



HAL
open science

Algorithmic Inference of Two-Parameter Gamma Distribution

Bruno Apolloni, Simone Bassis

► **To cite this version:**

Bruno Apolloni, Simone Bassis. Algorithmic Inference of Two-Parameter Gamma Distribution. *Communications in Statistics - Simulation and Computation*, 2009, 38 (09), pp.1950-1968. 10.1080/03610910903171821 . hal-00520670

HAL Id: hal-00520670

<https://hal.science/hal-00520670>

Submitted on 24 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Algorithmic Inference of Two-Parameter Gamma Distribution

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2009-0057.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	01-Jul-2009
Complete List of Authors:	Apolloni, Bruno; University of Milano, Department of Computer Science Bassis, Simone; University of Milano, Department of Computer Science
Keywords:	Two-parameters Gamma distribution, Algorithmic inference, Population bootstrap, Gibbs sampling, Adaptive rejection sampling, Maximum likelihood estimator
Abstract:	We provide an estimation procedure of the two-parameter Gamma distribution based on the Algorithmic Inference approach. As a key feature of this approach, we compute the joint probability distribution of these parameters without assuming any prior. To this end we propose a numerical algorithm which is often beneficial of a highly efficient speed up based on an approximate analytical expression of the probability distribution. We contrast our interval and point estimates with those recently obtained in Son and Oh (2006). We realize that our estimates are both unbiased and more accurate, albeit more dispersed than Bayesian methods.
<p>Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.</p>	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

infgamma.zip



For Peer Review Only

INFERENCE

ALGORITHMIC INFERENCE OF TWO-PARAMETER GAMMA DISTRIBUTION

B. Apolloni and S. Bassis

Dipartimento di Scienze dell'Informazione

University of Milano

Via Comelico 39/41, 20135

Milano, Italy

{apolloni,bassis}@dsi.unimi.it

Key Words: Two-parameters Gamma distribution; Algorithmic inference; Population bootstrap; Gibbs sampling; Adaptive rejection sampling; Maximum likelihood estimator.

ABSTRACT

We provide an estimation procedure of the two-parameter Gamma distribution based on the Algorithmic Inference approach. As a key feature of this approach, we compute the joint probability distribution of these parameters without assuming any *prior*. To this end we propose a numerical algorithm which is often beneficial of a highly efficient speed up based on an approximate analytical expression of the probability distribution. We contrast our interval and point estimates with those recently obtained in Son and Oh (2006) for the same problem. From this benchmark we realize that our estimates are both unbiased and more accurate, albeit more dispersed, in some cases, than the competitor methods', where the dispersion drawback is notably mitigated w.r.t. Bayesian methods by a greater estimate decorrelation. We also briefly discuss the theoretical novelty of the adopted inference paradigm which actually represents a brush up on a Fisher perspective dating to almost a century, made feasible today by the available computational tools.

1. INTRODUCTION

The Gamma distribution is a formidable touchstone for any inference framework. On

the one hand it is widely employed in many application fields, ranging from reliability (Jurán and Godfrey, 1999; Corp., 2009) to telecommunications (Titus and Wheatley, 1996), insurance (Ammeter, 1970) and so on. On the other, the estimation of its parameters still poses a challenging problem. Adopting an analogous notation as in Son and Oh (2006), we introduce the Gamma distribution through its probability density function (PDF) applied to the random variable X as follows ¹:

$$f_X(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}, \quad x > 0, \quad \alpha, \beta > 0, \quad (1)$$

where α and β are respectively a shape and scale parameter. Of the conventional point estimators, the one based on the method of moments (MM) is very simple but inefficient as well – with some exceptions in asymptotic conditions (Wiens et al., 2003). Maximum likelihood (ML) estimators may rely on a pair of joint sufficient statistics, but they are penalized by the absence of a closed analytical expression. To overcome this drawback, we find in the literature either approximate expressions (Thom, 1958; Greenwood and D., 1960) or recursive procedures to get a numerical solution (Bowman and Shenton, 1988). With the latter we work with a sequence of parameter values converging with some dynamics to the planned estimators. With a dual perspective, other researchers adopt the Bayes approach to collect a population of parameter estimators – possibly generated through a Gibbs sampler (Ripley, 1988) or more sophisticated techniques (Gilks and P., 1992) – and deduce point estimators as central values of this population.

Our work goes in the same direction, but without requiring to set any *a priori* distribution of the parameters. We have implemented this kind of procedure in many inference problems, mainly within the machine learning framework (Apolloni et al., 2008a) and in specific operational fields such as survival analysis (Apolloni et al., 2007b), human mobility modeling (Apolloni et al., 2007a) and statistical tests of hypotheses (Apolloni and Bassis, 2007). The goal of this paper is twofold: on the one hand we want to efficiently solve the

¹By default, capital letters (such as U , X) will denote random variables and small letters (u , x) their corresponding realizations; vectorial versions of the above variables will be denoted through bold symbols (\mathbf{X} , \mathbf{U} , \mathbf{x} , \mathbf{u}).

1 questioned estimation problem as a distinguished inference task. To this aim, we show our
2 procedures to quickly compute accurate and unbiased estimators based on independent ap-
3 proximations of the unknown parameters. On the other hand, we would use this solution as
4 a business card to introduce our inference framework, which we call Algorithmic Inference,
5 to a more probability methodologically oriented audience and invite the reader to read more
6 widely about it in a couple of books we have published on the subject (Apolloni et al., 2008c,
7 2006). Therefore the paper is organized as follows. In Section 2 we introduce our method
8 from a strictly operational perspective, through definitions and a few logical links, having
9 the questioned inference problem as lead example. In Section 3 we apply the proposed
10 procedure to the same benchmark as Son and Oh (2006), contrasting our results with the
11 competitor methods considered therein. We devote Section 4 to a discussion of the benefits
12 of our approach and its connections to some inference strategies animating the debate since
13 the early 20th century, to conclude the paper.

26 2. COMPATIBLE PARAMETERS

28 Given a distribution law, think of its unknown parameters as random parameters. But,
29 in place of a specific prior distribution, they inherit probabilities from a standard source
30 of random seeds that are also at the basis of the random variables they are deputed to
31 characterize. Like with a barrel with two interlocking taps, through the former you get,
32 from the source, samples of the random variable for given parameters; while from the latter
33 you get samples of the parameters for given observed variables. You measure what happens
34 with the former and compute what would happen with the latter. The basic steps of this
35 twisting task are the following:

- 43 1. *Sampling mechanism* \mathcal{M}_X . It consists of a pair $\langle Z, g_{\theta} \rangle$, where the *seed* Z is a random
44 variable without unknown parameters, while the *explaining function* g_{θ} is a function
45 mapping from samples $\{z_1, \dots, z_m\}$ of Z to samples $\{x_1, \dots, x_m\}$ of the random variable
46 X we are interested in. This function is indexed in θ which represents the not yet set
47 parameters of the random variable, i.e either a scalar value or a vector of values we want
48 to investigate. Thanks to the probability integral transformation theorem (Rosenblatt,
49
50
51
52
53
54

1952) we have that, by using the uniform variable U in $[0, 1]$ as a seed, the explaining function g_{θ} for X distributed according to (1) computes the following mapping:

$$x_i = F_X^{-1}(u_i) \quad (2)$$

where F_X^{-1} is the inverse $h_{\alpha,\beta}$ of the Gamma cumulative distribution function (CDF). What is interesting is that β plays the role of a scale factor, such that $h_{\alpha,\beta} = \beta h'_{\alpha}$, with $h'_{\alpha} = h_{\alpha,1}$. In turn h'_{α} has an analytical, though complex, expression that may be managed by widespread tools of symbolic computation. For instance in Mathematica (Wolfram Research Inc., 2008) it is computed by the routine `InverseGammaRegularized` $[\alpha, 0, u]$ where u is the argument of F_X^{-1} . Not always U is the most appropriate seed. For instance one prefers using the standard Gaussian variable Ψ as a seed of a Gaussian variable X with mean μ and standard deviation σ through the explaining function $x_i = (\mu + \psi_i\sigma)$, as F_X^{-1} is unavailable in its exact form for this X .

All this falls in the common practice of random variable simulation. The benefit of formalizing it in terms of sampling mechanism $\langle Z, g_{\theta} \rangle$ lies firstly in a clear partition of what is out of our handling and what may be the object of our inference. We can say nothing new about seeds $\{z_1, \dots, z_m\}$, since they are randomly tossed from a perfectly known distribution; hence they are completely unquestionable as for the single value, and completely known as for their ensemble properties. On the contrary, the explaining function groups into θ the free parameters that we want to infer from the sample. As for a second benefit, they are exactly the seeds that state links between observations and parameters of a given X . We cannot say which value has θ , since we do not know the seeds of the observations. Rather, we may transfer the probability mass of the seeds from the sample to the parameters realizing the sample, as we will see in the next point.

2. *Master equations.* The actual connection between the model and the observed data is tossed in terms of a set of relations between statistics on the data and unknown parameters that come as a corollary of the sampling mechanism. We call these relations master

equations. Pivoting around the statistic $\mathbf{s} = \mathbf{h}(x_1, \dots, x_m) = \mathbf{h}(g_{\theta}(z_1), \dots, g_{\theta}(z_m))$, where \mathbf{s} and \mathbf{h} are vectors in their general instantiation, the general form of a system of *master equations* is:

$$\mathbf{s} = \boldsymbol{\rho}(\boldsymbol{\theta}; z_1, \dots, z_m). \quad (3)$$

With this relation we may inspect the values of the parameter $\boldsymbol{\theta}$ that *could have generated* a sample with the observed statistic \mathbf{s} from a particular setting of the seeds $\{z_1, \dots, z_m\}$. Hence, if we draw seeds according to their known distribution – uniform in our case – we get a sample of parameters in response (Apolloni et al., 2006). In order to ensure this sample clean properties, we involve sufficient statistics w.r.t. the parameters (Stigler, 1973) in the master equations. In greater detail, with Gamma distributions we refer to the jointly sufficient statistics $s_1 = \sum_{i=1}^m x_i$ and $s_2 = \prod_{i=1}^m x_i$, which we prefer to represent through s_1 and $s_3 = s_2/s_1^m$ so that the master equation on s_3 depends only on α . Namely, we will use the following set of master equations:

$$s_1 = \beta \sum_{i=1}^m h'_{\alpha}(u_i) \quad (4)$$

$$s_3 = \frac{\prod_{i=1}^m h'_{\alpha}(u_i)}{(\sum_{i=1}^m h'_{\alpha}(u_i))^m} \quad (5)$$

3. *Parameters population.* Having fixed a set of master equations, you may draw seeds in an infinitely large number so as to map from a population of seeds into a population that is representative of the random parameter Θ . The specific features of the mapping translate the uniform distribution of the former into a properly shaped distribution of the latter. In this way we obtain the graph in Figure 1 which reports the joint empirical distribution law of the random parameters A and B on the basis of the male mice dataset of Gross and Clark (1975) (G&C dataset henceforth), consisting of the 20 randomly selected survival times of male mice exposed to 240 rads of gamma radiation as listed in Table 1.

On the one hand we have no problem realizing that the mentioned probability distribution transform may occur, provided that some properties are owned by the involved statistics,

152	152	115	109	137	88	94	77	160	165
125	40	128	123	136	101	62	153	83	69

Table 1: Male mice dataset of Gross and Clark (1975).

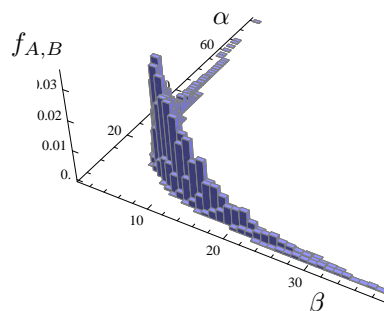


Figure 1: Empirical joint density function of the A and B parameters of Gamma distribution laws that are compatible with G&C dataset, on the basis of 10,000 bootstrap replicas.

such as the univocal invertibility of (3) w.r.t. θ or the feasibility of the inversion solution. We formally denoted a statistic with these properties as *well behaving* in Apolloni et al. (2008c) and realized that a statistic whose components are jointly sufficient is well behaving. On the other hand, the value of the new distribution stands in measuring a compatibility feature of the found parameters with the observed statistic that we formalize through the following definition.

Definition 1. For a random variable X and a sample drawn from it, a compatible distribution is a distribution having the same sampling mechanism $\mathcal{M}_X = \langle Z, g_{\check{\theta}} \rangle$ of X with a value $\check{\theta}$ of the parameter Θ derived from a system of master equations (3) rooted on a well behaving statistic \mathbf{s} . In turn, $\check{\theta}$ is a Θ value compatible with the sample.

As for the algorithms implementing the above procedure, we distinguish a numerical implementation from an analytical counterpart, leaving room for any intermediate specification. The lead example developed till now is an instance of the following Algorithm 1 that we call *p-bootstrap* in order to both evoking the original idea of Efron and Tibshirani (1993)

and highlighting the fact that we bootstrap population parameters and not samples.

Algorithm 1 Generating parameter populations through bootstrap.

1. Identify a well behaving statistic \mathbf{S} for the parameter $\boldsymbol{\theta}$ and the related system of master equations;
 2. compute a realization \mathbf{s}_{Θ} of \mathbf{S} from a sample of size m ;
 3. repeat for a satisfactory number n of iterations:
 - (a) draw a sample $\check{\mathbf{z}}_i$ of size m from the seed deputed random variable;
 - (b) get $\check{\boldsymbol{\theta}}_i = \text{Inv}(\mathbf{s}_{\Theta}, \check{\mathbf{z}}_i)$ as a solution of (3) in $\check{\boldsymbol{\theta}}$ with $\mathbf{s} = \mathbf{s}_{\Theta}$ and $(z_1, \dots, z_m) = \check{\mathbf{z}}_i$;
 - (c) add $\check{\boldsymbol{\theta}}_i$ to Θ population.
-

Since we have no limits to the number of seed replicas – apart from the computational time we are allowed to spend – we may approximate with any accuracy probabilities of events related to the parameters through the companion frequencies. We just highlight the different role of the sample size decreeing the shallowness of the parameter distribution, from the number of replicas determining the accuracy with which this distribution is computed.

Figure 1 was obtained exactly with this procedure, where we numerically realize that inversion of (5) is univocal since s_3 results to be a monotonically increasing function of α for whatever seeds $\{u_1, \dots, u_m\}$ we checked.

Things go better, as for computational loads, if we may identify the analytical form of the above distribution. In our case, the two master equations together with the well behaving-ness of s_1 and s_3 guarantee that for both distributions we may state the following relationships:

$$s_1 \leq \tilde{s}_1 \leftrightarrow \beta \leq \tilde{\beta} \quad \forall \alpha \quad (6)$$

$$s_3 \leq \tilde{s}_3 \leftrightarrow \alpha \leq \tilde{\alpha} \quad (7)$$

where \tilde{s}_1 is the value of the statistic we would have observed if the parameter had moved to $\tilde{\beta}$ in the master equation (3) for the same specifications $\{z_1, \dots, z_m\}$ of Z and the same α , and

similarly for \tilde{s}_3 but independently of β . Summing up these implications over all seeds and maintaining fixed $s_1, s_3, \tilde{\alpha}, \tilde{\beta}$, we have analogous relationships between events (see Apolloni et al. (2006) for a wider discussion):

$$(s_1 \leq \tilde{S}_1) = (B \leq \tilde{\beta}) \quad \forall \alpha \quad (8)$$

$$(s_3 \leq \tilde{S}_3) = (A \leq \tilde{\alpha}) \quad (9)$$

Hence, moving to probability we have:

$$F_{B|A=\alpha}(\beta) = 1 - F_{S_1}(s_1) = Q(\alpha m, s_1/\beta) \quad (10)$$

$$F_A(\alpha) = 1 - F_{S_3}(s_3) \sim Q(\phi_1(\alpha), s_3\phi_2(\alpha)) \quad (11)$$

where $Q(a, z)$ denotes the regularized incomplete Gamma function (Chaudhry and Zubair, 2001) and ϕ_1, ϕ_2 are suitable functions we are going to introduce. Namely, while S_1 follows a Gamma distribution with parameters $m\alpha$ and β , the distribution of S_3 has no immediate analytical closed form. If we read it in terms of the product of Dirichlet components, its distribution – while treatable for $m = 2$ (Nadarajah and Kotz, 2004) – becomes computationally costly with higher sample sizes, mainly due to its characterization in terms of the Fox's H-function (Garg et al., 2002) or its derivatives (Coelho and Mexia, 2007). Thus, abandoning the search for an exact analytical form, we find again in the Gamma distribution a good candidate function for approximating F_{S_3} . Then, with the mentioned limits of MM estimators for this distribution, we exploit the relative ease of computing the first and second S_3 moments (Nadarajah and Kotz, 2004), and obtain an approximate distribution of A by plugging in the above approximating function the MM estimators, say (t_1, t_2) , of its parameters:

$$t_1 = \phi_1(\alpha) = \left(\frac{\Gamma(\alpha)^{m-1} \Gamma(\alpha+1)^{2-2m} \Gamma(\alpha+2)^{m-1} \Gamma(\alpha m + m - 1)^2}{\Gamma(\alpha m) \Gamma((\alpha+2)m - 2)} - 1 \right)^{-1} \quad (12)$$

$$t_2 = \phi_2(\alpha) = \left(\frac{\Gamma(\alpha+1)^{1-m} \Gamma(\alpha+2)^{m-1} \Gamma(\alpha m + m - 1)}{\Gamma((\alpha+2)m - 2)} - \frac{\Gamma(\alpha)^{1-m} \Gamma(\alpha+1)^{m-1} \Gamma(\alpha m)}{\Gamma(\alpha m + m - 1)} \right)^{-1} \quad (13)$$

Does it make sense to use MM to approximate the A distribution law in order to overcome the drawbacks of the A MM estimator? In Figure 2 we show that this F_{S_3} approximation

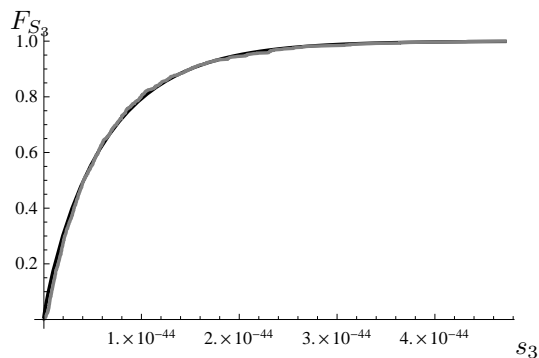


Figure 2: Comparison between CDFs of the S_3 statistic computed through a random sample with size 30 of a Gamma distribution with parameters $\alpha = 4$ and $\beta = 1$. Gray curve: empirical approximation computed on 1,000 sample realizations; black curve: analytical approximation.

may be very tight. On the one hand the shape of the Gamma CDF is well suited to the approximating purpose for $\alpha \geq 2$. On the other hand, however, we suffer from the bias of the MM estimators; therefore we must introduce some multiplicative coefficients γ_1 and γ_2 as tuning terms in (12) and (13). Since the ultimate goal is the approximation of A distribution, we get them directly from the interpolation of the A distribution function. Namely, denoting with $F_A^{-1}(r)$ the r -th quantile of A distribution computed via p-bootstrap method and with $Q_{(a,z)}^{-1}(r)$ the analogous quantile using (11), we use the 0.25 and 0.75 quantiles to obtain the tuning coefficients $\tilde{\gamma}_i$ s, i.e.:

$$Q_{(\tilde{\gamma}_1 t_1, s_3 \tilde{\gamma}_2 t_2)}^{-1}(0.25) = F_A^{-1}(0.25) \tag{14}$$

$$Q_{(\tilde{\gamma}_1 t_1, s_3 \tilde{\gamma}_2 t_2)}^{-1}(0.75) = F_A^{-1}(0.75) \tag{15}$$

We compute coefficients on a thousand of samples of a Gamma distribution with different sample size and distribution parameters. By interpolating these data, we obtain the general expression of γ_1 as a quadratic function of $\log s_3$ and m which works well for all our purposes in this paper:

$$\log \gamma_1 = 0.6476 - 0.1235m - 0.0077m^2 - 0.0548 \log s_3 - 0.0021m \log s_3 - 0.0001 \log^2 s_3 \tag{16}$$

Since we did not achieve a similar success with γ_2 , we decided to compute this coefficient directly in the single inference instances by fitting the empirical CDF obtained from a small set of α p-bootstrap replicas. With this strategy, the method we propose is a hybrid numerical-analytical procedure where the injection of the analytical knowledge, in terms of expressions (11) to (13), constitutes a speed up of the pure numerical solution allowing a notably reduction of the number of p-bootstrap replicas. Within the benchmarks used in this paper the narrowing rate ranges from 25 to 250. This entails an almost identical reduction in running times – since the time required to compute the above expression is negligible in comparison to the replicas' generation time. Thus the hybrid algorithm represents a relatively quick procedure attaining a good compromise between fastness and accuracy of the estimation algorithms, as will be shown in the next section. Focusing on the single components of the parameter vector and related scalar statistic, we may standardize the analytical computation with the following Algorithm 2. The discretization granule $\Delta > 0$ therein accounts for

Algorithm 2 Generating parameter populations through analytical arguments.

1. Identify a well behaving statistic S for the parameter θ and its discretization granule Δ (if any);
2. decide the monotony versus;
3. compute $F_{\Theta|S=s}(\theta) \in (q_1(F_{S|\Theta=\theta}(s)), q_2(F_{S|\Theta=\theta}(s)))$ where:

$$\begin{cases} q_1 = q_2 & \text{if } S \text{ continuous} \\ q_1(F_S(s)) = q_2(F_S(s - \Delta)) & \text{if } S \text{ discrete} \end{cases}$$

$$q_i(F_S) = \begin{cases} 1 - F_S & \text{if } s \text{ does not decrease with } \theta \\ F_S & \text{if } s \text{ does not increase with } \theta \end{cases}$$

for $i = 1, 2$.

the S value jumps in the case of discrete random variables. In this case, the confluence of

multiple seeds to the same sampling mechanism outputs induces an indeterminacy in the Θ distribution law reckoned by the pair (q_1, q_2) (see Apolloni et al. (2006) for wider discussion on the algorithm).

3. NUMERICAL EXPERIMENTS

Like dwarfs on the shoulders of giants, we decide to use the same benchmarks as in Son and Oh (2006). In principle we add a pair of new rows on the tables therein to contrast and compare our results with those of competitors, having the adaptive method proposed by the authors of the referred paper as the favorite benchmark.

Table 2 represents a set of reconstruction experiments. Starting with a set of Gamma variables with the same $\beta = 1$ and α spanning in the set $\{0.5, 1, 2, 3, 4, 5\}$, we generate a sample of 30 elements of each variable, on the basis of which we estimate the above parameters. Our competitors' methods are: i) two moment method implementations, one in the basic version (MM), and the other in the variant proposed in Wiens et al. (2003) and with a special selection of free parameters (WCB); ii) three maximum likelihood implementations, respectively with the approximations proposed in Thom (1958) (ML_{Thom}), in Greenwood and D. (1960) (ML_{GD}), and with the recursive procedure introduced in Bowman and Shenton (1988) (ML_{BS}); and iii) two Bayesian methods based on the prior:

$$f_{A,B}(\alpha, \beta) \propto 1/\beta \quad (17)$$

and Gibbs sampling of parameters α and β from the full conditional distributions. Namely, to achieve samples, a rejection method (Ripley, 1988) is used having a proposal distribution either normal (RS) (Tsionas, 2001) or adaptive (ARS) (Gilks and P., 1992). Bayesian methods generate samples of 1,000 realizations of the random parameters' posterior distribution – after a burn-in period of 100 iterations – and assume their mean as an estimate of the original parameters. The statistical behavior of the estimators is assessed over 1,000 samples of the Gamma variable for each parameter setting. The mean and the standard deviation are computed over the corresponding parameter estimates. Son and Oh focus only on α estimates, having the β estimation as a minor task. Indeed, with MM and ML methods

α	Method	Central value	Dispersion	α	Method	Central value	Dispersion
0.5	WCB _{k=0}	0.5282	0.1145	3.0	WCB _{k=0}	3.2364	0.7576
	MM	0.6091	0.1984		MM	3.2867	0.9018
	ML _{Thom}	0.5600	0.1105		ML _{Thom}	3.3624	0.7726
	ML _{GD}	0.5411	0.1139		ML _{GD}	3.3604	0.7731
	ML _{BS}	0.5411	0.1139		ML _{BS}	3.3609	0.7731
	RS	0.5569	0.1104		RS	3.3569	0.7675
	ARS	0.5400	0.1021		ARS	3.0451	0.3179
	AI _{p-boot}	0.4977	0.0766		AI _{p-boot}	3.0145	0.5518
				AI _{hybrid}	3.0210	0.5429	
1.0	WCB _{k=0}	1.0241	0.2412	4.0	WCB _{k=0}	4.2421	1.1319
	MM	1.0826	0.3224		MM	4.4578	1.2427
	ML _{Thom}	1.0670	0.2414		ML _{Thom}	4.3448	1.1676
	ML _{GD}	1.0580	0.2439		ML _{GD}	4.3433	1.1680
	ML _{BS}	1.0579	0.2439		ML _{BS}	4.3439	1.1680
	RS	1.0665	0.2419		RS	4.3456	1.1650
	ARS	1.0324	0.1514		ARS	4.0313	0.3718
	AI _{p-boot}	1.0044	0.1629		AI _{p-boot}	3.9231	0.7393
				AI _{hybrid}	3.8942	0.7405	
2.0	WCB _{k=0}	2.1185	0.5333	5.0	WCB _{k=0}	5.1577	1.3892
	MM	2.1857	0.6264		MM	5.1838	1.4469
	ML _{Thom}	2.1940	0.5497		ML _{Thom}	5.3459	1.4439
	ML _{GD}	2.1907	0.5506		ML _{GD}	5.3537	1.4443
	ML _{BS}	2.1909	0.5508		ML _{BS}	5.3542	1.4442
	RS	2.1920	0.5520		RS	5.3491	1.4477
	ARS	2.0757	0.2467		ARS	5.0503	0.4317
	AI _{p-boot}	2.0264	0.3459		AI _{p-boot}	4.9982	0.9631
				AI _{hybrid}	4.9864	1.0151	

Table 2: First and second order statistics for a set of reconstruction experiments for different values of α and $\beta = 1$. Central value and Dispersion columns refer respectively to sample mean and standard deviation, with the exception of the AI approach, where median and median deviation are used.

1 this estimation descends straightforwardly from the following relation between $\hat{\alpha}$ and $\hat{\beta}$, as
2 estimators of the corresponding parameters:
3
4

$$\hat{\beta} = s_1/(m\hat{\alpha}) \quad (18)$$

5
6
7
8 As to the Bayesian methods, for α given, sampling β from its conditional posterior distri-
9 bution looks trivial. The whole experimental campaign was repeated with samples of size
10 100. On our part, since we are working with interpreted languages, namely Mathematica
11 (Wolfram Research Inc., 2008), we preferred to avoid bearing the heavy computational effort
12 coming from the second set of simulations, in favor of specific experiments aimed at identi-
13 fying the different features of our approach w.r.t. the other ones. Hence, in the mentioned
14 table and with reference to the sole sample size 30, we add rows concerning central value
15 and dispersion of estimators from our approach (AI) using both p-bootstrap method and its
16 hybrid version.
17
18
19
20
21
22
23
24

25 On each sample we compute 1,000 bootstrap replicas $(\check{\alpha}, \check{\beta})$ of the parameters, as a
26 counterpart of analogous values obtained with the Bayesian methods. The hybrid version
27 uses subsamples of size 40 (just picking the first 40 of the above replicas) to compute γ_2 .
28 Unlike the Bayesian methods, we opt for the *sample median* in place of *sample mean* to
29 identify a central value. Analogously, we use the median of the medians coming from the
30 1,000 samples as indicator of their central value. Indeed, we assume this statistic to be
31 more suitable within the Algorithmic Inference approach. The problem is that, on the one
32 hand with this approach we refer to the parameter distribution laws compatible with the
33 observed sample. On the other, we generate a set of samples with exactly the same parameter
34 α . Hence, to adapt our method to the reconstruction problem we must use a statistic
35 that is substantially independent of the (either compatible or actual) parameter distribution
36 law. This occurs with the median, since, in the mechanism generating parameter replicas
37 from a sample $\{x_1, \dots, x_m\}$, the median of $\check{\alpha}$ depends only on the median of $\sum_{i=1}^m \log(1 -$
38 $u_i)$ and on no other detail of the distribution of this parameter. The same consideration
39 holds for the median of the medians coming from the samples. Thus, this statistic is not
40 biased by the actual parameter distribution law used to generate the sample, which is the
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

degenerate distribution with the support coinciding with one of the constant values of α . On the contrary, the adopted constant value coincides with the median of the degenerate distribution, in turn converging in probability with the median of medians (Zhu and Wang, 2005), which we use therefore as a resuming central value statistic.

As a result, we obtain in Table 2 central values that in 5 cases out of 6 with p-bootstrap and 3 cases out of 4 with the hybrid version are closer to the original parameter than those provided by ARS, and even better w.r.t. other competitors' methods. Above all, our estimates are *unbiased* in the sense that some are under the original parameter, others over it, while ARS and all other methods denote exclusively positive shifts of the estimate versus the original parameter (see Figure 3).

The worse results of p-bootstrap with $\alpha = 4$ may be related to unfavorable features of the superset of 1,000 samples we used. Despite this large number, indeed, in Table 3 we realize the anomaly of this set by considering the mean of sample means and the value ξ of the quantile level to which it corresponds. Namely, from the table we read this value for

α	0.5	1	2	3	4	5
ξ	0.064	0.371	0.416	0.829	0.125	0.599

Table 3: Quantile level ξ of the mean of the sample means observed on the 1,000 samples used to compute statistics in Table 2 for the different values of α .

each α ; for instance, with $\alpha = 2$ and $\beta = 1$ the probability that the mean of 1,000 Gamma sample means is less than or equal to the actual value observed on the samples used in our experiment is 0.416. Through these data, we realize that the supersets used with $\alpha = 0.5$ and $\alpha = 4$ are somehow anomalous in that they correspond to very low quantile levels with respect to the others. The higher dispersion in the second case enhances this drawback. The analytic approximation strictly follows the pure numerical results, possibly with small unfavorable drifts as a toll for the computational benefits we will discuss later on. We do not report performances in the first two cases since we found the analytical approximation working well only for $\alpha \geq 2$, as highlighted in Figure 4. It seems to be due to the unsuitability

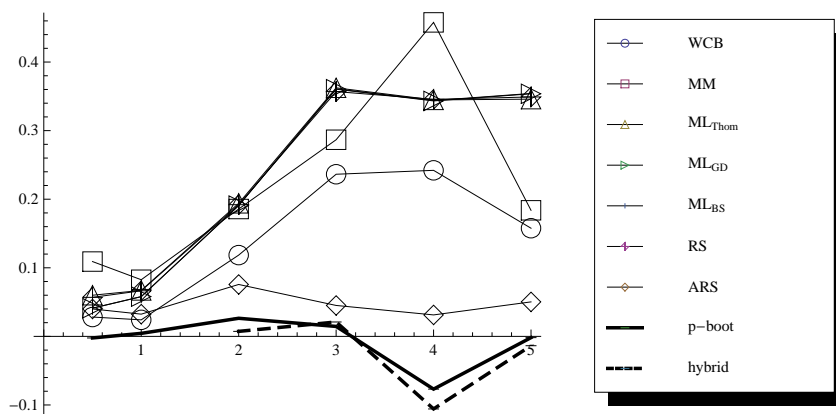


Figure 3: Estimators' bias for different values of α .

of the Gamma shape *per se*, since, for instance, in the picture with $\alpha = 0.05$ we left both γ_1 and γ_2 free in search of the best interpolation, which still proves unsuccessful.

As for estimates' dispersion, we essentially lose w.r.t. ARS and some of the other methods as well. Indeed, in line with the statistical meaning of median ν_X – namely it represents the parameter minimizing the sum of the absolute shifts, i.e.

$$\nu_X = \min_c \{E[|X - c|]\} \tag{19}$$

where E denotes the *expected value* operator – we use the median absolute deviation (MAD) as dispersion index which, with the above notation, reads:

$$MAD = \nu_{|x - \nu_X|} = \text{med}[|x - \nu_X|] \tag{20}$$

where med denotes the *median value* operator. Now MAD is generally a lower dispersion indicator than the standard deviation σ , where for a normal distribution we have in particular $MAD = \sqrt{2/\pi}\sigma$. Thus, in spite of the small MAD values we may read in the table, we conclude that the population of our α estimates is more dispersed than the analogous population obtained through ARS – while generally comparable with those of the other methods. We interpret this spreading as the counterpart of the estimator unbiasedness. To obtain this property, indeed, we do not rely on prior distributions, with a consequent broadening of the compatible parameter set.

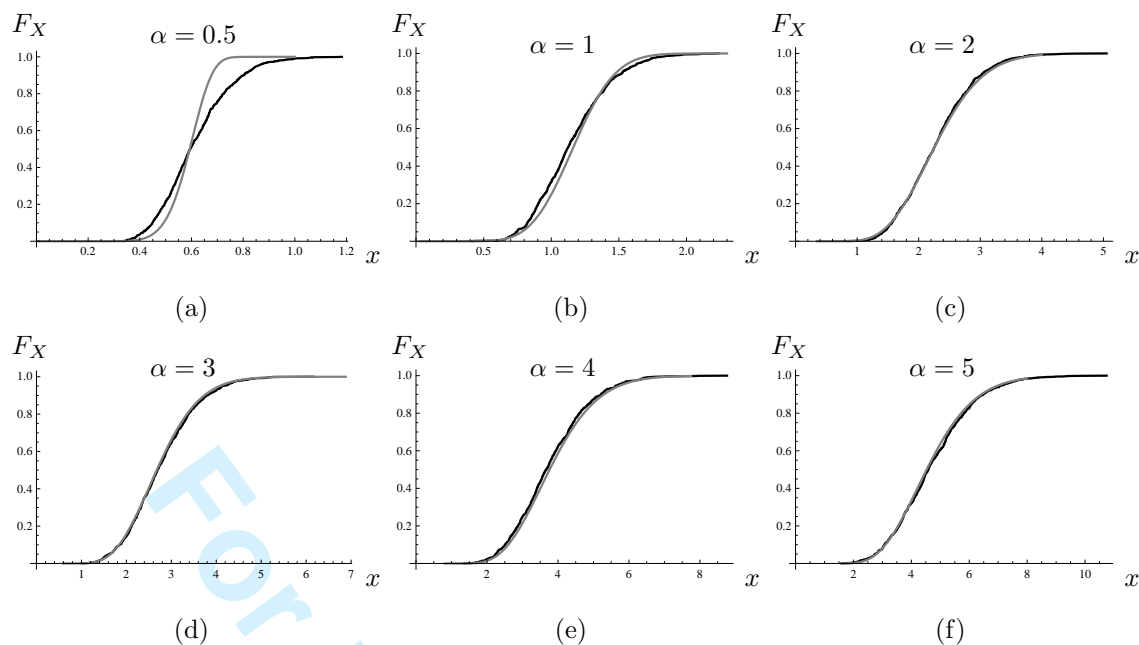


Figure 4: Templates of A distributions computed within Algorithmic Inference approach. Black curves: empirical CDFs drawn from a sample of p-bootstrap replicas; gray curves: approximate analytical CDFs computed from a subsample of the p-bootstrap replicas. The curve index α refers to the Gamma parameter with which the samples have been drawn.

As for computational loads, we compare p-bootstrap with the Bayesian procedures, since both base their estimates on a population of replicas. If we assume equivalent the costs of inverting either h'_α in (5) or the conditional distributions of A given β and $\{x_1, \dots, x_m\}$ in Son and Oh (2006), the computational complexities of the procedures as a function of the sample size and number of replicas are equal. On the contrary, the analytic solution (11) for F_A through the Gamma approximation with parameters (12) and (13) is two order faster. Then, since the reduction ratio of the number of replicas is 25, we may say that the hybrid procedure is one order faster.

In order to understand the influence of the *prior* in the parameter estimate, we delved into one sample with size 200 and $\alpha = 4$, causing performance differences in line with the above overall statistics, as shown in Table 4. In this table we report both Bayesian methods to contrast our results, where ARS is the direct competitor and RS helps identify a scaling of

Method	α			β		
	Central value	Dispersion	Confidence Interval	Central value	Dispersion	Confidence Interval
RS	4.0362	0.3917	0.99 [3.1038, 5.1351]	1.0496	0.1096	0.99 [0.8072, 1.3628]
			0.95 [3.3081, 4.8123]			0.95 [0.8585, 1.2856]
			0.90 [3.4149, 4.6828]			0.90 [0.8868, 1.2414]
ARS	4.038	0.3756	0.99 [3.0931, 5.032]	1.0483	0.1063	0.99 [0.8202, 1.3703]
			0.95 [3.3059, 4.7859]			0.95 [0.8662, 1.2833]
			0.90 [3.4272, 4.6672]			0.90 [0.8903, 1.2394]
AI _{p-boot}	3.929	0.2661	0.99 [2.9899, 5.0393]	1.0694	0.0757	0.99 [0.8253, 1.4351]
			0.95 [3.2206, 4.7472]			0.95 [0.8766, 1.319]
			0.90 [3.3249, 4.5836]			0.90 [0.9052, 1.2743]
AI _{hybrid}	3.923	0.3032	0.99 [3.0134, 4.8676]	1.0674	0.0894	0.99 [0.8462, 1.4175]
			0.95 [3.1922, 4.6538]			0.95 [0.8891, 1.329]
			0.90 [3.2972, 4.5411]			0.90 [0.9125, 1.2835]

Table 4: First and second order statistics, and confidence intervals of the parameters A and B for a reconstruction experiment with $m = 200$, $\alpha = 4$, and $\beta = 1$. Central value and Dispersion columns as in Table 2. Confidence Interval format: confidence level [lower bound, upper bound].

the results, and both Algorithmic Inference methods, referring to p-bootstrap as for statistic templates and the hybrid version for a cheaper variant. With all methods the statistics are collected from $n = 5,000$ replicas. Confidence interval extremes are exactly the $\delta/2$ and $1 - \delta/2$ quantiles, respectively, of the variously computed A and B distribution functions, where $1 - \delta$ is the confidence level. The comparison of the table values are as expected. In particular, we remark that the γ_2 coefficient of the analytical approximation is still based on 40 bootstrap replicas, thus realizing a speed up factor around 250. Instead, the exam of Figure 5 highlights a notably correlation between the replicas produced by ARS. Indeed, although the overall standard deviation is greater with p-bootstrap, the graph of the replica average by blocks of 100 consecutive replicas is more concentrated around the central value with our method than with the Bayesian one. Focusing on replicas distributions, in Figures 6

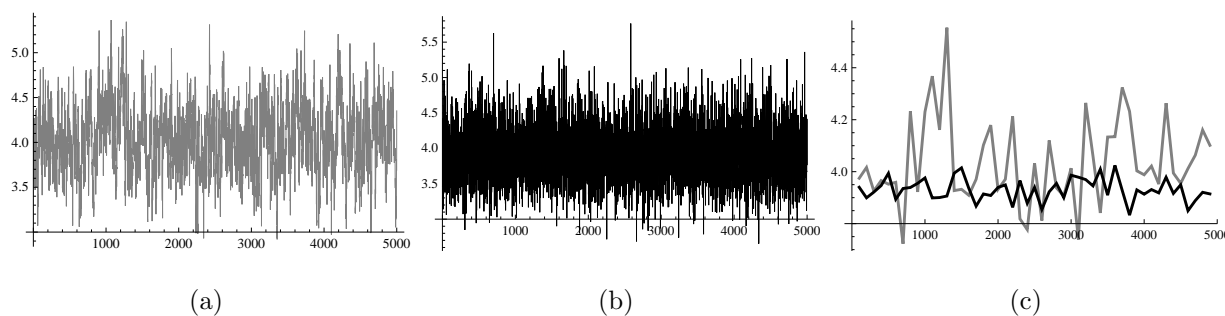


Figure 5: Courses of α replicas. (a) replicas generated by ARS; (b) replicas generated by p-bootstrap; (c) their averages by blocks of 100 consecutive replicas. Gray lines: ARS method; black lines: p-bootstrap method.

(a) and (b) we plot the 5,000 parameter estimates obtained solely with the p-bootstrap and ARS methods. We clearly recognize from the plots that distribution (17) inhibits high values of β that, *vice versa*, are acknowledged to be compatible with our method. This results in privileging high values of the α estimate (for instance in force of (18)) with the mentioned bias on their mean as a consequence.

In these figures we also draw confidence regions for the pair (α, β) that follow the thread of the parameter population (ARS regions narrower and shifted re p-bootstrap's) and in any case look more suitable than the cartesian product of the confidence intervals reported in Table 4. To get these regions we use a standard peeling procedure (Liu et al., 1999; Apolloni et al., 2007b) on the bundle of Gamma CDFs obtained by plugging the estimated parameters in their general expression. Starting from the whole population of curves in Figures 6 (c) and (d) we iteratively visited the upper and lower bounds of the region they span, circularly removing the bordering curves in number of $\delta n/2$ lying below the envelope of the remaining ones and $\delta n/2$ curves above it, having denoted with δ the preset confidence level.

Performing a similar analysis on the problem of fitting the Gross and Clark sample with a Gamma distribution lets to draw the following conclusions. Table 5 denotes a great difference between the estimates of the fitting parameters. The analysis of their scatterplots denotes a definite unbalancing of ARS parameters that completely miss the left-uppermost region

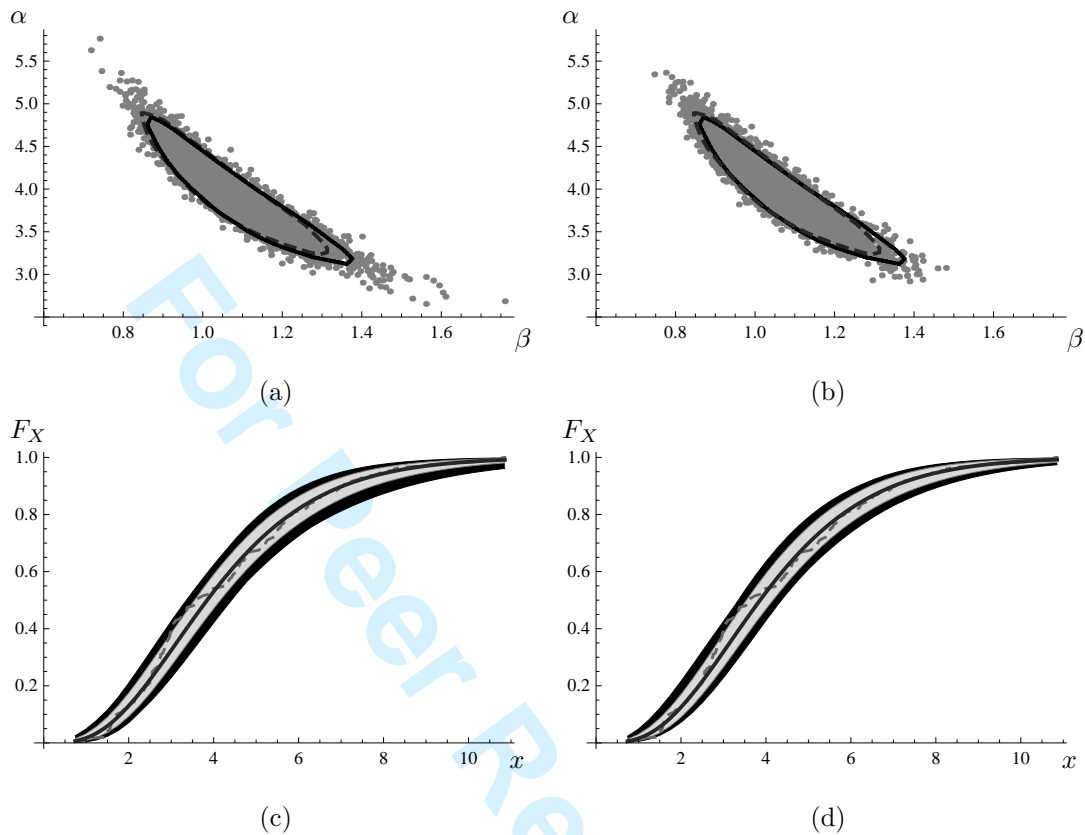


Figure 6: Distributions of Gamma parameters and related CDFs obtained with two inference approaches. Left column: Algorithmic Inference approach with p-bootstrap method; right column: Bayesian approach with ARS method. First row: parameters' scatterplot contoured by 0.90 confidence regions; plain line: contouring through p-bootstrap; dashed line: contouring through ARS. Second row: bundle of Gamma CDFs with the parameters plotted in the first row. Gray filled areas: 0.90 confidence regions, with the column method, corresponding to the companion regions in the first row.

Method	α			β		
	Central value	Dispersion	Confidence Interval	Central value	Dispersion	Confidence Interval
RS	9.2120	2.6430	0.99 [3.7912, 17.1581]	13.5246	4.4529	0.99 [6.4606, 30.5955]
			0.95 [4.7248, 14.9754]			0.95 [7.4934, 24.8940]
			0.90 [5.2521, 13.8601]			0.90 [8.0750, 22.0373]
ARS	9.3053	2.6796	0.99 [3.7041, 17.9137]	13.3868	4.5423	0.99 [6.2170, 32.3034]
			0.95 [4.6920, 15.3352]			0.95 [7.3226, 24.2030]
			0.90 [5.3633, 14.0487]			0.90 [7.9871, 21.6278]
AI _{p-boot}	13.4671	4.5923	0.99 [3.7973, 43.6594]	8.4424	2.8444	0.99 [2.5882, 30.7639]
			0.95 [4.8630, 34.1774]			0.95 [3.2778, 23.7440]
			0.90 [5.6707, 29.4574]			0.90 [3.8453, 20.3299]
AI _{hybrid}	12.6911	4.9257	0.99 [3.4563, 37.3762]	8.9295	3.6857	0.99 [3.0293, 35.0587]
			0.95 [4.6995, 29.651]			0.95 [3.8254, 24.3354]
			0.90 [5.4986, 26.188]			0.90 [4.2936, 20.6361]

Table 5: Same statistics as in Table 4 collected on G&C dataset.

of the parabolic-like distribution. This imbalance, which is opposite to the one studied before, has another explanation. The histogram in Figure 7 clearly shows that the Gross and Clark sample has a very low likelihood when it is attributed to a Gamma distribution. Actually, we do compute the ML estimate, for instance with the approximation proposed in Thom (1958), of these parameters. However, if we generate new samples exactly with these parameters and compute their likelihood, we realize that the likelihood of Gross & Clark sample represents around the fifth quantile of the likelihood distribution, thus getting a far location from the core in the histogram shown in the figure. In other words, as a sample of a Gamma distribution, it is a rather anomalous sample. Thus the p-bootstrap method gives rise to a broad parameter distribution that maintains, however, a due symmetry, as shown in Figure 8 (a) and (b). The plots of the distribution bundle denote a wider sheaf and confidence region with our method (see Figure 8 (c) and (d)). But, once again, we expect this to be rewarded by the unbiasedness of estimates, which in this case are definitely far from those obtained with conventional methods. As a final remark, the numerical results are taken on the basis of 10,000 parameter replicas, while hybrid method used only 200 replicas,

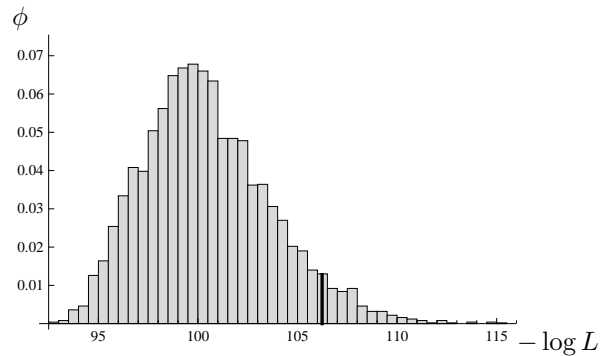


Figure 7: Histogram of the negative log-likelihood of samples drawn from a Gamma distribution with parameters given by the ML_{Thom} estimates from G&C dataset. Thick bar: likelihood of G&C dataset with this parameterization.

entailing a great saving of running time. Moreover, the γ_1 expression (16) derives from data interpolation when m uniformly ranges in $[20, 300]$, α in $[2, 6]$, and $\beta = 1$. Thus this case highlights a great extendability of the interpolated function to a range not covered by the training sample, even though closer results to full bootstrap method's may be expected with a more focused interpolation.

4. CONCLUSIONS

Facing the non trivial inference problem of estimating the two parameters of a Gamma distribution, we propose an estimator exhibiting the distinguishing feature of being unbiased, while a large set of reputed estimators does not enjoy the same property, with a further drawback of proving less accurate in a set of reconstruction problems. To obtain this result we adopted a different statistical framework showing the comparative benefit w.r.t. the Bayesian approach of not needing a prior parameter distribution. *Per se* contrasting the two approaches may appear no different from the Lilliputian contest described by Jonathan Swift in Gulliver Travels re *the way of breaking eggs: upon the larger or smaller end*. In place of the *six rebellions raised on that account* by Lilliputians, we just claim that missing prior distributions may improve the statistical properties of the estimators. This reopens a major dispute fought in the mid 20th century about the interpretation of parameter variability

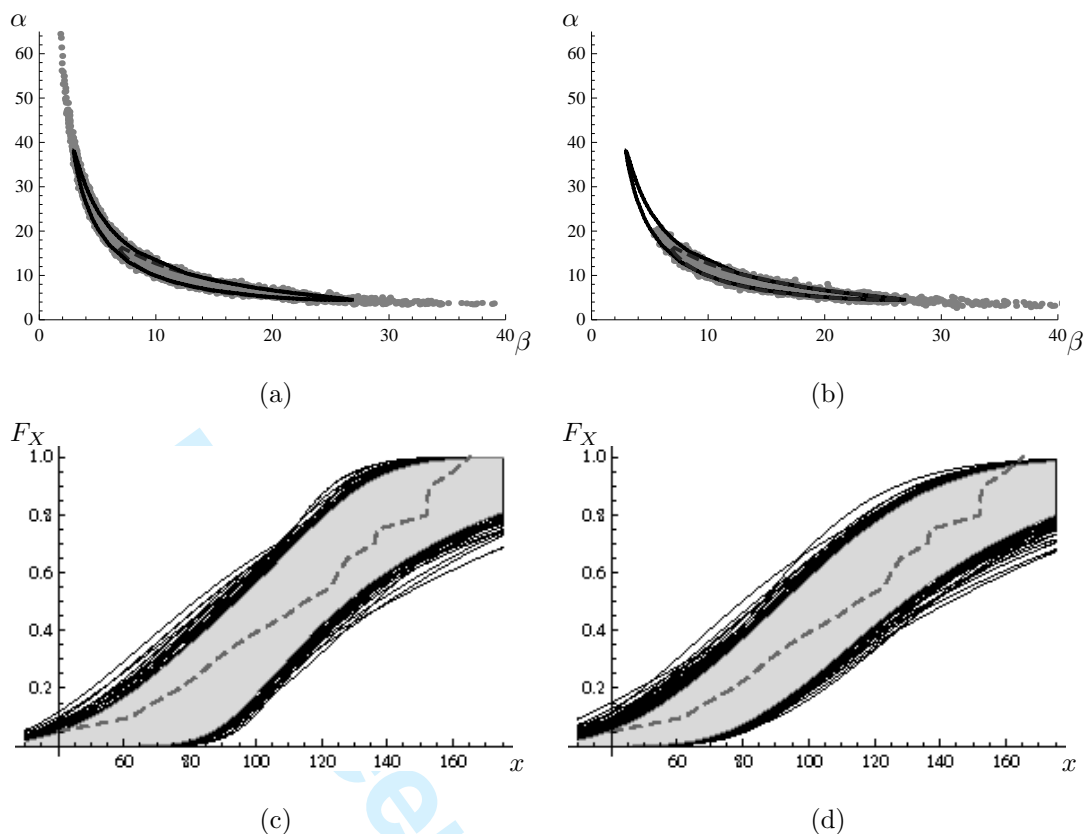


Figure 8: Same pictures and notation as in Figure 6 when the sample is G&C dataset.

(Zabell, 1992) in terms of *fiducial* distribution (Fisher, 1956), *structural* probabilities (Fraser, 1966), *priors/posteriors* (Ramsey, 1925), and so on. From an epistemologic viewpoint, this entailed a companion dispute as to the way of modeling probability: as a physical feature of the phenomena to be described through random variables or a tool for synthesizing data about a phenomenon? For a long time, the former model won out for all practical purposes. However the results we obtain in this way may prove narrow, in that almost harmless when the candidate models to explain data cannot involve Gaussian probability distributions or asymptotic limit theorems (Meeker and Escobar, 1995). Bayes approach offers a way to escape this drawback, by considering *a priori* parameter distribution laws – sometimes linked to the features of the subtending phenomena (Florens et al., 1990; Berger, 1999), other times just as an analytical artifact (the *uninformative* priors (Jaynes, 1968)) – which are modified

1 by the information content of the observed data. Very often this proves to be only the shift
2 of the hard part of the inference problem one step back, i.e. to the answer to the question:
3 Who tells us the priors?
4
5

6 We propose here and in some companion papers in the frame of computational learning
7 (Apolloni et al., 2007c, 2008b,a, 2007b; Apolloni and Bassis, 2008; Apolloni et al., 2006,
8 2008c) to return to the original thought of Fisher of considering the model parameters as
9 random variables *in light of the observed data* (Fisher, 1956). With this perspective, he
10 defines a fiducial distribution law of parameters of a given random variable that he deduces
11 from a sample of its realizations. Through this law he computes, for instance “the probability
12 that μ (mean of a Gaussian variable – our note) is less than any assigned value, or the
13 probability that it lies between any assigned values, or, in short, its probability distribution,
14 in the light of the sample observed” (Fisher, 1935). The Achilles heel of Fisher’s approach
15 lies in the joint distribution of more than one parameter (Zabell, 1992). In this paper
16 we show that this is not longer true with our revisitation, thanks to the wide availability
17 of computational power. Through (possibly heavy) numerical procedures indeed, we may
18 invert master equations like (5) to reverse the probability masses we jointly distribute on the
19 variable realizations for set parameters – a well assessed attitude – into the joint probability
20 distribution of the parameters compatible with the observed data. The compatibility notion
21 is very close to the fiducial probability or the structural probability proposed respectively by
22 Fisher (1956) and Fraser (1961). In this paper we both supply algorithmic procedures for
23 its computation and stress its properties on a severe reconstruction problem.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 References

- 44
45 H. Ammeter. The solvency problem in risk life insurance. *Mitteilungen der Vereinigung*
46 *Schweizerischer Versicherungsmatematiker*, 70(2):250–258, 1970.
47
48
49
50 B. Apolloni and S. Bassis. Algorithmic inference: from information granules to subtending
51 functions. *Nonlinear Analysis*, 2(2):648–664, 2008.
52
53
54

- 1 B. Apolloni and S. Bassis. A new goodness-of-fit statistical test. *Intelligent Decision Tech-*
2 *nologies*, 1:1–14, 2007.
- 3
4
5
6 B. Apolloni, D. Malchiodi, and S. Gaito. *Algorithmic Inference in Machine Learning*. Ad-
7 vanced Knowledge International, Magill, Adelaide, 2nd edition, 2006.
- 8
9
10 B. Apolloni, S. Bassis, and S. Gaito. Fitting opportunistic networks data with a pareto
11 distribution. In *Knowledge-Based Intelligent Information and Engineering Systems*, vol-
12 ume 4694/2007 of *Lecture Notes in Computer Science*, pages 812–820, Berlin/Heidelberg,
13 2007a. Springer.
- 14
15
16
17
18 B. Apolloni, S. Bassis, S. Gaito, and D. Malchiodi. Appreciation of medical treatments by
19 learning underlying functions with good confidence. *Current Pharmaceutical Design*, 13
20 (15):1545–1570, 2007b.
- 21
22
23
24
25 B. Apolloni, S. Bassis, S. Gaito, and D. Malchiodi. Solving complex regression problems
26 via algorithmic inference: a new family of bootstrap algorithms. *Far East Journal of*
27 *Theoretical Statistics*, 22(2):141–180, 2007c.
- 28
29
30
31 B. Apolloni, S. Bassis, S. Gaito, and D. Malchiodi. Bootstrapping complex functions. *Non-*
32 *linear Analysis*, 2(2):665–683, 2008a.
- 33
34
35
36 B. Apolloni, S. Bassis, D. Malchiodi, and W. Pedrycz. Interpolating support information
37 granules. *Neurocomputing*, 71:2433–2445, 2008b.
- 38
39
40
41 B. Apolloni, S. Bassis, D. Malchiodi, and P. Witold. *The Puzzle of Granular Computing*,
42 volume 138 of *Studies in Computational Intelligence*. Springer Verlag, 2008c.
- 43
44
45 J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Verlag, New York,
46 2nd edition, 1999.
- 47
48
49
50 K. O. Bowman and L. R. Shenton. *Properties of Estimators for the Gamma Distribution*.
51 Marcel Dekker, New York, 1988.
- 52
53
54
55
56
57
58
59
60

- 1 M. A. Chaudhry and S. M. Zubair. *On a Class of Incomplete Gamma Functions with*
2 *Applications*. Chapman & Hall CRC, 2001.
- 3
4
5
6 C. A. Coelho and J. T. Mexia. On the distribution of the product and ratio of independent
7 generalized gamma-ratio random variables. *The Indian Journal of Statistics*, 69(2):221–
8 255, 2007.
- 9
10
11 Renesas Technology Corp. Reliability handbook, 2009. URL [http://www.renesas.com/
12 fmwk.jsp?cnt=reliability_handbook.jsp&fp=/products/common_info/reliability/
13 handbook/](http://www.renesas.com/fmwk.jsp?cnt=reliability_handbook.jsp&fp=/products/common_info/reliability/handbook/).
- 14
15
16
17
18 B. Efron and R. Tibshirani. *An introduction to the Bootstrap*. Chapman and Hall, Freeman,
19 New York, 1993.
- 20
21
22
23 M. A. Fisher. The fiducial argument in statistical inference. *Annals of Eugenics*, 6:391–398,
24 1935.
- 25
26
27
28 M. A. Fisher. *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh and
29 London, 1956.
- 30
31
32
33 J. P. Florens, M. M. Mouchart, and J. M. Rolin. *Elements of Bayesian statistics*. Marcel
34 Dekker, Inc., New York, 1990.
- 35
36
37
38 D. A. S. Fraser. The fiducial method and invariance. *Biometrika*, 48(3/4):261–280, 1961.
- 39
40
41
42 D. A. S. Fraser. Structural probability and a generalization. *Biometrika*, 53(1/2):1–9, 1966.
- 43
44
45
46 M. Garg, V. Katta, and M. K. Gupta. The distribution of the products of powers of gener-
47 alized dirichlet components. *Kyungpook Mathematical Journal*, 42:429–436, 2002.
- 48
49
50
51 W. R. Gilks and Wild. P. Adaptive rejection sampling for gibbs sampling. *Applied Statistics*,
52 41:337–348, 1992.
- 53
54
55
56 J. A. Greenwood and Durand. D. Aids for fitting the gamma distribution by maximum
57 likelihood. *Technometrics*, 2:55–65, 1960.
- 58
59
60

- 1 A. Gross and V. Clark. *Survival Distribution: Reliability Applications in the Biomedical*
2 *Sciences*. John Wiley & Sons, New York, 1975.
- 3
4
5
6 E. T. Jaynes. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*,
7 4:227–241, 1968.
- 8
9
10 J. M. Juran and A. B. Godfrey. *Juran's Quality Handbook*. McGraw-Hill Professional Pub-
11 lishing, 5th edition, 1999.
- 12
13
14
15 R. Y. Liu, J. M. Parelius, and K. Singh. Multivariate analysis by data depth: Descriptive
16 statistics, graphics and inference. *The Annals of Statistics*, 27:783–858, 1999.
- 17
18
19
20 W.Q. Meeker and L. A. Escobar. Teaching about approximate confidence regions based on
21 maximum likelihood estimation. *The American Statistician*, 49(1):48–53, 1995.
- 22
23
24
25 S. Nadarajah and S. Kotz. Exact and approximate distributions for the product of dirichlet
26 components. *Kybernetika*, 40:735–744, 2004.
- 27
28
29
30 F. P. Ramsey. The foundations of mathematics. *Proc. of London Mathematical Society*,
31 1925.
- 32
33
34
35 B. D. Ripley. *Stochastic Simulation*. John Wiley & Sons, New York, 1988.
- 36
37
38
39 M. Rosenblatt. Remarks on a multivariate transformation. *The Annals of Mathematical*
40 *Statistics*, 23(3):470–472, 1952.
- 41
42
43
44 Y. S. Son and M. Oh. Bayesian estimation of the two-parameter gamma distribution. *Com-*
45 *munications in Statistics - Simulation and Computation*, 35:285–293, 2006.
- 46
47
48
49 S. Stigler. Studies in the history of probability and statistics. XXXII: Laplace, Fisher and
50 the discovery of the concept of sufficiency. *Biometrika*, 60(3):439–445, 1973.
- 51
52
53
54
55
56
57
58
59
60 H. C. Thom. A note on the gamma distribution. *Monthly Weather Review*, 86:117–122,
1958.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- J.L. Titus and C.F. Wheatley. Experimental studies of single-event gate rupture and burnout in vertical power MOSFETs. *IEEE Transactions on Nuclear Science*, 43(2):533–545, 1996.
- E. G. Tsionas. Exact inference in four-parameter generalized gamma distributions. *Communications in Statistical Theoretical Methods*, 30(4):747–756, 2001.
- D. P. Wiens, J. Cheng, and N. C. Beaulieu. A class of method of moments estimators for the two-parameter gamma family. *Park. Journal of Statistics*, 29(1):129–141, 2003.
- Wolfram Research Inc. Mathematica 7, 2008. URL <http://www.wolfram.com/products/mathematica/index.html>.
- S. L. Zabell. R. A. Fisher and the fiducial argument. *Statistical Science*, 7(3):369–387, 1992.
- M. Zhu and Y.-G. Wang. Quantile estimation from ranked set sampling data. *The Indian Journal of Statistics*, 67(2):295–304, 2005.

1
2
3 Response to referee comments in the paper:
4

5 ID LSSP-2009-0057, Algorithmic Inference of Two-Parameter Gamma Distribution
6

7 *General
8

9 We have revised the paper according to the suggestions of the referee, whom we thank
10 sincerely for his consideration.
11

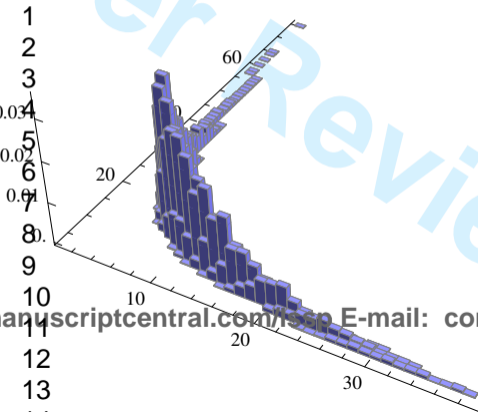
12 Specifically, the changes we have made on the basis of the referee's comments are as
13 follows:
14

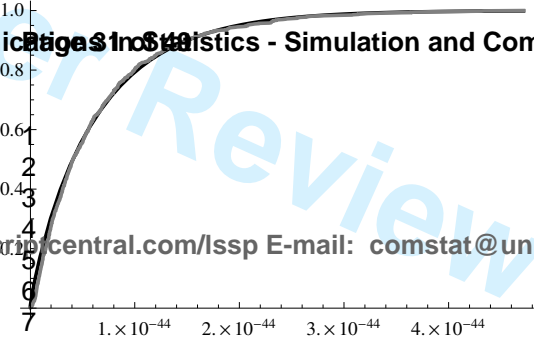
15 C(omment). Different seeds than uniform variable; the rationale of "seed" notion.
16

17 A(nswer). On page 4, line 10 and ff. we mention that the standard Gaussian variable is
18 very often used as a seed in the practice of variable simulation. We also explain the
19 benefit of introducing the notion of seed as a link between parameter and variable
20 randomness.
21
22

23 C. s_3 monotony.
24

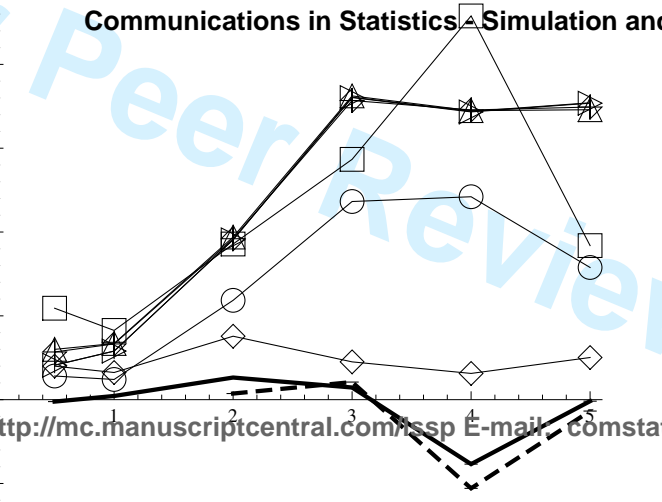
25 A. On page 7, line -11 we clarify that we numerically realized this property. Actually, we
26 already remark on page 5, line 11, that (s_1, s_3) constitute a different representation of
27 the jointly sufficient statistics (s_1, s_2) , but we have no direct analytical proof of the s_3
28 monotony with α .
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

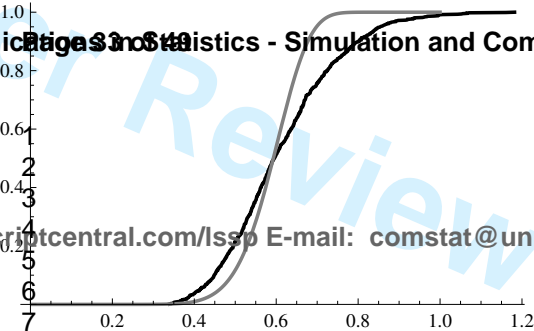


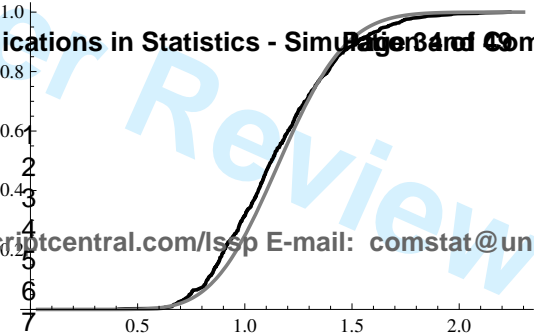


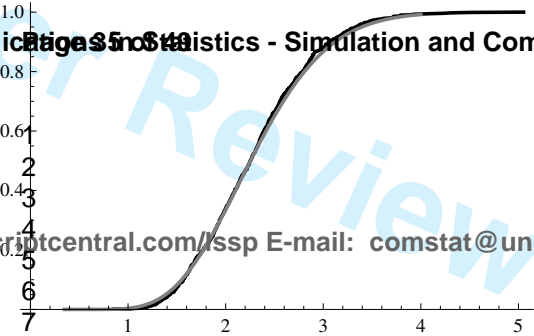
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4

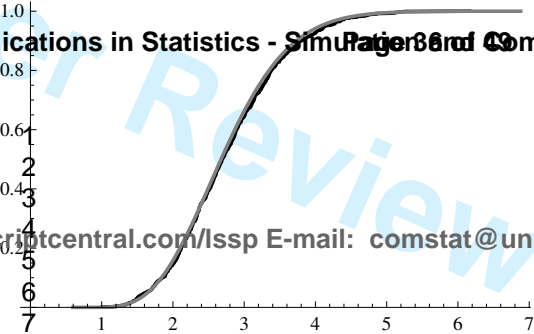
- WCB
- MM
- △ ML_{Thom}
- ▽ ML_{GD}
- + ML_{BS}
- ✦ RS
- ◇ ARS
- p-boot
- - - hybrid

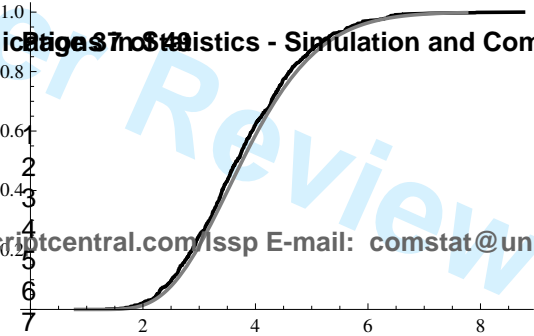


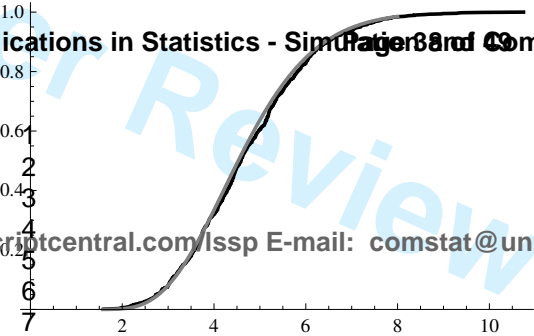




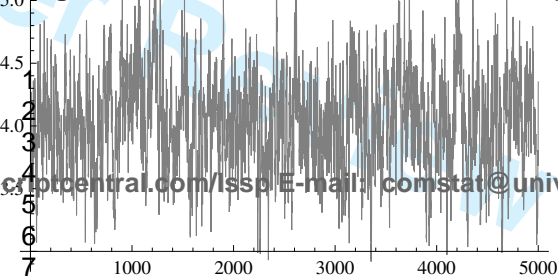








Journal of Statistics - Simulation and Computation



cryptcentral.com/issp E-mail: comstat@univ

Communications in Statistics - Simulation Page 40 of 49

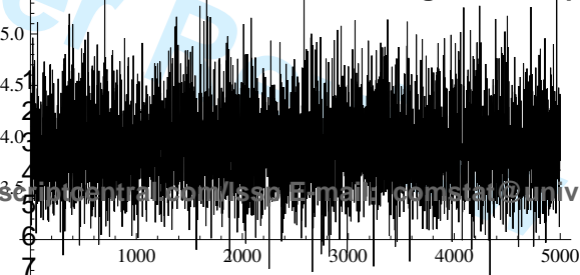
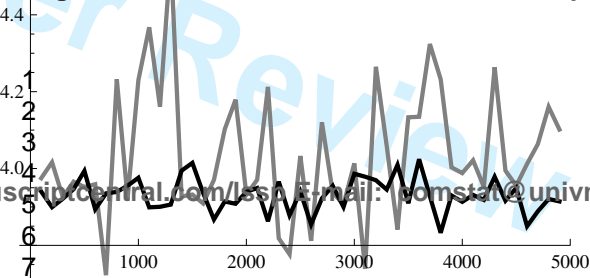
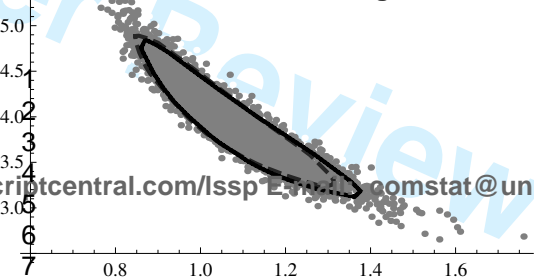


Figure 4.9: Comparison of Statistics - Simulation and Computation

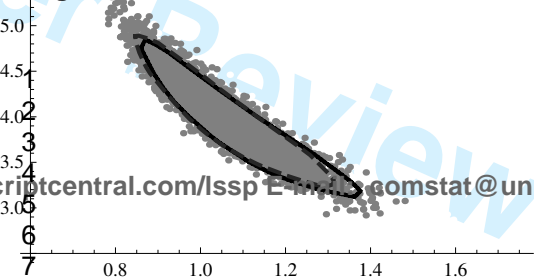


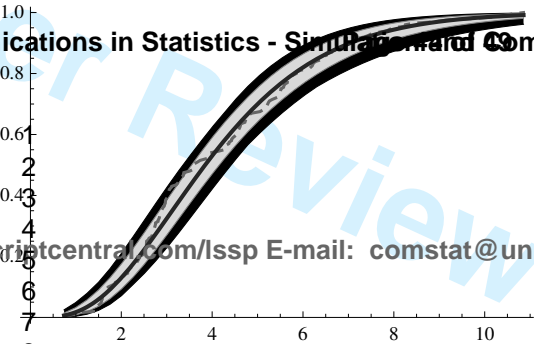
scriptcentral.com/lssp E-mail: comstat@univ

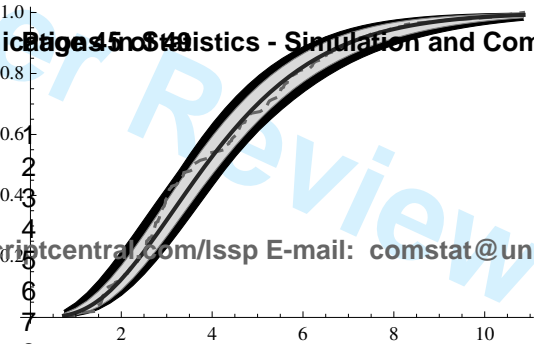
Applications in Statistics - Simulation 42nd Com

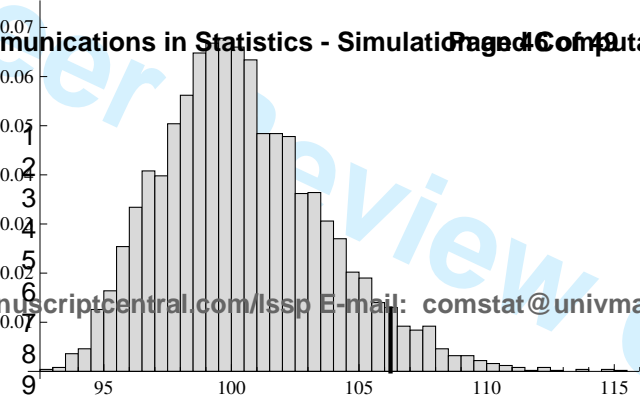


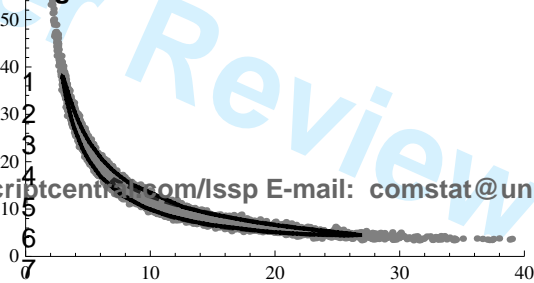
Chapter 10 Statistics - Simulation and Com

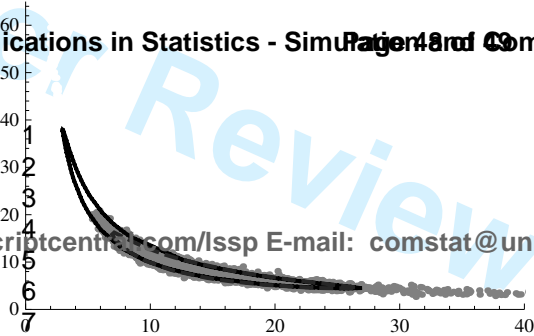












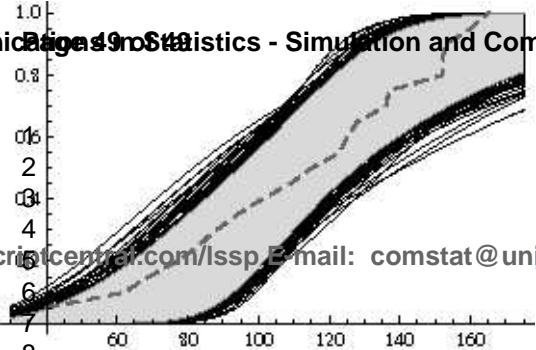
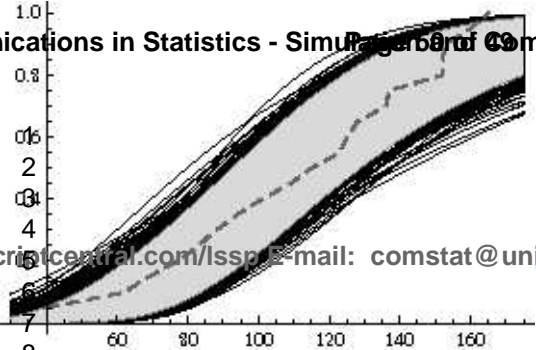


Figure 19. Statistics - Simulation and Comparison

critcentral.com/Issp E-mail: comstat@un



Communications in Statistics - Simulation and Computation

http://www.tandf.co.uk/journals/ISSN/0162-1459 E-mail: comstat@univ