

Supplementary Materials

Mazet Olivier, Rodríguez Valcarce Willy, Chikhi Lounès

1 Proof of the lemma

Lemma 1.1 *Given set of n_L independent observations $(t_1, t_2, \dots, t_{n_L})$, the log-likelihood function $\log(\mathbb{L}_{SSPSC})$ has no critical points in \mathbb{R}^2 .*

Proof Given n_L independent values $t = (t_1, t_2, \dots, t_{n_L})$, the likelihood is:

$$\mathbb{L}_{SSPSC}(\alpha, T) = \prod_{i=1}^{n_L} \mathbb{L}_{t_i}(T, \alpha),$$

and taking the log:

$$\log(\mathbb{L}_{SSPSC}(\alpha, T)) = \sum_{i=1}^{n_L} \log(\mathbb{L}_{t_i}(\alpha, T)), \quad (1)$$

where

$$\mathbb{L}_{t_i}(T, \alpha) = \frac{1}{\alpha} e^{-T - \frac{1}{\alpha}(t_i - T)} \mathbb{I}_{T \leq t_i} + e^{-t_i} \mathbb{I}_{T > t_i} \quad (2)$$

First note that:

- For $T > t_i$, $\mathbb{L}_{t_i}(\alpha, T) = e^{-t_i}$ is **constant** (with respect to (α, T)).
- If $\alpha \neq 1$ ($\alpha = 1$ means there is no change in the population's size) then $\mathbb{L}_{t_i}(T, \alpha)$ has a discontinuity at $T = t_i$.

As we are interested in the case $\alpha \neq 1$, the log-likelihood function has discontinuities at each t_i , $i = 1 \dots n_L$.

For $i \in \{0, 1, \dots, n_L + 1\}$, let $C_i = \{(\alpha, T) \in \mathbb{R}^+ \times \mathbb{R}^+, t_i < T < t_{i+1}\}$ with $t_0 = 0$ and $t_{n_L+1} = +\infty$.

Now let be $C = \bigcup_{i=0}^{n_L} C_i$ (Figure 1). We can see that $\mathbb{L}_{SSPSC}(\alpha, T)$ is continuously differentiable in the interior of C .

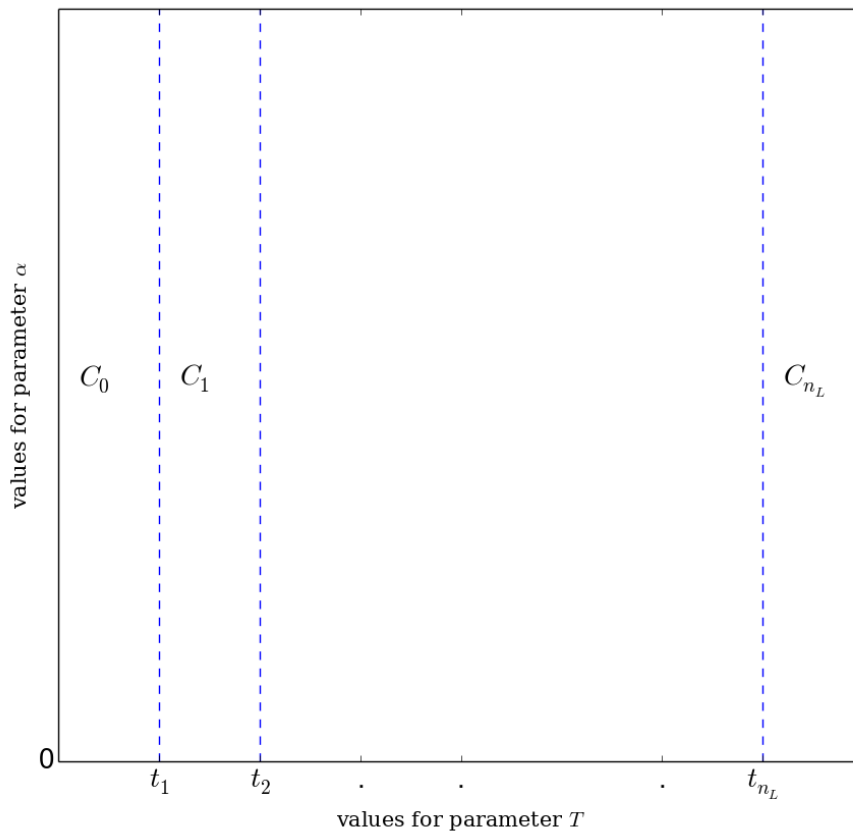


Figure 1: Cutting the parameter space, taking into account the discontinuities of the log-likelihood function in \mathbb{R}^2

Given that we do not consider negative values for α or T , we split the parameter space into two subsets C and $\mathbb{R}_+^2 \setminus C$. If $(\alpha, T) \in C$, taking the log and the derivative with respect to T in equation (2) gives:

$$\frac{\partial}{\partial T} \mathbb{L}_{t_i}(\alpha, T) = \begin{cases} -1 + \frac{1}{\alpha} & \text{if } T < t_i \\ 0 & \text{otherwise.} \end{cases}$$

As we can see, if $\alpha < 1$ then $\frac{\partial}{\partial T} \mathbb{L}_{t_i}(\alpha, T) > 0$ for all i and if $\alpha > 1$ then $\frac{\partial}{\partial T} \mathbb{L}_{t_i}(\alpha, T) < 0$ for all i . A consequence, $\nabla \mathbb{L}_{SSPSC}(\alpha, T)$ will never be zero in the interior of C if $\alpha \neq 1$. \square

This fact suggests that the *min* and *max* values of \mathbb{L}_{SSPSC} (if they exist) have the form (α, t_i) .

Let's find the critical points of \mathbb{L}_{SSPSC} over the lines (α, t_i) with $t_i \in \{t_1, t_2, \dots, t_{n_L}\}$. When we fix the value of T the function becomes a function of the single variable α . If $T = t_a$ for $a \in \{1, \dots, n_L\}$ it follows from (1) that:

$$\log(\mathbb{L}_{SSPSC}(\alpha, t_a)) = \log\left(\frac{1}{\alpha}\right) - t_a - \sum_{i=1}^{n_L} \left[\log\left(\frac{1}{\alpha}\right) - t_a - \frac{1}{\alpha}(t_i - t_a) \right] \mathbb{I}_{t_a < t_i} - \sum_{i=1}^{n_L} t_i \mathbb{I}_{t_a > t_i}.$$

Denoting $K = \sum_{i=1}^{n_L} \mathbb{I}_{t_i < t_a}$, we then have:

$$\log(\mathbb{L}_{SSPSC}(\alpha, t_a)) = (K+1) \left(\log\left(\frac{1}{\alpha}\right) - t_a \right) - \frac{1}{\alpha} \sum_{i=1}^{n_L} (t_i - t_a) \mathbb{I}_{t_a < t_i} - \sum_{i=1}^{n_L} t_i \mathbb{I}_{t_a > t_i}.$$

Let us find the zeros of the derivative in α :

$$\begin{aligned} \log(\mathbb{L}_{SSPSC}(\alpha, t_a)) = 0 &\Leftrightarrow - (K+1) \frac{1}{\alpha} + \frac{1}{\alpha^2} \sum_{i=1}^{n_L} (t_i - t_a) \mathbb{I}_{t_a < t_i} = 0 \\ &\Leftrightarrow - \alpha(K+1) + \sum_{i=1}^{n_L} (t_i - t_a) \mathbb{I}_{t_a < t_i} = 0 \\ &\Leftrightarrow \alpha = \frac{\sum_{i=1}^{n_L} t_i \mathbb{I}_{t_a < t_i} - K t_a}{K+1}. \end{aligned}$$

Hence, the maximum value of the log-likelihood function (if it exists) is of the form:

$$m_a = \left(\frac{\sum_{i=1}^{n_L} t_i \mathbb{I}_{t_a < t_i} - K t_a}{K + 1}, t_a \right), a \in \{1, 2, \dots, n_L\}.$$

We then take $(\hat{\alpha}, \hat{T}) = \operatorname{argmax}_{a \in \{1, \dots, n_L\}} \{\log(\mathbb{L}_{SSPSC}(m_a))\}$ as the Maximum Likelihood Estimation.

Remark: Note that the function \mathbb{L}_{SSPSC} does not have an upper bound. Let's say $(T_1, T_2, \dots, T_{n_L})$ are the n_L observations of T_2 sorted from the lower value to the higher value.

For $T = T_{n_L}$ we have from (1) and (2):

$$\mathbb{L}_{SSPSC}(\alpha, T_{n_L}) = \log\left(\frac{1}{\alpha}\right) - T_{n_L} - \sum_{i=1}^{n_L-1} T_i$$

which clearly goes to $+\infty$ as α goes to zero. The same problem is present for every $T \in]T_{n_L-1}, T_{n_L}]$

Nevertheless, in practice, our m_a is a good estimate for the parameters.

2 Proportion of rejections

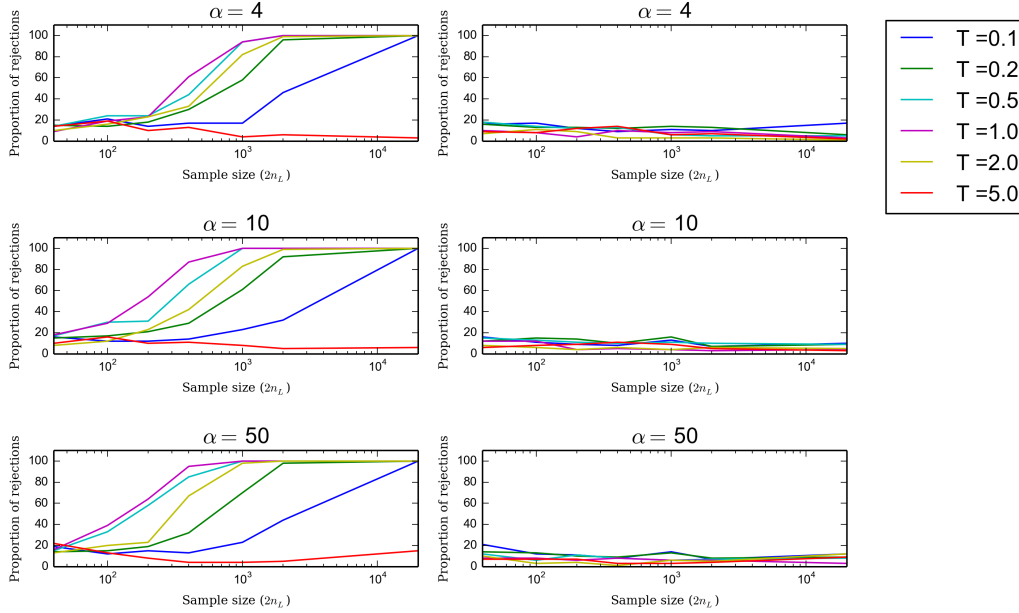


Figure 2: Proportion of rejected data sets simulated under the SSPSC model. Left panels: the reference model is the StSI model (the wrong model). Right panels: the reference model is the SSPSC (the right model), *i.e.* the model under which the data were simulated. Note that for the abscissa we used $2n_L$ instead of n_L because in order to perform the KS test it is necessary to first estimate the parameters using n_L loci and then an independent set of n_L values of T_2 .

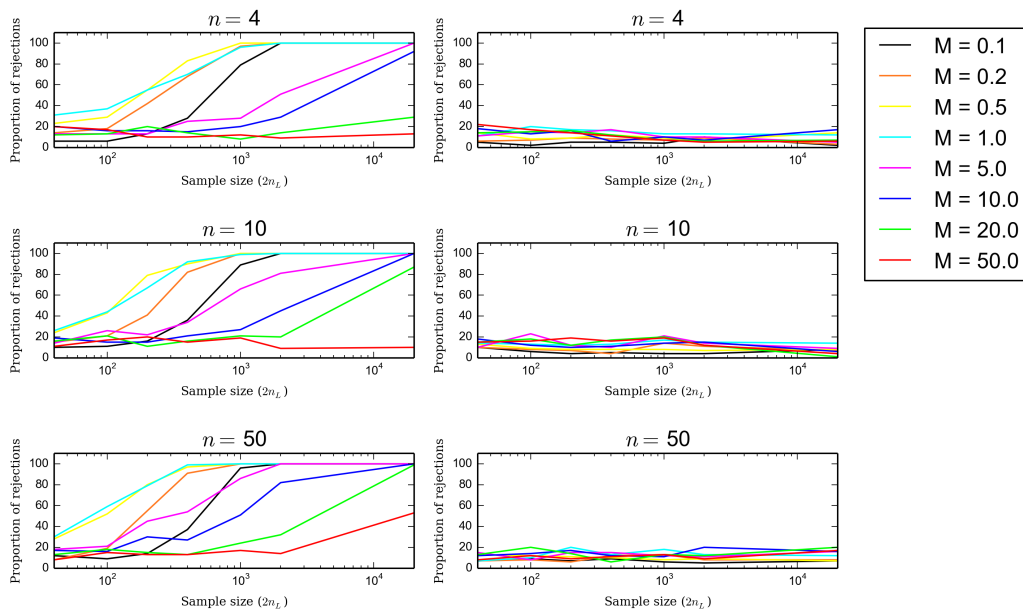


Figure 3: Proportion of rejected data sets simulated under the StSI model. Left panels: the reference model is the SSPSC (the wrong model). Right panels: the reference model is the StSI model (the right model), *i.e.* the model under which the data were simulated. Note that for the abscissa we used $2n_L$ instead of n_L because in order to perform the KS test it is necessary to first estimate the parameters using n_L loci and then an independent set of n_L values of T_2 .

3 Estimation accuracy

All the scripts used in the simulations are available from http://github.com/willyrv/SSPSC_vs_StSI

Figures 4, 5, 6, 7, 8 and 9 show the results of the estimation of α for all the values of T and all the sample sizes, under the SSPSC model. The values where $\alpha = (2, 4, 10, 20, 50, 100)$, $T = (0.1, 0.2, 0.5, 1, 2, 5)$ and $n_L = (20, 50, 100, 200, 500, 1000, 10000)$

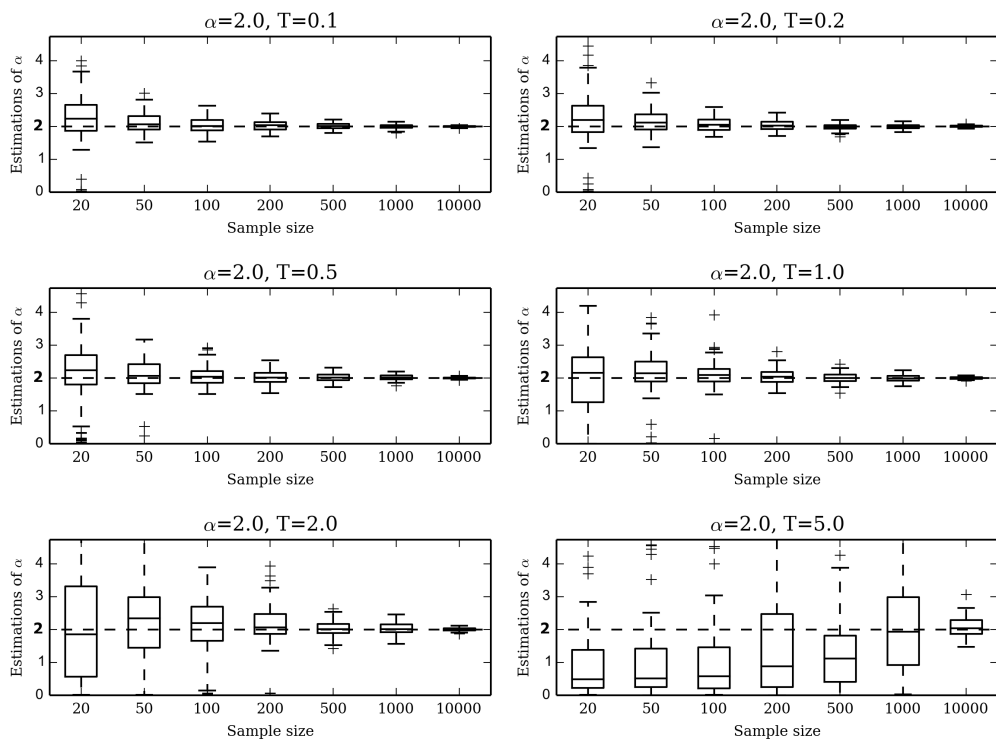


Figure 4: Estimations of α (real value $\alpha = 2.0$)

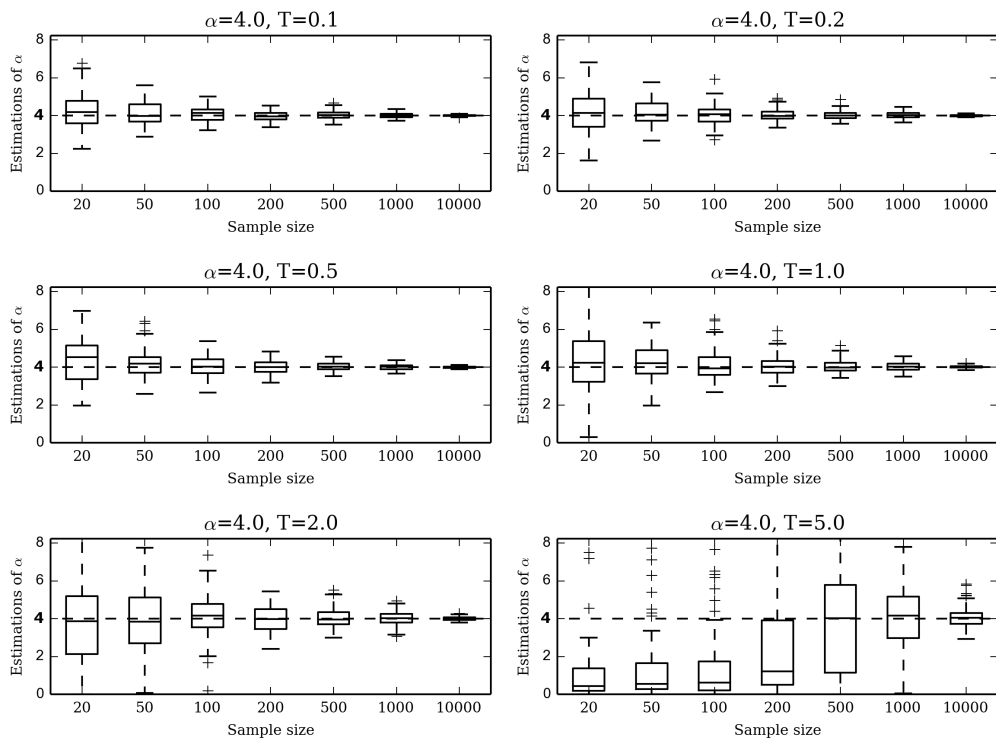


Figure 5: Estimations of α (real value $\alpha = 4.0$)

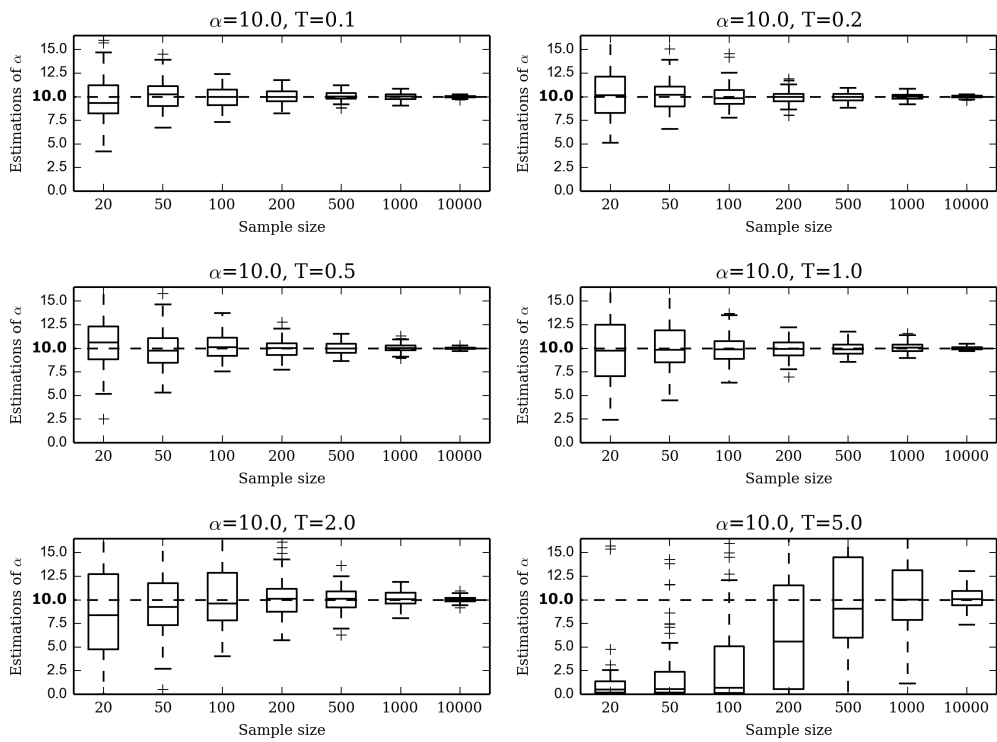


Figure 6: Estimations of α (real value $\alpha = 10.0$)

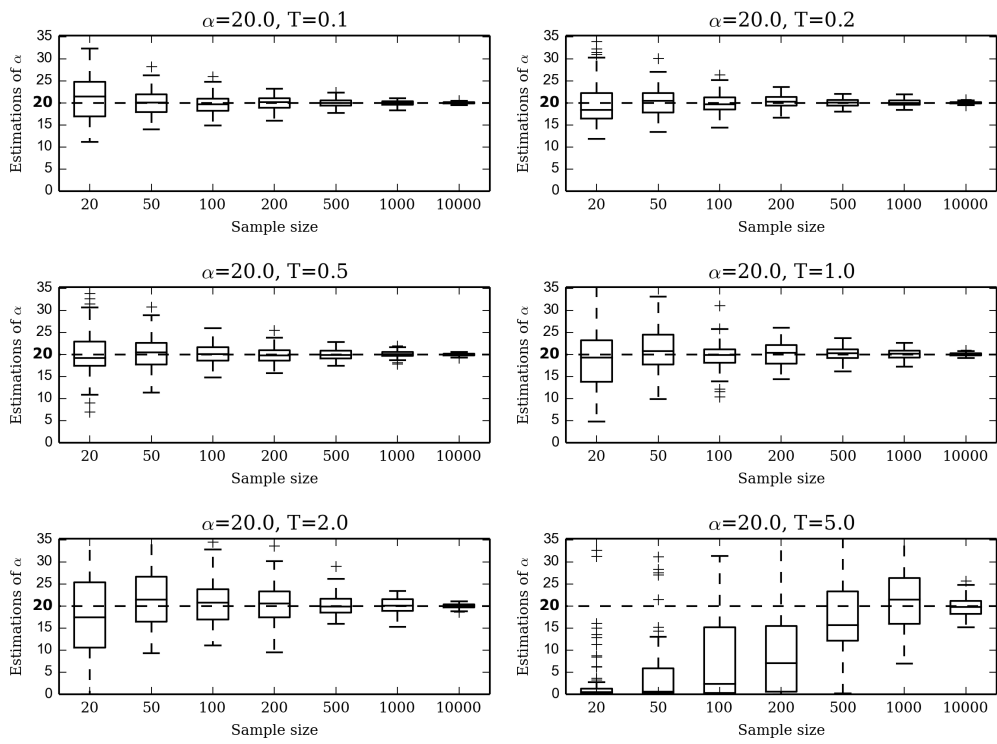


Figure 7: Estimations of α (real value $\alpha = 20.0$)

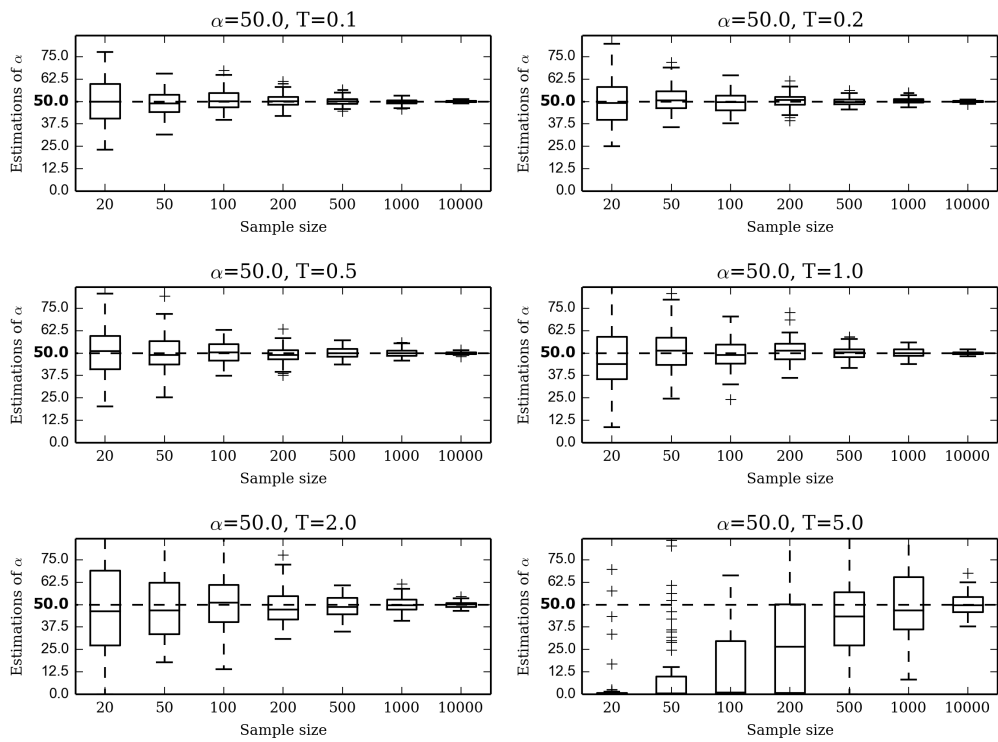


Figure 8: Estimations of α (real value $\alpha = 50.0$)

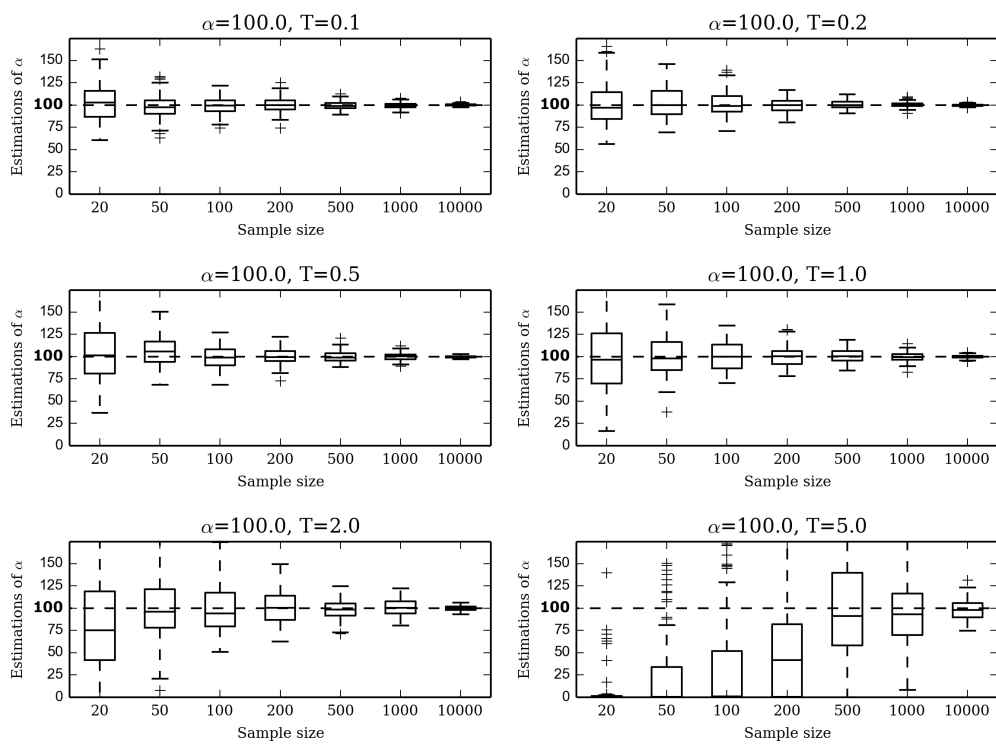


Figure 9: Estimations of α (real value $\alpha = 100.0$)

Figures 10, 11, 12, 13, 14 and 15 show the results of the estimation of T for all the values of α and all the sample sizes, under the SSPSC model. The values where $\alpha = (2, 4, 10, 20, 50, 100)$, $T = (0.1, 0.2, 0.5, 1, 2, 5)$ and $n_L = (20, 50, 100, 200, 500, 1000, 10000)$

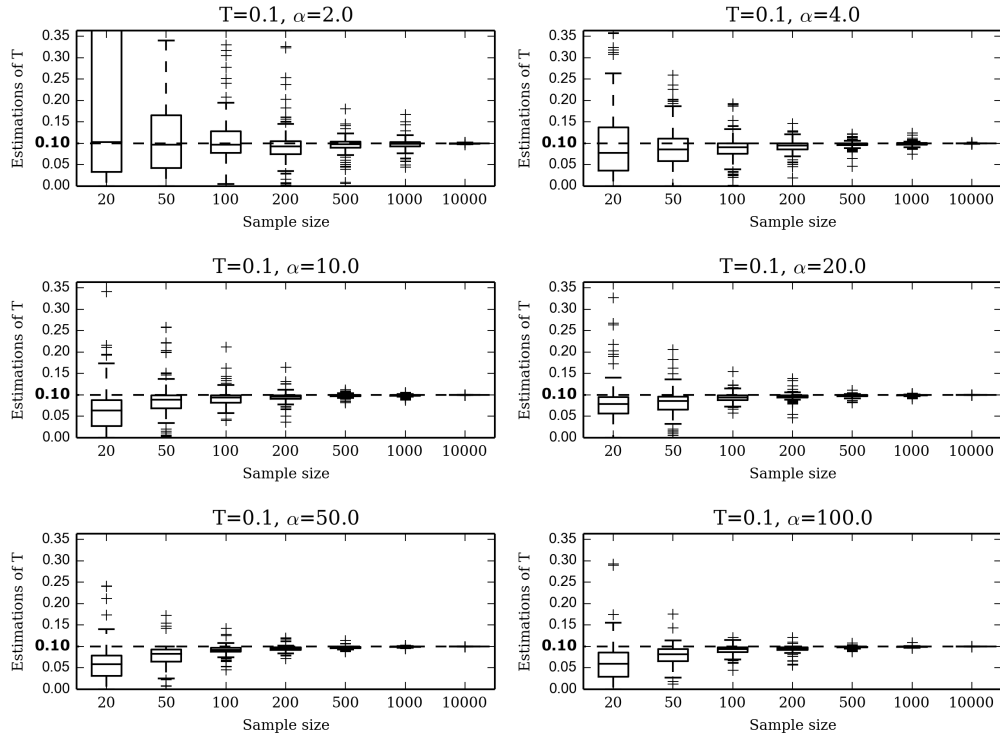


Figure 10: Estimations of T (real value $T = 0.1$)

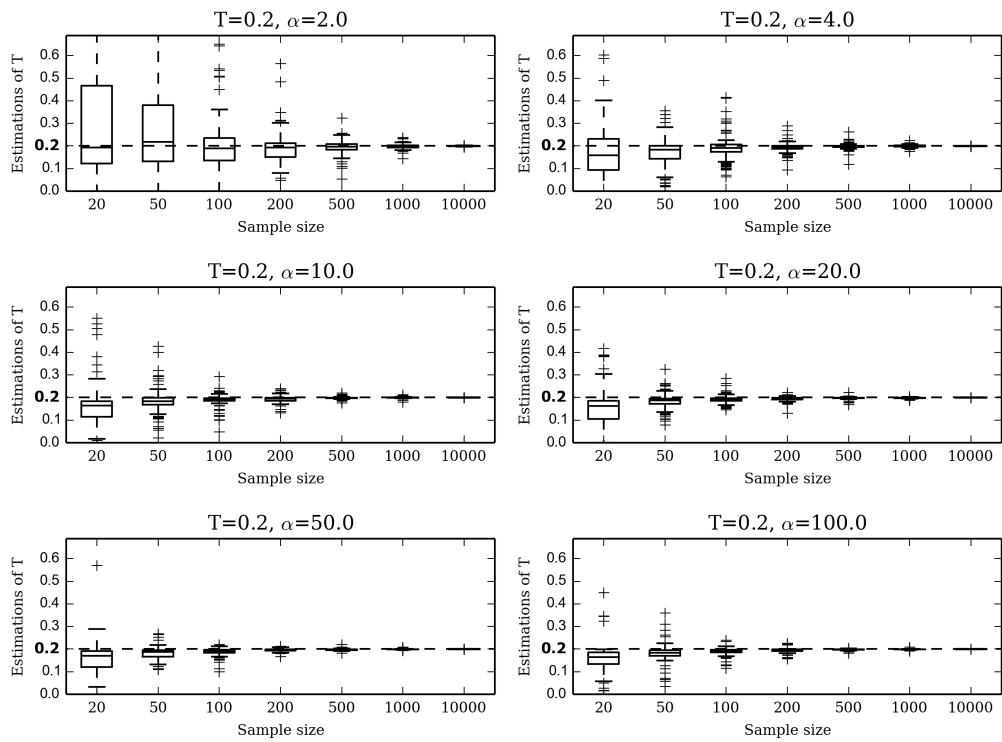


Figure 11: Estimations of T (real value $T = 0.2$)

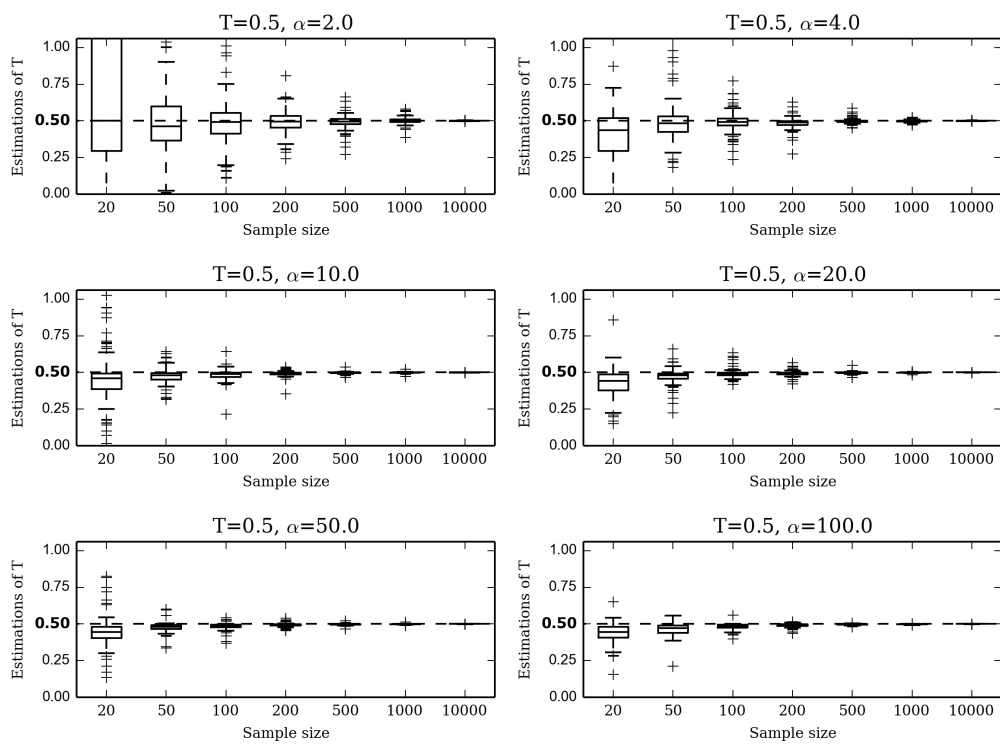


Figure 12: Estimations of T (real value $T = 0.5$)

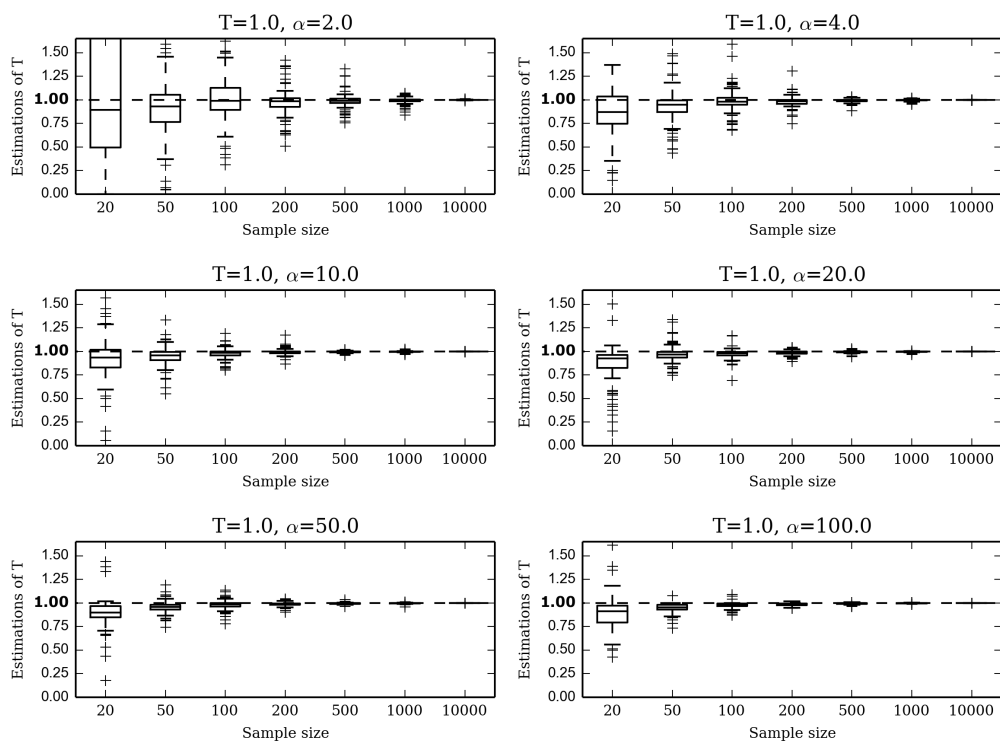


Figure 13: Estimations of T (real value $T = 1.0$)

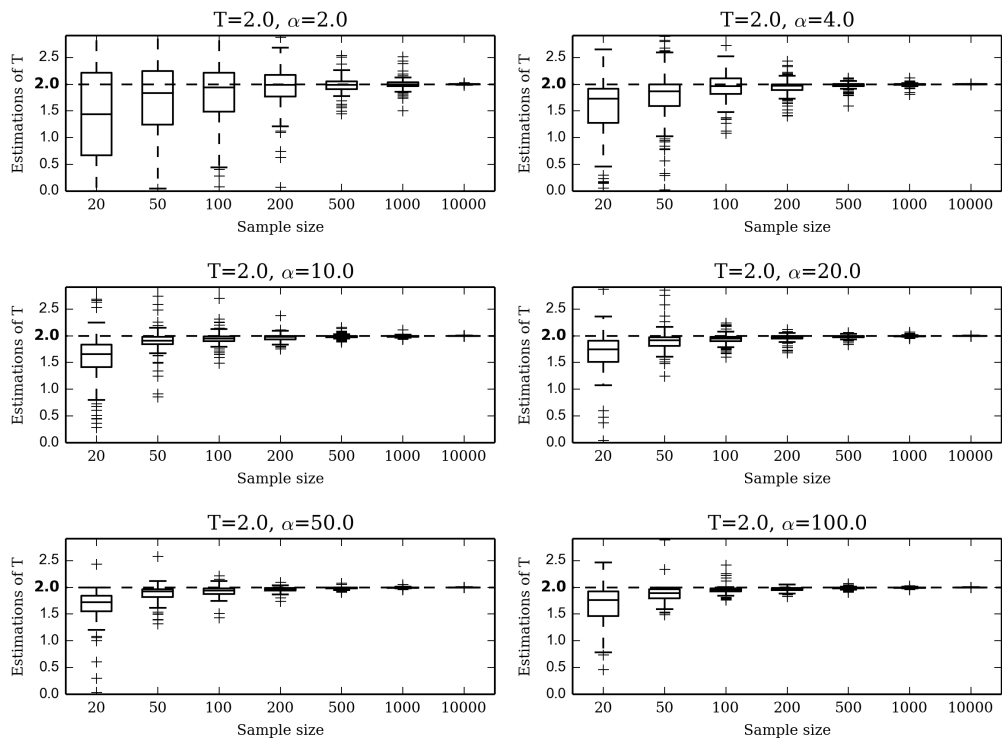


Figure 14: Estimations of T (real value $T = 2.0$)

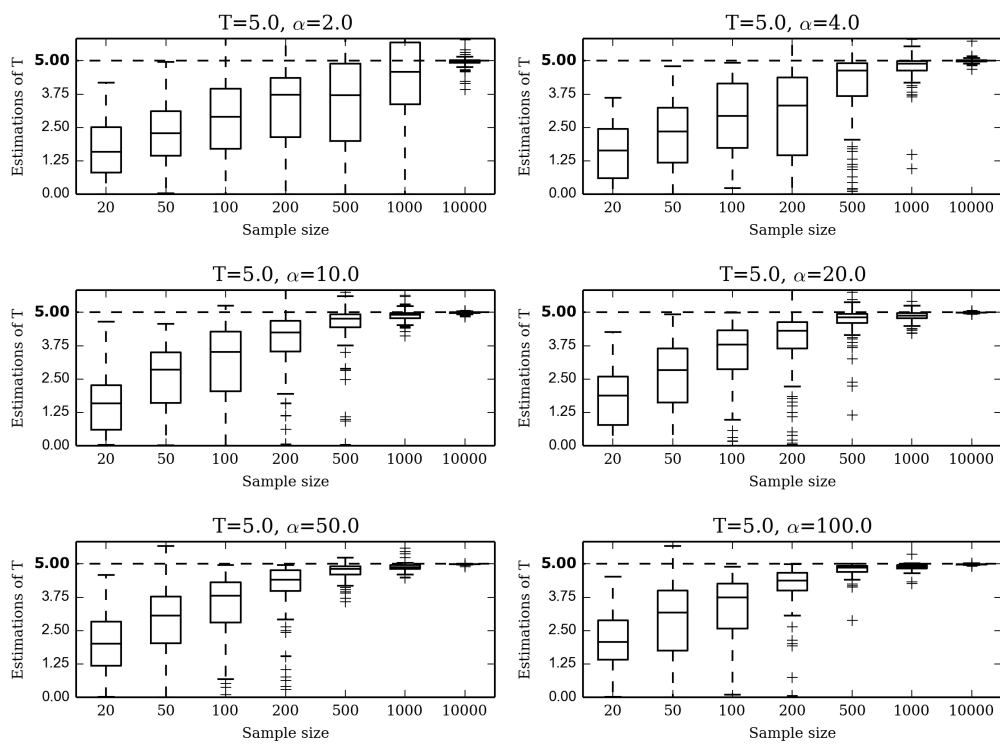


Figure 15: Estimations of T (real value $T = 5.0$)

Figures 16, 17, 18, 19, 21, 22, 23 and 24 show the results of the estimation of M for all the values of n and all the sample sizes, under the StSI model. The values where $M = (0.1, 0.2, 0.5, 1, 5, 10, 20, 50)$, $n = (2, 4, 10, 20, 50, 100)$ and $n_L = (20, 50, 100, 200, 500, 1000, 10000)$

$(0.1, 0.2, 0.5, 1, 5, 10, 20, 50)$ and $(2, 4, 10, 20, 50, 100)$

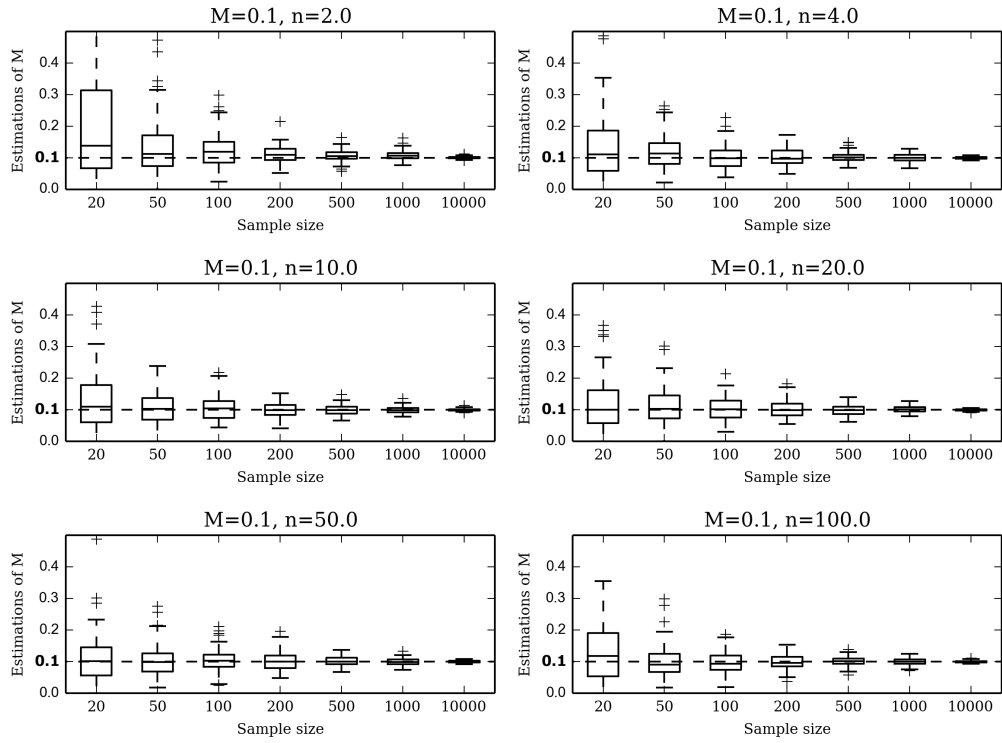


Figure 16: Estimations of M (real value $M = 0.1$)

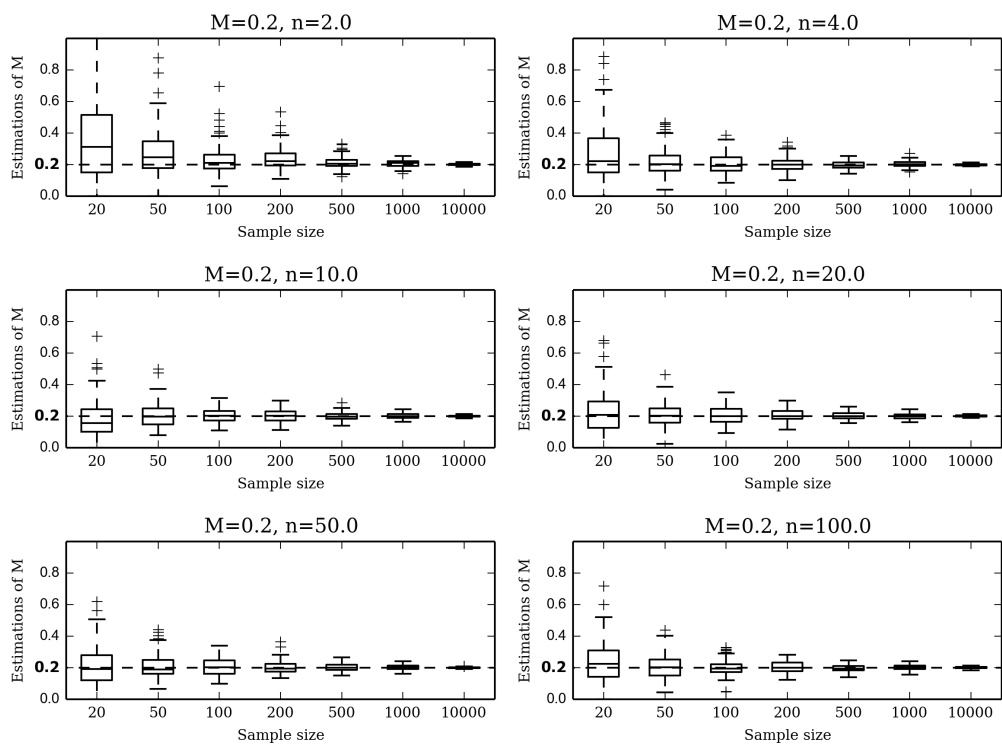


Figure 17: Estimations of M (real value $M = 0.2$)

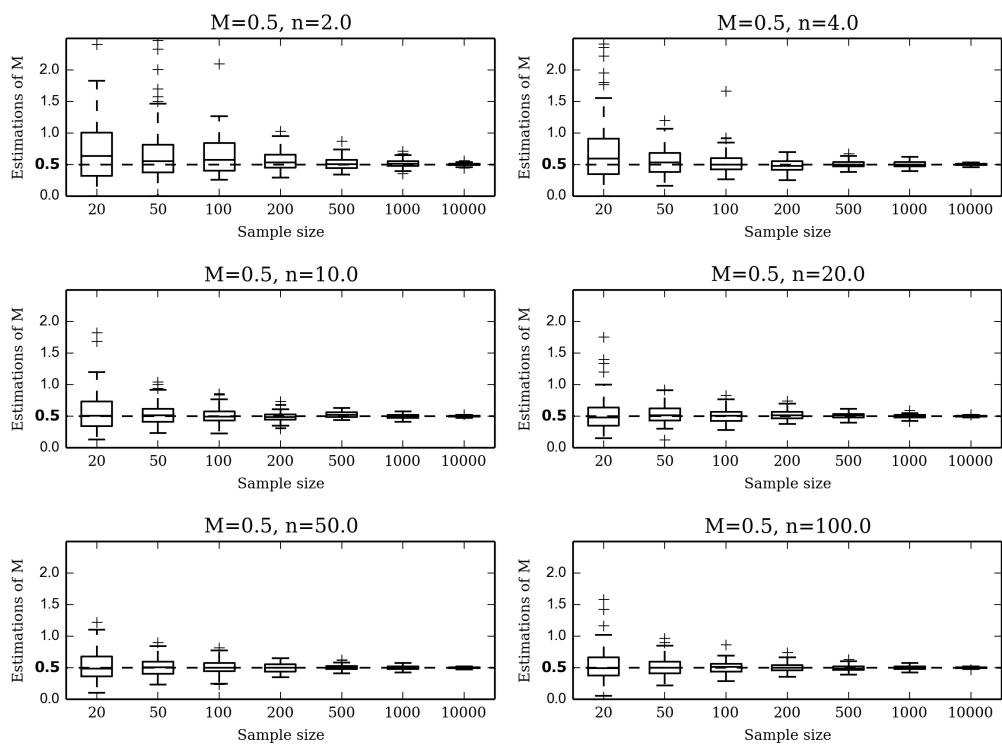


Figure 18: Estimations of M (real value $M = 0.5$)

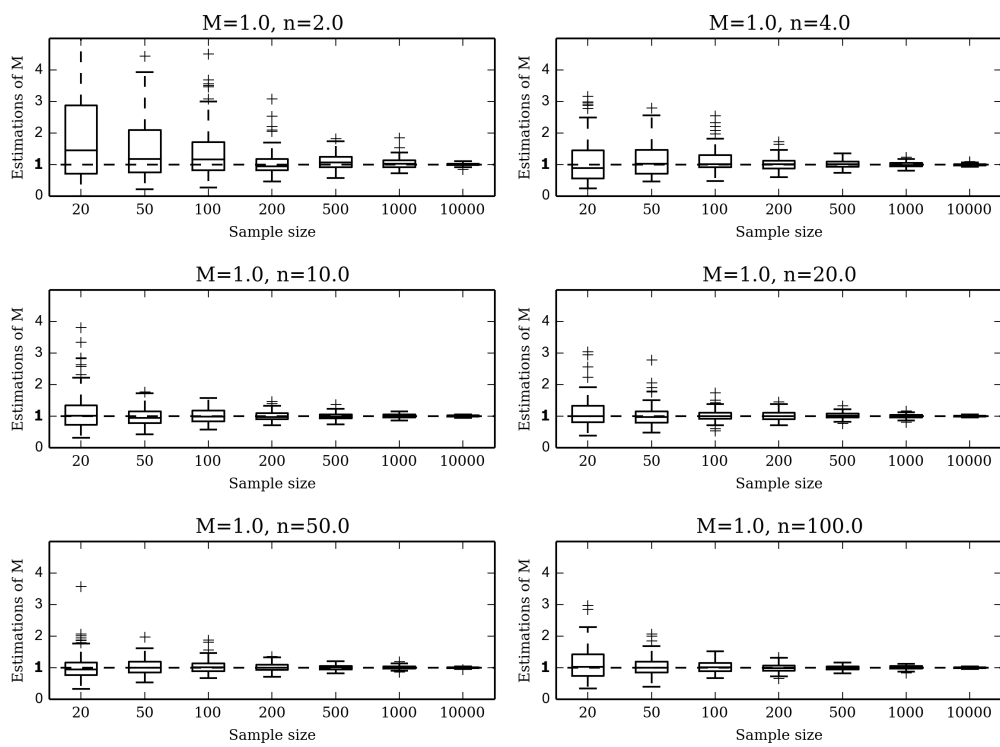


Figure 19: Estimations of M (real value $M = 1.0$)

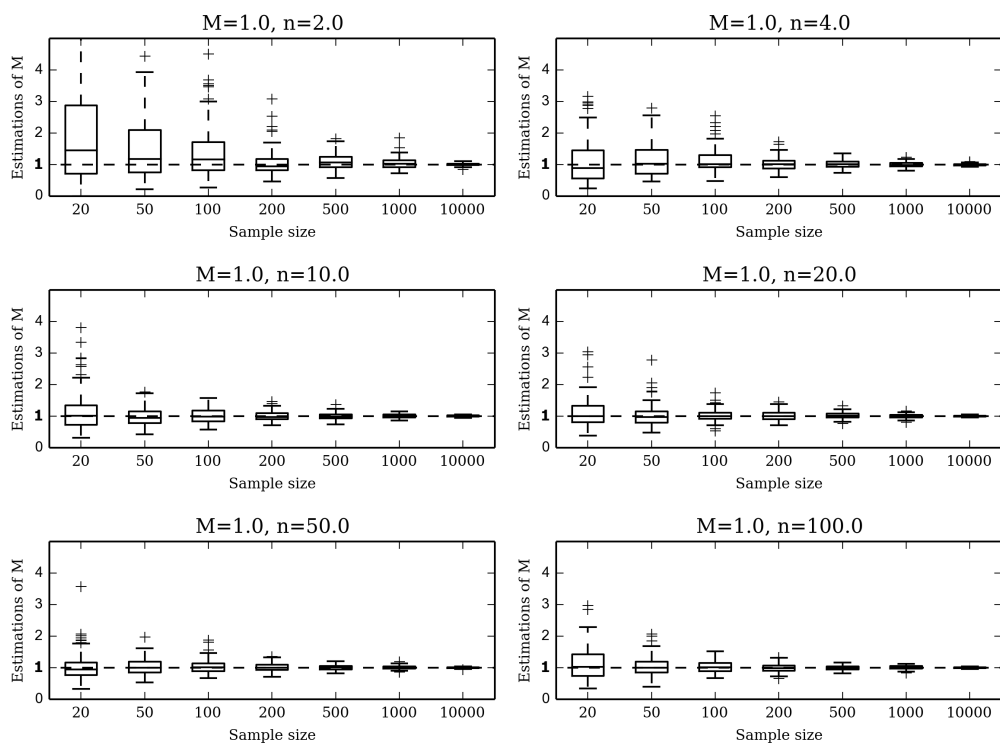


Figure 20: Estimations of M (real value $M = 2.0$)

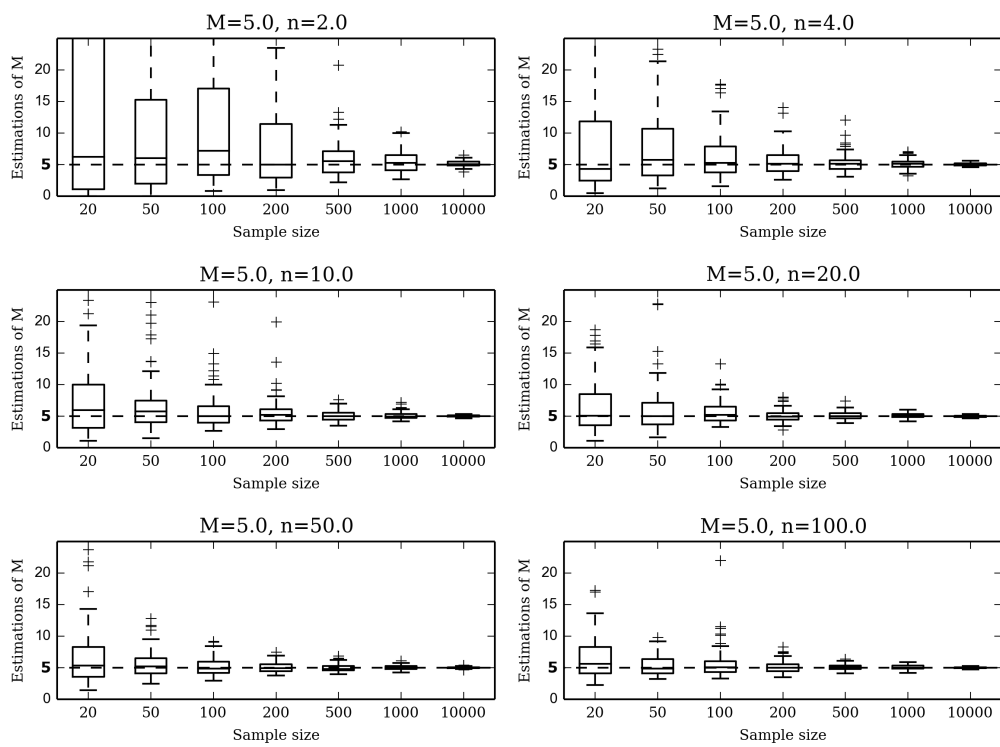


Figure 21: Estimations of M (real value $M = 5.0$)

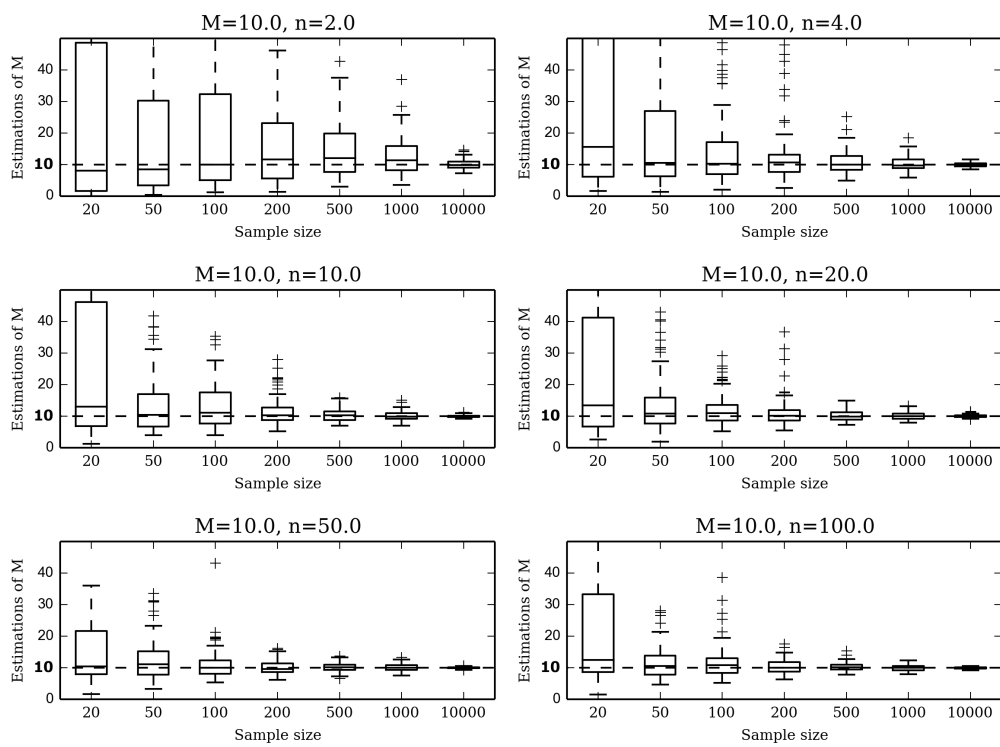


Figure 22: Estimations of M (real value $M = 10.0$)

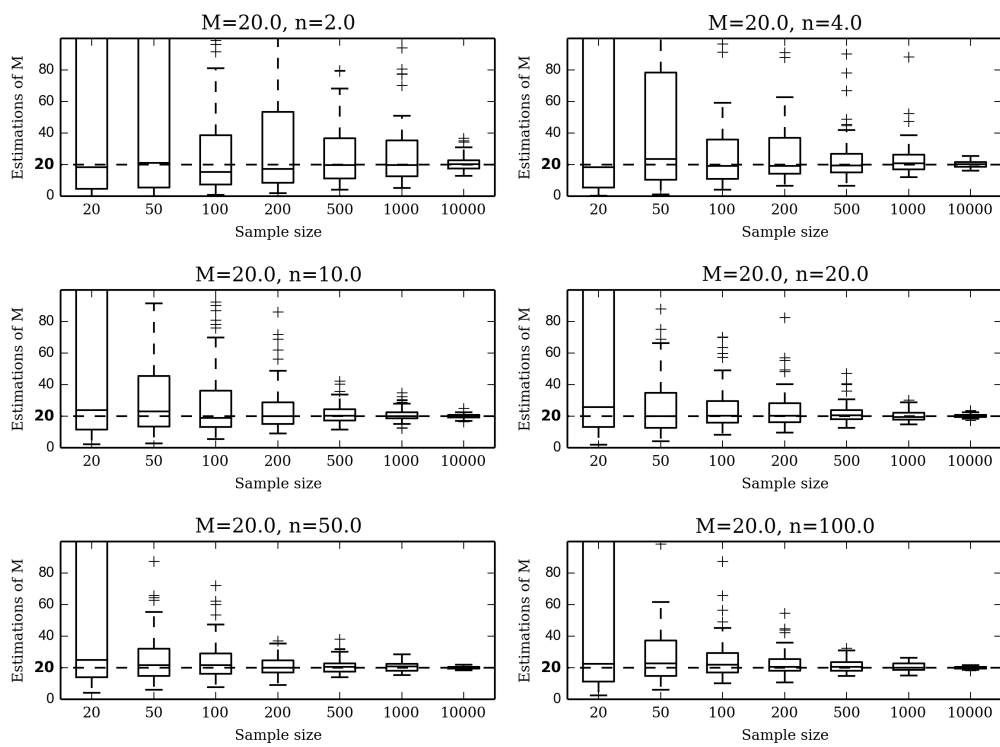


Figure 23: Estimations of M (real value $M = 20.0$)

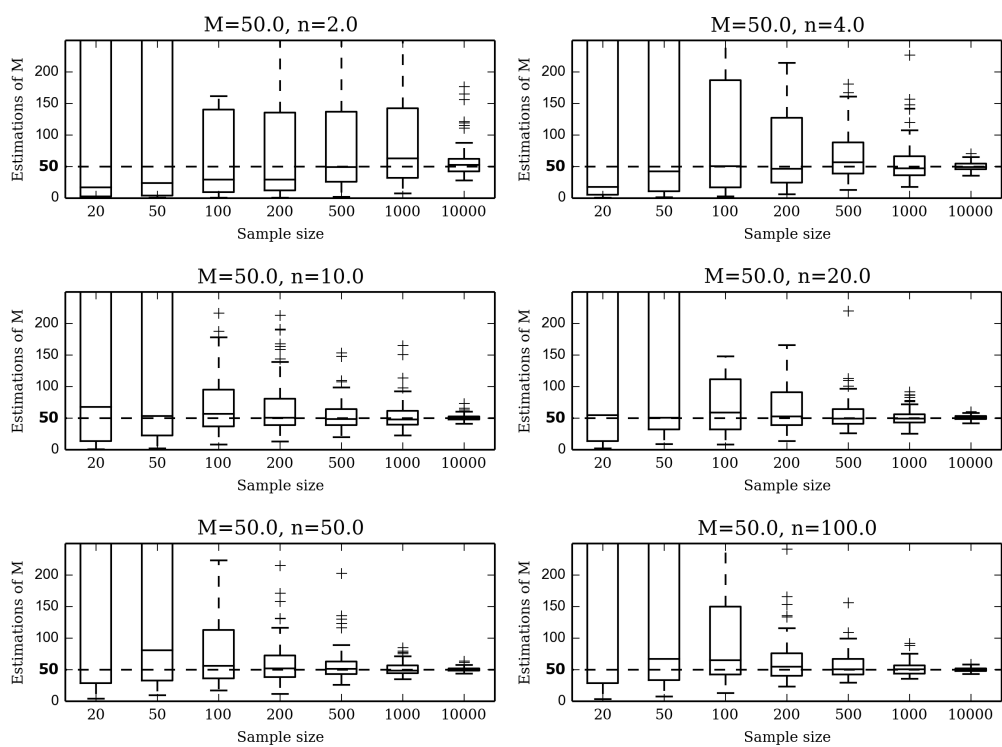


Figure 24: Estimations of M (real value $M = 50.0$)

Figures 25, 26, 27, 28, 29 and 30 show the results of the estimation of n for all the values of M and all the sample sizes, under the StSI model. The values where $M = (0.1, 0.2, 0.5, 1, 5, 10, 20, 50)$, $n = (2, 4, 10, 20, 50, 100)$ and $n_L = (20, 50, 100, 200, 500, 1000, 10000)$

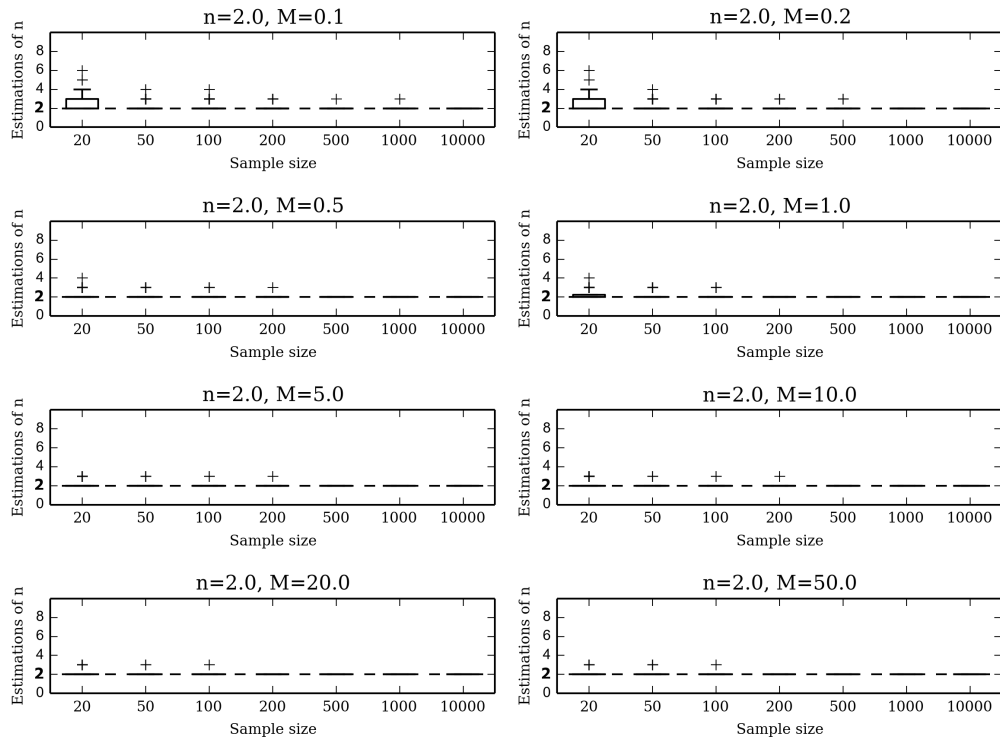


Figure 25: Estimations of n (real value $n = 2.0$)

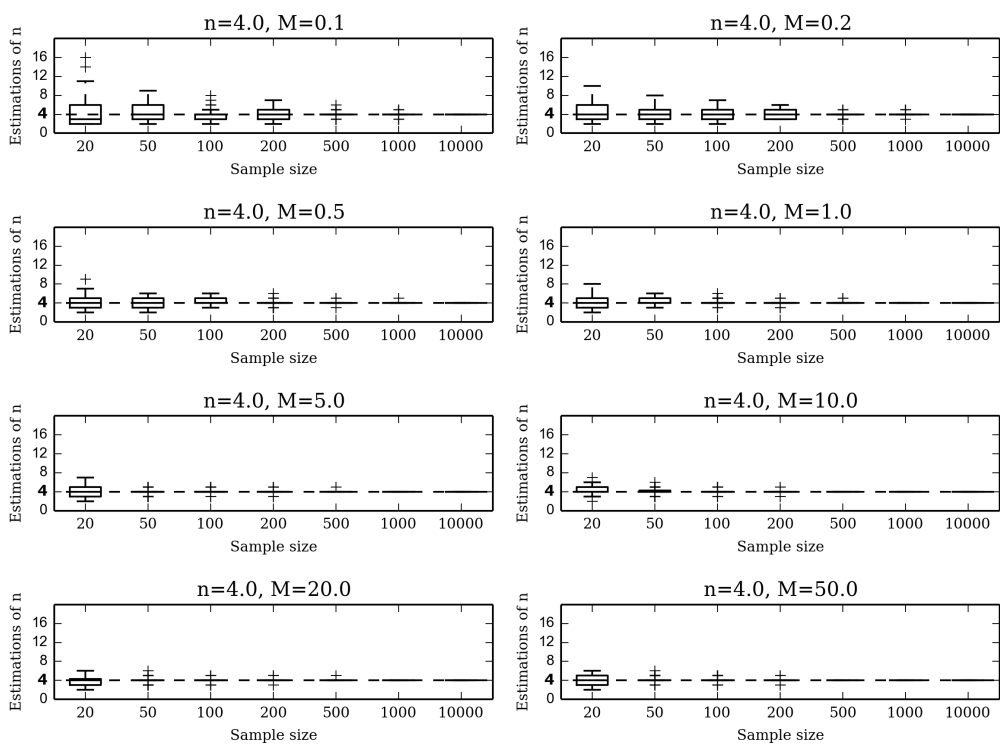


Figure 26: Estimations of n (real value $n = 4.0$)

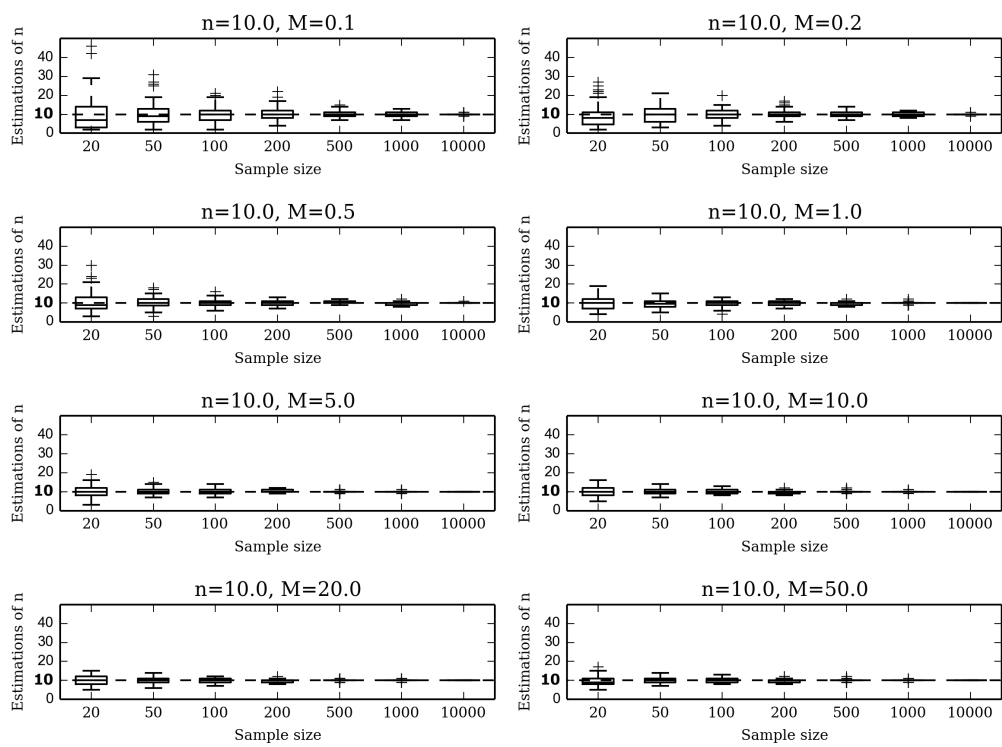


Figure 27: Estimations of n (real value $n = 10.0$)

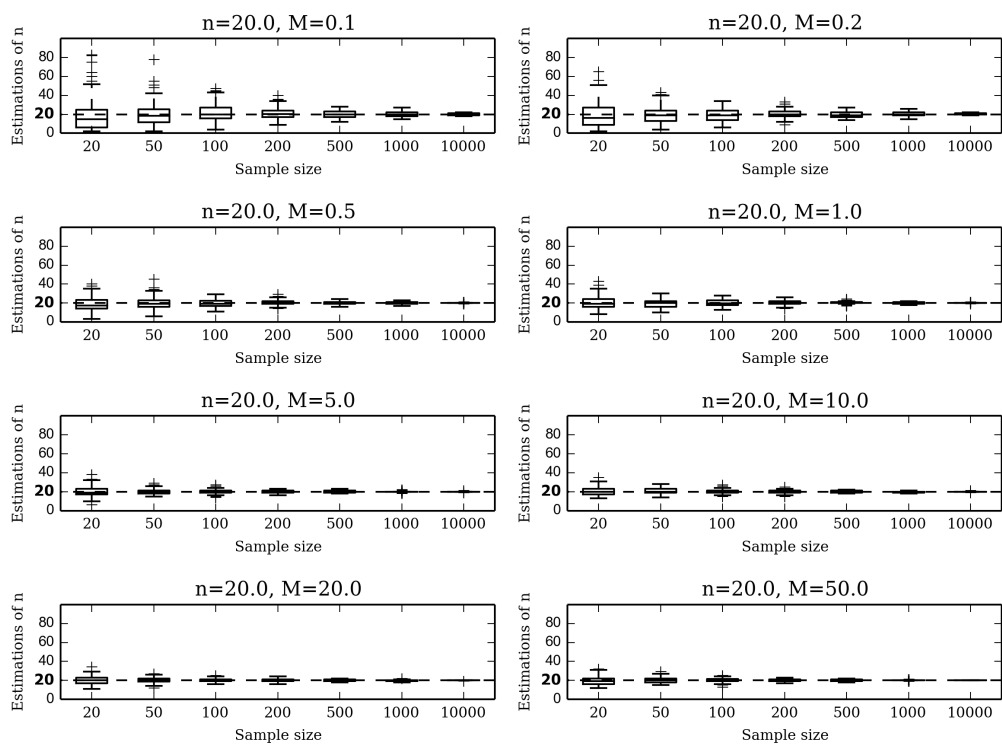


Figure 28: Estimations of n (real value $n = 20.0$)

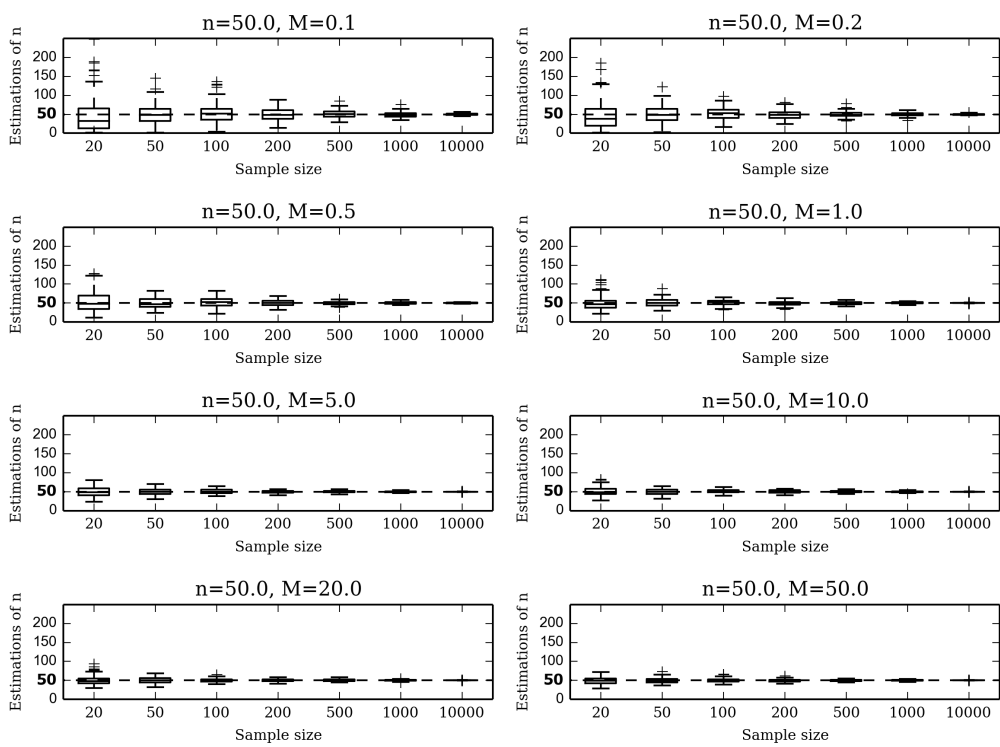


Figure 29: Estimations of n (real value $n = 50.0$)

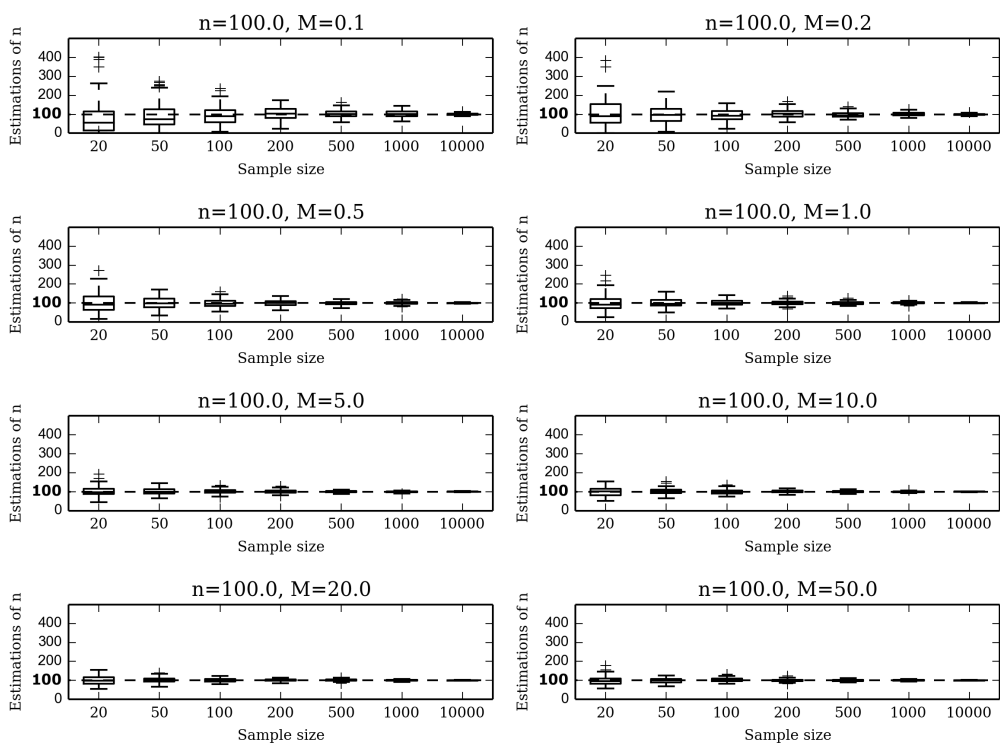


Figure 30: Estimations of n (real value $n = 100.0$)