Empirical Bayes Methods for Discrete Event Simulation Performance Measure Estimation*

Shona Blair[1], Tim Bedford[1], and John Quigley[1]

1 **Dept. of Management Science, University of Strathclyde**
40 George Street, Glasgow, G1 1QE, UK
{shona.blair, tim.bedford, j.quigley}@strath.ac.uk

Abstract

Discrete event simulation (DES) is a widely-used operational research methodology facilitating the analysis of complex real-world systems. Although, generally speaking, simplicity is greatly desirable in DES modelling applications, in many cases the nature of the underlying system results in simulation models which are large in scale, complex, and expensive to run. As such, the careful design and analysis of simulation experiments is essential to ensure valid and efficient inference concerning DES model performance measures. It is envisaged that empirical Bayes (EB) methods, which enable data to be pooled across a set of populations to support inference of the parameters of a single population, may be of use within this context. Despite this potential, EB has so far been neglected within the DES literature. This paper presents a preliminary computational investigation into the efficacy of EB procedures in the estimation of DES performance measures. The results of this investigation, and their significance, are explored. Additionally, likely directions for future research are also addressed.

1998 ACM Subject Classification I.6.8 Types of Simulation

Keywords and phrases Discrete Event Simulation, Analysis Methodology, Empirical Bayes

Digital Object Identifier 10.4230/OASIcs.SCOR.2012.21

1 Introduction and Motivation

Discrete event simulation (DES) is a powerful and flexible methodology, widely utilized in OR applications for the design, analysis and improvement of complex, dynamic and stochastic real-world systems. At its core, DES involves abstracting the fundamental structure of the system of interest and using this information to construct a computer model of the system. A process of experimentation is conducted with the computer model in order to gain insight into and understanding of the performance of the real system. One of the key advantages of discrete event simulation is its ability to incorporate a "realistic" level of system complexity into the analysis process, when compared with the more rigid assumptions of alternative modeling techniques; indeed, DES is frequently referred to as a "method of last resort" [9]. Whilst the benefits of simple models are well understood and widely disseminated (see, for example, [11, 13]), there are many instances when the nature of the underlying system being studied necessitates the use of simulation models which are large-scale, structurally complex, difficult to interpret and computationally expensive to run. As such, the careful design and analysis of simulation experiments is necessary to ensure valid and efficient inference concerning model performance [8].

* This work was partially supported by an EPSRC CASE Studentship and SIMUL8 Corporation.

3rd Student Conference on Operational Research (SCOR 2012).
Editors: Stefan Ravizza and Penny Holborn; pp. 21–30
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Empirical Bayes (EB) procedures offer a structured and theoretically sound framework for the pooling of data obtained across a set of populations to support inference concerning the parameters of an individual population. This often enables more efficient inference in situations which feature a repeated structure, providing that sufficient "similarity" exists between component populations. (For a general EB reference, see [2]). It seems intuitively reasonable that such an approach may be of benefit in simulation model experimentation, owing to the underlying similarity between simulation model configurations. In light of the computational expense involved in executing simulation models (touched on above), such increased efficiency in estimation would likely prove highly advantageous in practice. Yet, in spite of this apparent potential, EB has so far been neglected within the simulation literature.

This paper presents the results of a preliminary computational investigation into the use of EB procedures in the estimation of DES model performance measures. It begins, in Section 2, with the presentation and brief derivation of the EB procedures which are to be applied. Then, in Section 3, the DES model selected for testing is introduced, the reasons behind this choice outlined and certain theoretical results regarding performance measures of interest are presented. After this, in Section 4, the experimental design of the study is described, before the results are summarised in Section 5. The paper concludes, in Section 6, with some discussion of how this research area might be further explored in the future.

2 Introducing the Empirical Bayes Procedures

Empirical Bayes procedures feature a hierarchical model structure, identical to that of a traditional Bayesian analysis. As such, we typically have a situation in which model parameters are themselves represented by probability distributions, termed "prior" distributions. In a Bayesian analysis, the prior distribution would be subjectively determined, usually elicited from subject matter experts. However, in an empirical Bayes setting the data themselves determine the prior distribution. As mentioned in the previous section, EB methods are well-suited to applications featuring a large number of "similar" populations or processes. In this situation, the data obtained from each of the populations are pooled and used to provide inference on a general prior distribution. This general prior distribution is then combined with the individual samples from each of the populations using standard Bayesian updating to obtain a "posterior" distribution specific to each of the populations. (For a detailed overview of the above theory and terminology, please refer to [2].)

EB methods have a long history, with their roots in actuarial work on credibility theory, the first major publication by Robbins [12] in 1955 and a series of landmark papers in the 1970's by Efron and Morris [4, 5, 6] (see [10] for a more detailed account of their development). However, recent years have seen a huge upsurge in the volume of EB publications. This is due predominantly to scientific advances such as microarray technology, facilitating high-throughput biological screening and generating massive datasets that demand a fresh approach to statistical analysis [3]. A frequent feature of such datasets is their large number of populations, contrasted with relatively few observations from each. Such structures, as mentioned before, are ideally suited to an empirical Bayes analysis. Indeed, many successful applications have been published; a recent survey being [1]. Not surprisingly, this renewed interest has led to methodological and theoretical developments, as well as applications.

Here, however, we focus on some specific results which show particular promise in terms of their potential applicability within the context of DES model analysis. The empirical Bayes estimator shall employ is the "double-shrinkage" estimator presented in article [15]. This estimator assumes a normal/lognormal model of the data and its derivation is presented in the following subsections.

2.1 Assumptions

For $i = 1, 2, \ldots, p$, we assume:

$$X_i|\theta_i, \sigma_i^2 \overset{ind}{\sim} N(\theta_i, \sigma_i^2) \quad (1)$$

$$\theta_i \overset{iid}{\sim} N(\mu, \tau^2) \quad (2)$$

$$\log \sigma_i^2 \overset{iid}{\sim} N(\mu_v, \tau_v^2) \quad (3)$$

$$\log(S_i^2/\sigma_i^2) \overset{iid}{\sim} N(m, \sigma_{ch}^2) \quad (4)$$

where $m = E[\log(\chi_d^2)] = \psi(\frac{d}{2}) - \log(\frac{d}{2})$ and $\sigma_{ch}^2 = Var[\log(\chi_d^2)] = \psi'(\frac{d}{2})$, with d denoting the degrees of freedom. Initially, we suppose that hyperparameters μ, τ^2, μ_v and τ_v^2 are known.

2.2 Derivation

Given a sample of n observations, x_i, from sampling distribution (1) for population i, the prior distribution (2) on θ_i, and, for the moment, the additional assumption that σ_i^2 is known, a standard application of Bayes rule yields:

$$\theta_i|x_i, \sigma_i^2 \sim N(M_i\bar{x}_i + (1 - M_i)\mu, M_i\sigma_i^2), \quad (5)$$

where $M_i = \tau^2/(\tau^2 + \sigma_i^2/n)$ and $\bar{x}_i = \frac{1}{n}\sum_j x_{ij}$, for the posterior distribution of θ_i.

In such a case, we would use the posterior mean:

$$\hat{\theta}_i = M_i\bar{x}_i + (1 - M_i)\mu, \quad (6)$$

as a point estimator of θ_i.

However, the true population variances σ_i^2 for $i = 1, 2, \ldots, p$ are unknown, and as with [15] we adopt a lognormal prior (3) for σ_i^2, with the additional assumption (4) that S_i^2/σ_i^2 is also lognormally distributed (with parameters selected to coincide with those of χ_d^2/d, the standard distributional assumption regarding the quantity S_i^2/σ_i^2).

From (4), it follows that:

$$\log S_i^2 | \log \sigma_i^2 \sim N(m + \log \sigma_i^2, \sigma_{ch}^2), \quad (7)$$

and combining (3) and (7) using Bayes rule yields:

$$\log \sigma_i^2 | \log S_i^2 \sim N(M_v(\log S_i^2 - m) + (1 - M_v)\mu_v, M_v\sigma_{ch}^2), \quad (8)$$

where $M_\upsilon = \tau_\upsilon^2/(\tau_\upsilon^2 + \sigma_{ch}^2)$.

As in [15], we estimate this quantity using:

$$\hat{\sigma}_i^2 = \exp(M_\upsilon(\log S_i^2 - m) + (1 - M_\upsilon)\mu_\upsilon). \tag{9}$$

Thus, assuming known hyperparameters, we have the estimator:

$$\hat{\theta}_i = M_i \bar{x} + (1 - M_i)\mu, \tag{10}$$

with $M_i = \tau^2/(\tau^2 + \hat{\sigma}_i^2/n)$, where $\hat{\sigma}_i^2$ is as given by equation (9) above.

2.3 Estimation of Hyperparameters

All that remains is the estimation of the hyperparameters $\mu, \tau^2, \mu_\upsilon$ and τ_υ^2. As with [15], we adopt the following estimators.

For μ_υ and τ_υ^2, we have:

$$\hat{\mu}_\upsilon = \frac{1}{p}\sum_i (\log S_i^2 - m) \quad \text{and} \quad \hat{\tau}_\upsilon^2 = (\frac{1}{p}\sum_i (\log S_i^2 - m)^2 - \sigma_{ch}^2 - \hat{\mu}_\upsilon^2)_+,$$

from which we obtain:

$$\hat{M}_\upsilon = \frac{\hat{\tau}_\upsilon^2}{\hat{\tau}_\upsilon^2 + \sigma_{ch}^2} \quad \text{and} \quad \hat{\sigma}_{EB,i} = \exp(\hat{M}_\upsilon(\log S_i^2 - m) + (1 - \hat{M}_\upsilon).$$

To estimate μ and τ^2, we use:

$$\hat{\mu} = \frac{\sum_i (\bar{x}_i/\hat{\sigma}_{EB,i})}{\sum_i (1/\hat{\sigma}_{EB,i})} \quad \text{and} \quad \hat{\tau}^2 = (\frac{\sum_i (\bar{x}_i - \hat{\mu})^2}{p})_+.$$

Thus, the ‘double-shrinkage’ empirical Bayes point estimator is given as:

$$\hat{\theta}_{EB,i} = \hat{M}_{EB,i}\bar{x}_i + (1 - \hat{M}_{EB,i})\hat{\mu}, \tag{11}$$

where $M_{EB,i} = \hat{\tau}^2/(\hat{\tau}^2 + \hat{\sigma}_{EB,i}^2/n)$, and $\hat{\mu}$, $\hat{\tau}^2$ and $\hat{\sigma}_{EB,i}^2$ are as given above.

3 Introducing the DES Test Model

The purpose of this study is to evaluate the application of the EB methodology to the estimation of DES model performance measures. Having already introduced the EB procedures which are to be evaluated, this section aims to discuss an appropriate DES model upon which to test the EB procedures.

The model to be used is a computer-based implementation of an $M/M/1$ queuing model. This simple model consists of a single-server queuing system with exponentially distributed inter-arrival and service times and a first in - first out (FIFO) queuing discipline.

Simple "artificial" DES models such as this are frequently used in research for the evaluation of DES model analysis techniques [7]. These test models offer the key advantage that theoretical values are available for many performance measures of interest, and this knowledge greatly facilitates the testing of output analysis methods. One criticism which can be leveled at this approach is that such models bear little resemblance to the majority of DES models encountered in practice (those being significantly more complex). In light of this point, it is helpful to highlight the exploratory nature of the study; it should be emphasized that much of the value of this investigation lies in the issues it raises and the directions for further research which surface. This discussion is taken up again in Section 6 after the results are presented.

For now, some relevant queuing model theory and results, extracted from [14] and used in later sections of the paper, are presented.

3.1 $M/M/1$ Theory

As discussed above, an appropriate choice of DES test model for this investigation appears to be an $M/M/1$ model with a first in - first out queuing discipline. In this case, there are only two additional model parameters which may be varied, the arrival rate, denoted by λ, and the service rate, denoted by μ. In order to simplify our analysis, we note that these parameters may be combined to give a single parameter, namely the traffic intensity, which uniquely specifies a particular $M/M/1$ configuration. The traffic intensity parameter, denoted by ρ, is given by $\rho = \frac{\lambda}{\mu}$. We note that in future discussions, the particular $M/M/1$ configuration will be specified solely by reference to the traffic intensity, ρ.

Our performance measure of interest will be the steady-state (or long-run) expected time in system, which we denote by W. The exact value of this quantity for any $M/M/1$ model is given as a simple function of the arrival (λ) and service rate (μ) parameters, $W = \frac{1}{\mu-\lambda}$. This enables us to calculate the steady-state expected time in system exactly for any given $M/M/1$ model configuration.

4 Experimental Design

The aim of this section is to provide an overview of the experimental design employed during the computational testing. Although the key points will be presented, a more comprehensive picture of the research design may be obtained from visiting the following web address: `http://personal.strath.ac.uk/shona.blair/research/SCOR2012/`

Initially, it is important to mention that in this experiment, the performance of the EB estimator, $\hat{\theta}_{EB,i}$, (as given by (11) in Section 2), is evaluated relative to that of the standard estimator of a population mean, the sample mean $\bar{X}_i$. The sample mean is most prevalent point estimator of DES performance measures [9]; it offers the advantage of being easily interpretable and constitutes a convenient standard by which other methods may be compared.

In order to estimate the steady-state time in system, the test model had to be executed and time in system data collected. The relevant details concerning this stage of the experimentation are as follows:

- **Traffic intensities**: a mesh with lower limit $\rho = 0.02$, upper limit $\rho = 0.9$ and step-size $\rho = 0.02$ was used in the experiment. This resulted in 45 model configurations and gave a nice, fine grid.
- **Number of replications**: $200,000$ independent replications were made at each traffic intensity configuration to ensure sufficient data was collected.
- **Warm-up period**: a warm-up of 500 customers was used, ensuring that the model was in steady-state prior to data collection.
- **Run-length**: a run-length of 600 customers was used. The relatively short run-length (in comparison to the warm-up period) ensured the experiment reflected possible real-life data collection constraints.

This scheme yields a $200,000$ by 45 data matrix, where each data value represents an average time in system based on the first 100 customers after the aforementioned warm-up period.

Having obtained the necessary $M/M/1$ time in system data, we now discuss how this data can be used to calculate both EB and standard estimates. In the standard setting, we simply collect a batch of data from each traffic intensity configuration and calculate the sample mean to make inference regarding the true population mean. The only decision to be made concerns the size of the batch of data. In the empirical Bayes setting, however, data obtained from other model configurations can be pooled to support inference of the population mean of a given configuration. Thus immediate decisions need to be made concerning the size of the batches to be used and the pooling strategy to be employed.

To gain as much insight as possible, a range of batch sizes were explored, from 3 to 20. This enabled us to understand how the batch size affected the performance of the estimation technique. Additionally, in terms of the pooling strategy employed, a simple two sample t-test was conducted systematically for each pair of traffic intensity configurations to test for homogeneity. This approach was adopted as it avoids subjective decisions, based on theoretical knowledge of the system, biasing the results of the study. A range of significance levels, from $0.05, 0.1, 0.15, ..., 0.95$, were used to explore the relationship between pooling strategy and EB performance. This design resulted in 342 possible combinations of batch size and significance level, each of which was evaluated in the course of the computational testing. For each of these batch size / significance level combinations, the large volume of $M/M/1$ data collected permitted the calculation of $10,000$ pairs (both EB and standard) of estimates for each of the 45 traffic intensities. It is thus convenient to consider two $10,000$ by 45 matrices (one for each estimator) associated with each of the 342 batch size / significance level combinations.

To assess the performance of each estimator, the estimated values were subtracted from the true values, the errors squared, and the averages calculated over the 45 traffic intensities. This created, for each estimator, $10,000$ mean squared error (MSE) values. These were averaged, to remove the issue of stochastic variability, and the square root was taken to obtain the root mean squared error (rMSE) for each estimator. In order to gauge the relative efficacy, the ratio of the EB rMSE over the standard rMSE was taken as our critical statistical metric of interest. Finally, note that both the $M/M/1$ model and the EB and standard data analysis procedures were implemented in Matlab R2011a and run on a standard desktop PC (Intel Core-i5, 3GHz, 8GB RAM).

5 Summary and Discussion of Experimental Results

This section of the paper briefly presents and discusses the key results obtained from the aforementioned program of experimentation. As in the previous section, we begin by noting that a comprehensive set of numerical results and a detailed analysis may be found at the same web address. As mentioned in the previous section, the key statistical metric of interest is the ratio of the rMSE of the EB estimator to the rMSE of the standard estimator. The value of this quantity for each combination of batch size and significance level can found in Table 1 of Appendix A, which has been illustrated in Figure 1.

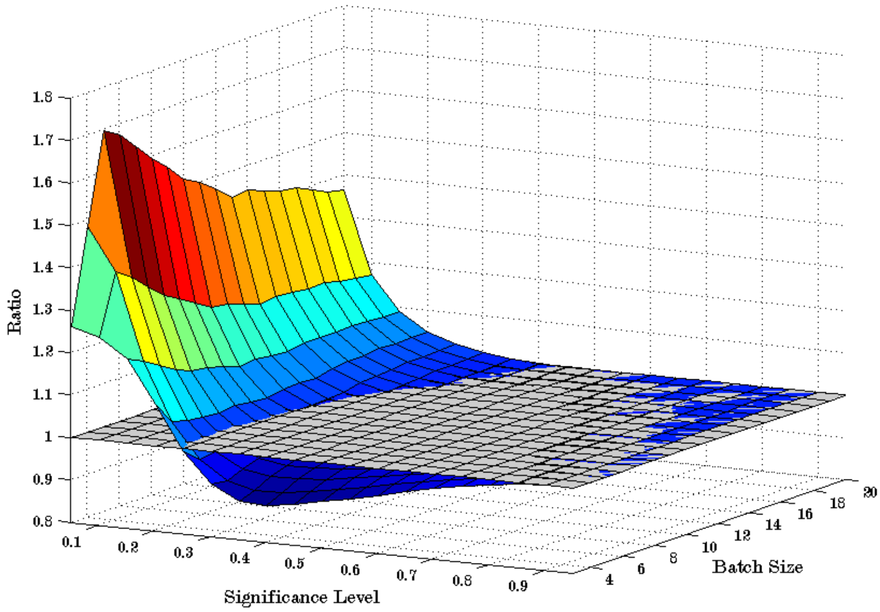

Figure 1 Ratio of EB rMSE to standard rMSE: This depicts the relative performance of the estimators, and how this varies with batch size and significance level. The plane at ratio= 1 enables us to easily identify areas where the EB estimator outperforms the standard estimator.

From examination of Table 1, the most favourable value of this ratio (0.8852) was obtained using a batch size of 3 and a significance level of 0.4, whilst the least favourable (1.6980) occurred for batch size 5 and a significance level of 0.05. Additionally, it appears that the optimal significance level, in terms of EB performance, is 0.4. As such, Figure 2 illustrates the relative EB performance for this significance level over varying batch size.

6 Conclusion and Future Research

As may be seen from the results illustrated in the previous section, quantifiable benefit can be achieved from the application of EB procedures to DES performance measure estimation.

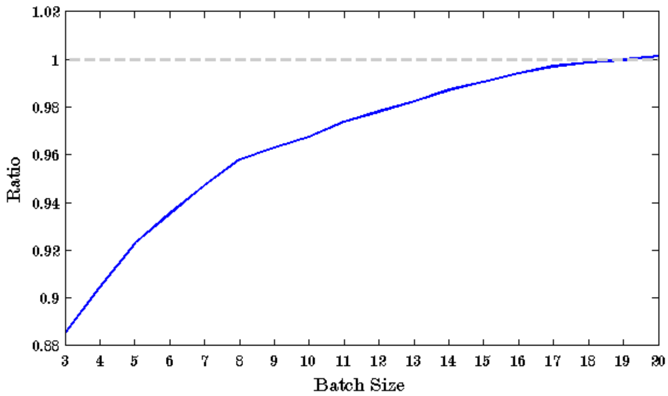

Figure 2 rMSE ratio over varying batch size for optimal significance level.

This is a positive outcome to our pilot study which goes a significant way towards establishing the feasibility of the application of empirical Bayes to DES model performance measure estimation. However, it is also apparent that the batch size used and the pooling strategy adopted are critical to the realization of this benefit. In this study, the decision as to whether or not to pool model configurations was made on the outcome of a simple two sample t-test. Although there are many more options for statistical tests of homogeneity, little in the way of formal guidance exists specifically concerning empirical Bayes pooling strategies. More exploration should be done on this subject with the aim of providing statistical indicators to guide practitioners in how to apply this method.

DES models exhibit great variety, differing vastly in characteristics such as purpose, structure, scale, nature of input parameters and nature of output distributions. This complexity increases the challenges involved in attempting to apply empirical Bayes to model analysis and in attempting to provide detailed guidelines enabling practitioners to make use of this methodology. A comprehensive study evaluating the benefits of EB in relation to more "realistic" DES models forms part of our ongoing research.

Acknowledgements We would like to thank Stephen Chick for several fruitful discussions on this topic and for his involvement as external supervisor of S. Blair, and the Simul8 Corporation for their technical advice and general project support.

A Detailed Results

Table 1 rMSE ratios for 342 combinations: EB outperforming standard highlighted.

Batch	Significance Level									
Size	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
3	1.2616	1.2425	1.1975	1.0992	**0.9934**	**0.9226**	**0.8887**	**0.8852**	**0.8939**	**0.9085**
4	1.4835	1.3842	1.1791	1.0449	**0.9616**	**0.9177**	**0.9026**	**0.9044**	**0.9134**	**0.9242**
5	1.6980	1.3588	1.1560	1.0339	**0.9666**	**0.9334**	**0.9215**	**0.9227**	**0.9290**	**0.9382**
6	1.6748	1.3238	1.1351	1.0287	**0.9711**	**0.9447**	**0.9346**	**0.9352**	**0.9393**	**0.9481**
7	1.6339	1.2898	1.1187	1.0255	**0.9782**	**0.9555**	**0.9471**	**0.9472**	**0.9506**	**0.9575**
8	1.5932	1.2686	1.1127	1.0302	**0.9877**	**0.9670**	**0.9585**	**0.9580**	**0.9602**	**0.9657**
9	1.5593	1.2473	1.1020	1.0265	**0.9891**	**0.9715**	**0.9635**	**0.9629**	**0.9651**	**0.9695**
10	1.5176	1.2244	1.0913	1.0243	**0.9907**	**0.9751**	**0.9682**	**0.9674**	**0.9697**	**0.9742**
11	1.4973	1.2189	1.0944	1.0311	**0.9990**	**0.9839**	**0.9763**	**0.9739**	**0.9755**	**0.9790**
12	1.4691	1.2055	1.0893	1.0323	1.0019	**0.9867**	**0.9801**	**0.9780**	**0.9790**	**0.9824**
13	1.4381	1.1922	1.0863	1.0312	1.0032	**0.9897**	**0.9842**	**0.9825**	**0.9826**	**0.9849**
14	1.4411	1.1983	1.0938	1.0391	1.0126	**0.9970**	**0.9900**	**0.9872**	**0.9870**	**0.9883**
15	1.4262	1.1947	1.0946	1.0444	1.0166	1.0018	**0.9939**	**0.9906**	**0.9904**	**0.9924**
16	1.4135	1.1900	1.0945	1.0439	1.0179	1.0042	**0.9976**	**0.9942**	**0.9932**	**0.9936**
17	1.4074	1.1910	1.0972	1.0492	1.0239	1.0092	1.0015	**0.9971**	**0.9961**	**0.9961**
18	1.3900	1.1816	1.0937	1.0476	1.0231	1.0094	1.0009	**0.9986**	**0.9965**	**0.9971**
19	1.3735	1.1769	1.0912	1.0484	1.0249	1.0119	1.0043	**0.9997**	**0.9979**	**0.9985**
20	1.3641	1.1718	1.0913	1.0496	1.0269	1.0131	1.0057	1.0013	**0.9992**	**0.9996**

Batch	Significance Level								
Size	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
3	**0.9216**	**0.9359**	**0.9509**	**0.9651**	**0.9774**	**0.9878**	**0.9954**	**0.9993**	1.0000
4	**0.9368**	**0.9499**	**0.9618**	**0.9735**	**0.9838**	**0.9921**	**0.9975**	**0.9999**	1.0000
5	**0.9490**	**0.9591**	**0.9698**	**0.9803**	**0.9886**	**0.9947**	**0.9989**	1.0000	1.0000
6	**0.9569**	**0.9669**	**0.9761**	**0.9849**	**0.9915**	**0.9966**	**0.9993**	1.0000	1.0000
7	**0.9657**	**0.9737**	**0.9812**	**0.9885**	**0.9945**	**0.9983**	**0.9997**	1.0000	1.0000
8	**0.9718**	**0.9788**	**0.9861**	**0.9922**	**0.9967**	**0.9990**	**0.9999**	1.0000	1.0000
9	**0.9754**	**0.9820**	**0.9886**	**0.9937**	**0.9974**	**0.9995**	**0.9999**	1.0000	1.0000
10	**0.9788**	**0.9845**	**0.9907**	**0.9953**	**0.9981**	**0.9994**	1.0000	1.0000	1.0000
11	**0.9831**	**0.9884**	**0.9925**	**0.9964**	**0.9990**	**0.9999**	1.0000	1.0000	1.0000
12	**0.9859**	**0.9906**	**0.9945**	**0.9975**	**0.9995**	**0.9999**	1.0000	1.0000	1.0000
13	**0.9878**	**0.9928**	**0.9961**	**0.9985**	**0.9997**	1.0000	1.0000	1.0000	1.0000
14	**0.9913**	**0.9946**	**0.9974**	**0.9991**	**0.9999**	**0.9999**	1.0000	1.0000	1.0000
15	**0.9939**	**0.9962**	**0.9983**	**0.9996**	**0.9999**	1.0001	1.0000	1.0000	1.0000
16	**0.9950**	**0.9972**	**0.9991**	**0.9999**	1.0000	1.0001	1.0000	1.0000	1.0000
17	**0.9976**	**0.9995**	1.0003	1.0002	1.0000	1.0000	1.0000	1.0000	1.0000
18	**0.9979**	**0.9991**	**0.9997**	1.0001	1.0000	1.0000	1.0000	1.0000	1.0000
19	1.0001	1.0006	1.0002	1.0000	1.0001	1.0000	1.0000	1.0000	1.0000
20	**0.9996**	1.0002	1.0002	1.0001	1.0000	1.0000	1.0000	1.0000	1.0000

References

1 Y. Benjamini. Discovering the false discovery rate. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):405–416, 2010.

2 B.P. Carlin and T.A. Louis. *Bayesian methods for data analysis*. Texts in statistical science. CRC Press, 2009.

3 B. Efron. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2010.

4 B. Efron and C. Morris. Limiting the risk of bayes and empirical bayes estimators - part ii: The empiricial bayes case. *Journal of the American Statistical Society*, 67:130–139, 1972.

5 B. Efron and C. Morris. Stein's estimation rule and it competitors - an empiricial bayes approach. *Journal of the American Statistical Society*, 68:117–130, 1973.

6 B. Efron and C. Morris. Dta analysis using stein's estimator and its generalizations. *Journal of the American Statistical Society*, 70:311–319, 1975.

7 K. Hoad, S. Robinson, and R. Davies. Classification of discrete event simulation models and output: creating a sufficient model set. In *Operational Research Society Simulation Workshop 2010 (SW10)*, 2010.

8 J.P.C. Kleijnen, S.M. Sanchez, T.W. Lucas, and T.M. Cioppa. State-of-the-art review: A user's guide to the brave new world of designing simulation experiments. *INFORMS Journal on Computing*, 17(3):263–289, 2005.

9 A.M. Law. *Simulation modeling and analysis*. McGraw-Hill series in industrial engineering and management science. McGraw-Hill, 2007.

10 J.S. Maritz. *Empirical Bayes methods*. Methuen's monographs on applied probability and statistics. Methuen, 1970.

11 M. Pidd. *Computer simulation in management science*. Wiley, 2004.

12 H. Robbins. An empirical bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pages 157–164. University of California Press, 1955.

13 S. Robinson. Conceptual modelling for simulation part i: definition and requirements. *Journal of the Operational Research Society*, 59(3):278–290, 2008.

14 S.M. Ross. *Introduction to Probability Models*. Academic Press, 2009.

15 Z. Zhao. Double shrinkage empirical bayesian estimation for unknown and unequal variances. *Statistics and Its Interface*, 3(4):533–541, 2010.