SC/62/A10/MSYR2

A Revised Bayesian Meta-Analysis for Estimating a Posterior Distribution for the rate of increase for an "Unknown" Stock

ANDRÉ E. PUNT[1]

Contact e-mail: aepunt@uw.edu

1 - School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, WA 98195-5020, USA

ABSTRACT
An approach is outlined which can be used to construct a probability distribution for the rate of increase for an 'unknown' stock in the limit of zero population size, r_0. This approach is based on a beta distribution prior for the ratio of r_0 to the maximum demographically feasible rate of increase, r_{max}, and also accounts for environmental impacts on the population growth rate as well as uncertainty in the estimate of the realized growth rate. Estimation is based on Bayesian methods. Analyses based on simulation are conducted to evaluate the performance of this approach in data-rich and data-poor cases when estimates of r_0 are based on time-series of 20 years. As expected, performance is best when sample sizes are large and observation error is low.

KEYWORDS: MSYR, BAYESIAN META-ANALYSIS, SIMULATION

INTRODUCTION

The Scientific Committee of the International Whaling Commission is conducting a review of the range of MSYR values to include in simulation trials when selecting among variants of the Revised Management Procedure (RMP). A table of estimates of MSYR and rates of increase at low population size was developed by IWC (2009a) as a part of this review and revised by IWC (2009b). Punt (2009) outlined an approach for conducting a meta-analysis of rates of increase at low population size, r_0, for whales stocks based on the Bayesian paradigm. Punt and Allison (2010) applied this approach to the set of estimates of rates of increase identified by IWC (2009b) and developed a distribution for r_0 for an 'unknown' stock (i.e. not one of the stocks included in the meta-analysis).

Cooke (2010) noted that the results obtained by Punt and Allison (2010) were implausible because the prior for r_0 in the Bayesian analysis did not impose the lower bound that r_0 must exceed 0. In addition, Cooke (2010) noted that the approach of Punt (2009) did not account for the impact of environmental variation in the population growth rate. Cooke (2010) outlined a way to overcome both of the concerns raised with the approach of Punt (2009).

This paper outlines a revised method for conducting a Bayesian meta-analysis for estimating a posterior distribution for r_0 for the 'unknown' stock which imposes a lower bound of 0 for r_0 and accounts for the uncertainty caused by environmental variation in population growth rates. The revised method is then tested using simulated data.

METHODS

Estimating a posterior distribution for r_0 for an unknown stock

In the following $\hat{r}_i$ is the estimate of the rate of increase for stock i, and σ_i is the (estimate of) the observation error standard deviation for $\hat{r}_i$. Let us first define $r_{0,i}^{true}$ as the expectation of the rate of increase for stock *i* at low stock size and $r_{\max,i}$ as the maximum demographically possible rate of increase for stock *i* (assumed to be known exactly). Now, $r_{0,i}^{true} / r_{\max,i} = \chi_i$ is assumed to be beta-distributed, i.e. $\chi_i \sim Be(\alpha, \beta)$ [1]. Now let $\hat{r}_i$ be assumed to be distributed about a "realized" rate of increase subject to observation error, i.e. $\hat{r}_i \sim r_i^{real} + v_i$ where $v_i \sim N(0;\sigma_i^2)$. The realized rate of increase is related to true rate of increase, accounting for process uncertainty caused by environmental variation, i.e. the distribution of $(1+r_i^{real})^n$ is:

$$(1+r_i^{real})^{n_i} = \prod_{y=1}^{n_i} \exp\{r_{\max,i}(1-e^{-(\tau w_y - \tau^2/2)}(1-q_i)^z)\} \tag{1}$$

where n_i is the number of data points for stock *i*, $w_y = \rho w_{y-1} + \sqrt{1-\rho^2}\varepsilon_y$, $\varepsilon_y \sim N(0;1)$, ρ is the extent of auto-correlation in the environmental impact on *r*, and τ is the standard deviation for the environmental impact on *r*.

Now, given $q_i = (1-(1-\chi_i)^{1/z})$, *z*, ρ, and τ (assumed known) one can generate a distribution for $(1+r_i^{real})^{n_i}$ numerically. For estimation purposes, the mean of r_i^{real} can be approximated using the formula:

$$E(r_i^{real}) = \alpha_1 R_{\max,i} + \alpha_2 \chi_i + \alpha_3 (R_{\max,i})^2 + \alpha_4 (\chi_i)^2 + \alpha_5 R_{\max,i}\,\chi_i + \alpha_6 (R_{\max,i})^2 \chi_i + \alpha_7 (\chi_i)^2 R_{\max,i}$$

The standard deviation of r_i^{real}, $Var(r_i^{real})$ is approximated similarly.

The likelihood function is then:

$$L(D\,|\,\alpha,\beta) = \prod_i \int_0^1 \frac{\Gamma(\alpha,\beta)}{\Gamma(\alpha)\Gamma(\beta)} \chi_i^{\alpha-1}(1-\chi_i)^{\beta-1} \frac{1}{\sqrt{2\pi}\tilde{\sigma}_i} e^{-[\hat{r}_i - E(r_i^{real})]^2/(2\tilde{\sigma}_i^2)} d\chi_i \tag{2}$$

where $\tilde{\sigma}_i^2 = \sigma_i^2 + Var(r_i^{real})$

The integrals in Equation 2 are evaluated numerically (in this case using a application of the trapezoidinal rule with 100 steps). The priors for α and β are assumed to be uniform, i.e. U[0,10].

Simulation evaluation

The performance of the estimator outlined above is evaluated by generating simulated data sets. For computational ease, n_i is set to 20 for all stocks, $r_{\max}$=0.1, and *z*=2.39. Table 1 lists the values for the number of stocks for which data are generated, the extent of environmental variation, the extent of autocorrelation in environmental process error, the standard deviation of the observation error (assumed to be the same for all stocks), and true values for α and β. As noted above, the standard deviation of the extent of observation error, the extent of environmental process error, the extent of autocorrelation in environmental process error, and *z* are all assumed to be known exactly. The values for the parameters which define the simulation scenarios are

[1] The beta distribution is selected here because it provides a flexible way to model bounded random variables.

deliberately chosen so that whether the method performs adequately in principle can be tested. However, cases 5 and 10-12 explore scenarios in which the sample size is more representative of that be expected in reality.

Figure 1 shows the distributions for $r_0^{true} / r_{\max}$ for each of the three choices for α and β in Table 1, while Figure 2 shows some example trajectories for population size for the reference case simulations (case 1 in Table 1) as well as for case 10 for which environmental process error is set to 1. The performance of the estimation method is summarized by the relative bias and relative root-mean-square-error (RMSE) for the lower 1st, 2nd, 5th, 10th, 25th and 50th percentiles for the distribution for $r_0^{true} / r_{\max}$ for an unknown stock. Table 2 lists the "true" values for these percentiles for the three choices for α and β.

RESULTS AND DISCUSSION

The method performs well (low bias, < 2.1% and RMSE, < 11%) for all percentiles for the most informative cases (cases 1-3 in Tables 3-4 and Figure 3). The distributions of the estimates of the lower percentiles are, however, not symmetric (Figure 3).

As expected, the performance of the method deteriorates as the number of stocks included in the analysis decreases from 1,000 to 100 to 20 (cases 1, 4 and 5 in Tables 3 and 4). The estimates for case 5 are both biased and fairly imprecise (Figure 4). The impact of high process error (case 7) is to increase both bias and imprecision. Increasing the extent of observation error from a standard deviation of 0.011 to 0.1 (an almost 100-fold increase in variance) (cases 8-10) leads to markedly higher RMSE values and also to bias (Figure 5). The bias for case 10 is, however, generally negative (Figure 5) suggesting that if the method is applied when the extent of environmental process error is high, the autocorrelation in process error is high and the number of stocks is low, the method is "conservative".

The results in Tables 3-4 and Figures 3-5 are based on the assumption that the extent of environmental process error and the autocorrelation in environmental process error is known. The impact of violation of this assumption has not been evaluated quantitatively. However, this impact would logically lead to poorer estimation performance.

ACKNOWLEDGEMENTS

Justin Cooke is acknowledged for providing much of the thinking which led to the algorithm evaluated in this paper. NOAA/NMFS/NMML are thanked for providing funding support through grant number AB133F-09-SE-4556.

REFERENCES

Cooke, J.G. 2010. Towards a plausible range of MSYR values: a suggested way forward. Appendix 5 to the Report of the Sub–committee on the Revised Management Procedure. *J. Cet. Manage. Res*. 00: 00-00.

IWC. 2009a. Report of the MSYR Workshop. Document SC/60/Rep5 presented to the IWC Scientific Committee, June 2008. (unpublished). 16pp.

IWC. 2009b. Report of the MSYR Workshop. Document SC/61/Rep6 presented to the IWC Scientific Committee, June 2009. (unpublished). xxpp.

Punt, A.E. 2009. Constructing a posterior distribution for the rate of increase of whale stocks at low population size. IWC Document SC/F09/MSYR1 (7pp).

Punt, A.E. and C. Allison. 2010. Revised outcomes from the Bayesian meta-analysis. Appendix 2 to the Report of the Sub–committee on the Revised Management Procedure. *J. Cet. Manage. Res*. 00: 00-00.

Table 1
The values for the parameters which define the simulation experiments

Case No	N_{stock}	τ	ρ	σ	α	β
1	1000	0.2	0	0.011	3	3
2	1000	0.2	0	0.011	2	4
3	1000	0.2	0	0.011	4	2
4	100	0.2	0	0.011	3	3
5	20	0.2	0	0.011	3	3
6	1000	1	0	0.011	3	3
7	1000	0.2	0.7	0.011	3	3
8	1000	0.2	0	0.1	3	3
9	1000	1	0.7	0.1	3	3
10	20	1	0.7	0.1	3	3
11	20	0	0	0.011	3	3
12	20	0	0	0.05	3	3

Table 2
The true percentiles for the $r_0^{true} / r_{\max}$ for the three choices for α and β

α, β	Percentiles					
	0.01	**0.02**	**0.05**	**0.1**	**0.25**	**0.5**
3, 3	0.105	0.135	0.189	0.247	0.360	0.500
2, 4	0.033	0.047	0.076	0.112	0.194	0.314
4, 2	0.223	0.269	0.344	0.417	0.547	0.686

Table 3
Biases (expressed relative to the true percentile)

Case No	Percentiles					
	0.01	**0.02**	**0.05**	**0.1**	**0.25**	**0.5**
1	0.019	0.020	0.012	0.002	0.001	0.001
2	0.003	-0.002	0.018	0.007	0.005	0.000
3	0.005	0.021	0.015	0.005	0.006	0.002
4	0.147	0.121	0.097	0.069	0.036	0.003
5	0.110	0.125	0.129	0.095	0.041	-0.011
6	0.090	0.089	0.056	0.041	0.019	0.004
7	0.021	0.012	0.004	-0.001	0.003	0.000
8	0.362	0.363	0.316	0.238	0.104	0.001
9	0.232	0.273	0.247	0.198	0.091	0.001
10	-0.541	-0.412	-0.243	-0.140	-0.054	-0.017
11	0.121	0.122	0.106	0.084	0.037	-0.012
12	-0.114	-0.020	0.038	0.043	0.017	-0.023

Table 3
Root mean square errors (expressed relative to the true percentile)

Case No	**Percentiles**					
	0.01	**0.02**	**0.05**	**0.1**	**0.25**	**0.5**
1	0.078	0.066	0.046	0.031	0.013	0.003
2	0.105	0.086	0.065	0.047	0.025	0.011
3	0.067	0.055	0.041	0.029	0.016	0.008
4	0.311	0.258	0.201	0.151	0.093	0.056
5	0.471	0.406	0.316	0.236	0.143	0.094
6	0.157	0.140	0.096	0.071	0.036	0.014
7	0.079	0.064	0.045	0.031	0.015	0.004
8	0.486	0.444	0.363	0.272	0.136	0.063
9	0.480	0.440	0.339	0.255	0.132	0.064
10	0.664	0.579	0.470	0.391	0.286	0.208
11	0.489	0.408	0.305	0.231	0.140	0.094
12	0.489	0.446	0.380	0.323	0.235	0.173

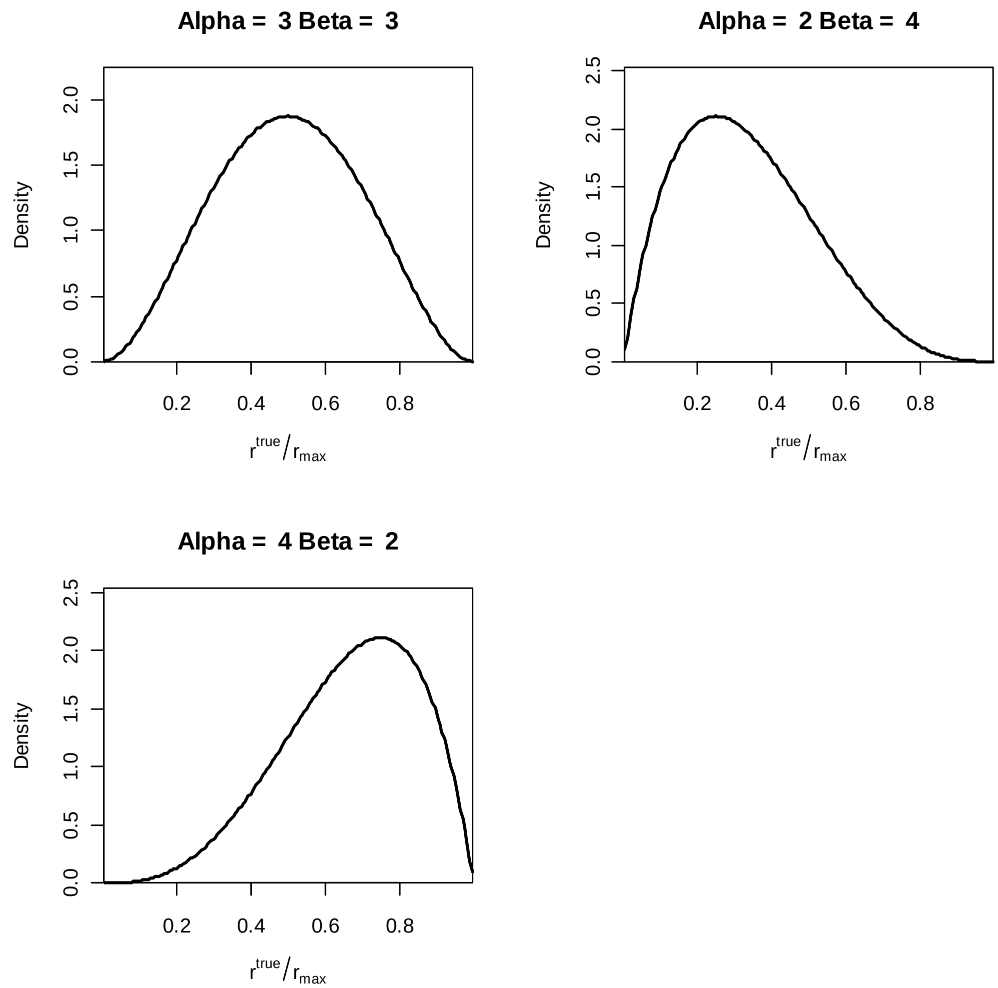
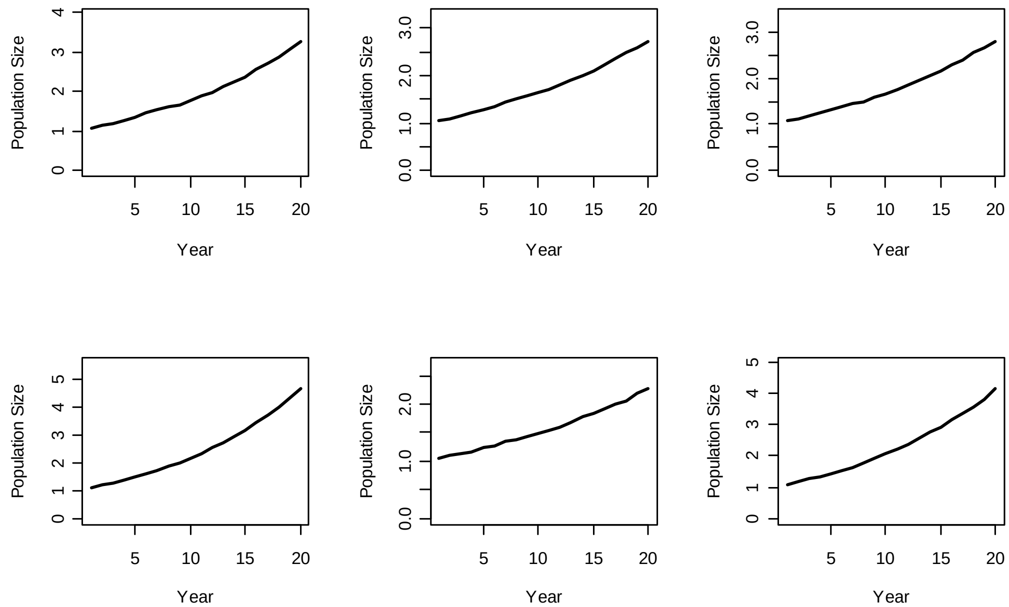


Figure 1. Distributions for $r_0^{true} / r_{\max}$ for the three choices for α and β considered in the simulation study.

(a) Case 1 (τ=0.2; ρ=0)



(b) Case 10 (τ=1; ρ=0.7)

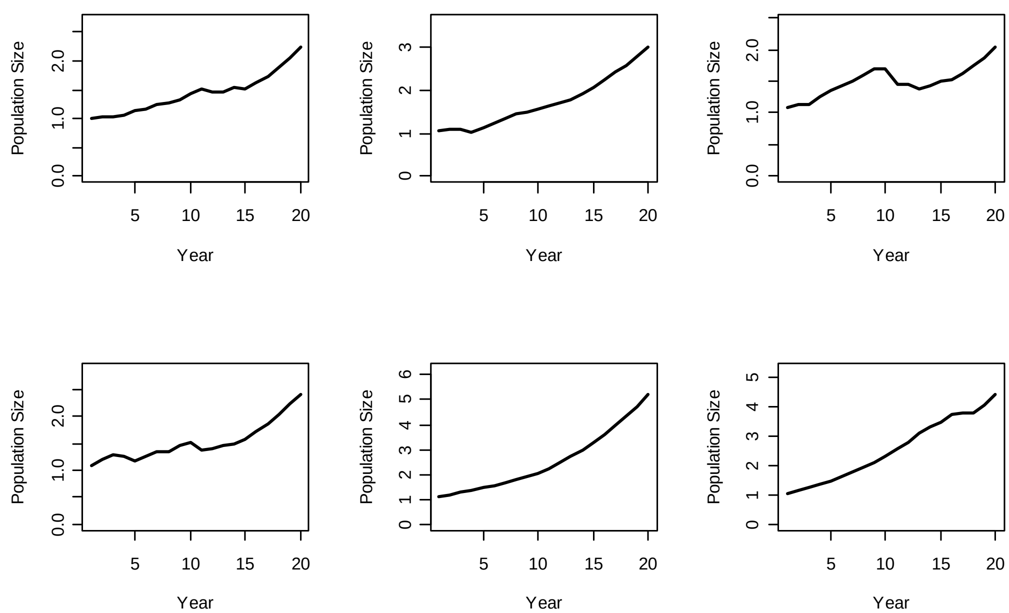


Figure 2. Example time-trajectories of population size for cases 1 and 10.

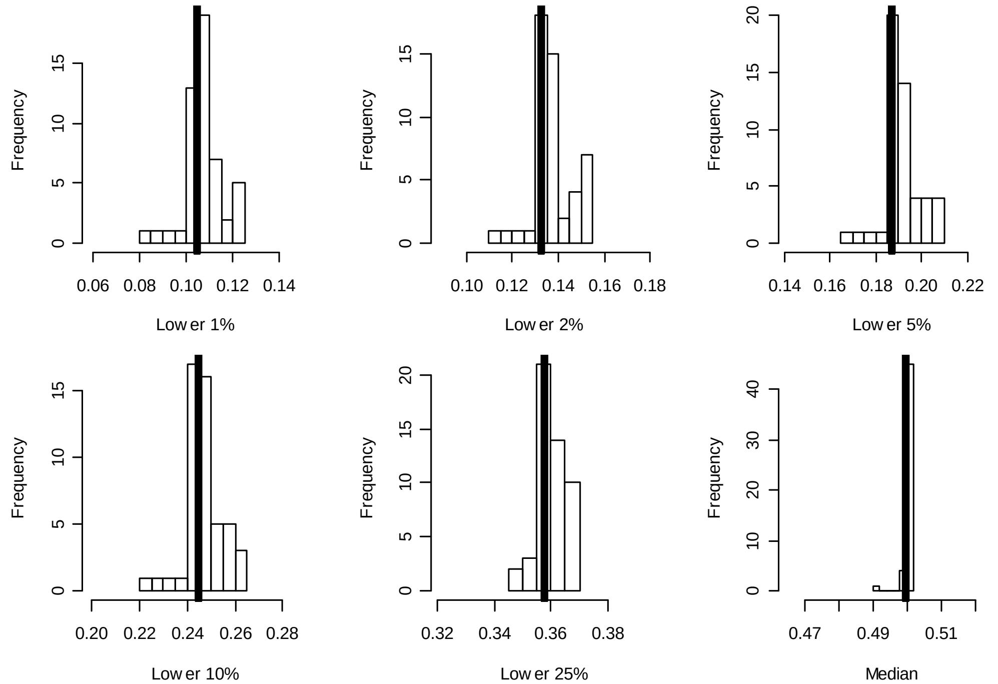


Figure 3. Estimates of the lower percentiles of the distribution for $r_0^{true} / r_{\max}$ for case 1. The solid line denotes the true percentile.

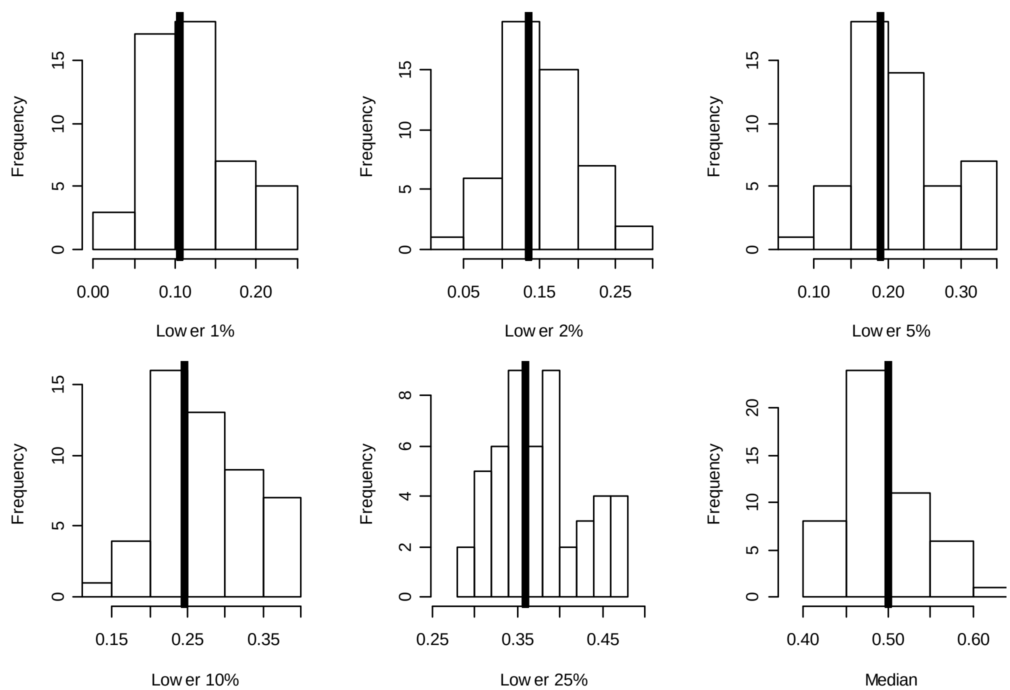


Figure 4. Estimates of the lower percentiles of the distribution for $r_0^{true} / r_{\max}$ for case 5. The solid line denotes the true percentile.

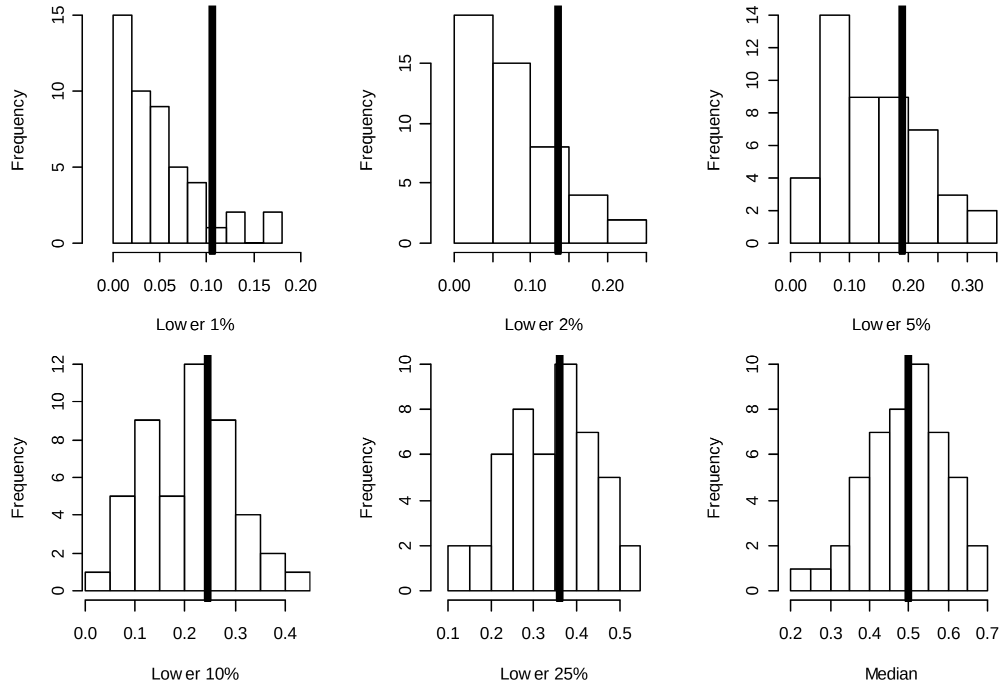


Figure 5. Estimates of the lower percentiles of the distribution for $r_0^{true} / r_{\max}$ for case 10. The solid line denotes the true percentile.