{n_Y\lambda_Y}{n_Y\alpha_Y - 1}\right)^2$$

And that we leave as a barbarian expression,

$$V(T) = \frac{\lambda_Y^2}{\lambda_X^2} \cdot \left(\frac{\alpha_X/n_X + \alpha_X^2}{(\alpha_Y - 1/n_Y)(\alpha_Y - 2/n_Y)} - \frac{\alpha_X^2}{(\alpha_Y - 1/n_Y)^2}\right) \tag{14}$$

Interestingly, if $n_X = n_Y = n$ then when $n \to +\infty$, $V(T) \to 0$. This is illustrated on Figure 2.