

A Simple Approximation to the Convolution of Gamma Distributions (Revision of DP 2006-27)

Stewart, T.; Strijbosch, L.W.G.; Moors, J.J.A.; van Batenburg, P.

Publication date:
2007

[Link to publication](#)

Citation for published version (APA):

Stewart, T., Strijbosch, L. W. G., Moors, J. J. A., & van Batenburg, P. (2007). *A Simple Approximation to the Convolution of Gamma Distributions (Revision of DP 2006-27)*. (CentER Discussion Paper; Vol. 2007-70). Operations research.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.



No. 2007–70

**A SIMPLE APPROXIMATION TO THE CONVOLUTION OF
GAMMA DISTRIBUTIONS**

By Trevor Stewart, Leo Strijbosch, Hans Moors,
Paul van Batenburg

September 2007

This is a revised version of CentER Discussion Paper
No. 2006-27

April 2006

ISSN 0924-7815

A SIMPLE APPROXIMATION TO THE CONVOLUTION OF GAMMA DISTRIBUTIONS

Trevor Stewart (Deloitte & Touche LLP)

Leo Strijbosch*(Tilburg University)

Hans Moors (Tilburg University)

Paul van Batenburg (Deloitte Accountants B.V. and University of Warwick)

Abstract

The exact expression for the convolution of gamma distributions with different scale parameters is quite complicated. The approximation by means of another gamma distribution is shown to be remarkably accurate for wide ranges of the parameter values, especially if more than two random variables are involved. The approximation is particularly good for the upper quantiles that play an important role in auditors' decisions.

JEL-code: C16

Keywords: approximating distributions, convolution, gamma distribution

*Corresponding author. Tel.: + 31 13 4663233, fax: +31 13 4663280. E-mail address: L.W.G.Strijbosch@uvt.nl. Postal address: Tilburg University, Warandelaan 2, P.O.Box 90153, 5000 LE, Tilburg, The Netherlands.

1 Introduction

Consider r.v.'s X_i having gamma distributions with shape parameter $\alpha_i > 0$ and scale parameter $\beta_i > 0$; notation

$$X_i \sim \Gamma(\alpha_i, \beta_i).$$

Then the pdf of X_i reads

$$p(x_i) = \frac{x_i^{\alpha_i-1}}{\beta_i^{\alpha_i} \Gamma(\alpha_i)} e^{-x_i/\beta_i}, x_i > 0$$

and mean and variance equal

$$\mu_i = E(X_i) = \alpha_i \beta_i, \quad \sigma_i^2 = V(X_i) = \alpha_i \beta_i^2.$$

A convolution of two or more independent random variables is the probability distribution of the sum of those variables. Convolutions of gamma-distributed r.v.'s occur often: a brief overview from the literature follows. Moschopoulos (1985) gives two examples: queuing problems where total waiting time is composed of individual waiting times having gamma distributions; and engineering problems, like total excess water flow, built up by gamma-distributed individual excess flows. Sim (1992) discusses point processes with interarrival times having gamma densities. Thom (1968) and recently Fleurant et al. (2004) present applications regarding precipitation data and the morphometric description of (the lengths of) trees, respectively. To this list we add production processes where each production stage has a gamma-type duration; inventory models where demand per period has a gamma density; and auditing, where gamma distributions may be used to describe the potential misstatement of financial statements. In Bayesian inference, applications may be even more abundant, since, apart from the normal, the gamma is the most widespread prior. E.g., Mathai & Moschopoulos (1991) and Tsonas (2004) used convolutions of gamma r.v.'s to construct multivariate gamma distributions. Note that, whenever an exponential distribution is appropriate, the more general gamma distribution may be useful as well.

In the special case that all n individual X_i have the same scale parameter β , it is well-known that the convolution is

$$\Gamma\left(\sum_{i=1}^n \alpha_i, \beta\right) \tag{1}$$

For the general case of unequal β_i , the convolution is not gamma and a simple expression for the pdf is not known. Both Moschopoulos (1985) - based upon Mathai (1982) - and

Sim (1992) express the exact convolution as an infinite sum. As only a finite number of terms can be calculated in practice, a small rounding error remains. Simulation is another method to find the nearly-exact convolution: great precision can be achieved very fast nowadays. And recently, Akkouchi (2005) expressed the cdf as a multiple integral; numerical integration would again lead to small rounding errors.

If the convolution is an end in itself, these methods are quite satisfactory. If it is just a step in a continuing argument, however, a simple closed expression to approximate the convolution would be much more useful than a numerical method, even at the cost of a small approximation error. Thom (1968) used a normal approximation; we here propose as new approximation the gamma distribution with the exact mean and variance. Keeping the argument within the family of gamma distributions of course has many theoretical advantages.

More precisely, denote

$$Y = X_1 + X_2 + \dots + X_n$$

with independent $X_i \sim \Gamma(\alpha_i, \beta_i)$; mean and variance equal

$$\mu = E(Y) = \sum_{i=1}^n \alpha_i \beta_i, \quad \sigma^2 = V(Y) = \sum_{i=1}^n \alpha_i \beta_i^2$$

Then we propose the following approximate distribution for Y :

$$\Gamma(\alpha, \beta), \text{ with } \alpha = \mu^2/\sigma^2, \quad \beta = \sigma^2/\mu \quad (2)$$

Note that (2) reduces to (1) in case $\beta_i = \beta$ for all i . Throughout the paper we will take $\beta_1 = 1$. Since the β_i are scale parameters, this causes no loss of generality.

In what follows we will show that (2) approximates the convolution surprisingly well if the scale parameters β_i differ by no more than a factor of 10 and if the shape parameters α_i are not too small. Furthermore, the approximation improves if n increases. This holds for the cdf in general, and in particular in the neighbourhood of the 95th-percentile - the most important feature in statistical inference - where the error is usually no more than 0.1 percentage point.

This paper originated from a Bayesian audit assurance model, developed by Stewart (2007). This model is described below in some more detail.

An auditor's professional judgment about the potential misstatement of a financial population can usually be satisfactorily represented by a gamma probability distribution. In a Bayesian sense, the audit process consists of an initial prior that is transformed by the gathering and analysis of successive layers of evidence until the auditor has

achieved a reasonable level of assurance (usually taken to mean 95% confidence) that misstatement is not material (i.e., would not be enough to matter to the users of the financial statements). Since evidence is typically provided by statistical samples in which the number of observed misstatements is Poisson-distributed, the fact that the gamma and Poisson distributions are natural conjugates keeps the model describing this audit process within the gamma family.

In the audit of a group of entities, the group financial statements are the aggregation (consolidation) of the financial statements of the component entities. In many cases those entities are separately and independently audited. This leads naturally to consideration of the aggregate misstatement, and this in turn leads to a consideration of the convolution of gamma distributions. Component entities may themselves be groups of sub-components.

Because audit work may be performed and evidence obtained at any level, prior and posterior distributions may be associated with each component and may aggregate with others up to the next level. At the audit planning stage, the process works in reverse as parameters are established for audit work at various levels such that, if the audit goes as expected, the auditor will have obtained reasonable assurance that the group financial statements are not materially misstated.

Because the component-level probability distributions in this application are not ends in themselves but part of a continuing argument and aggregation process, it is very convenient to be able to regard each convolution as a gamma distribution. And it is even worth sacrificing some accuracy in the process, particularly since a large amount of subjective professional judgment enters into the establishment of the priors in the first place.

The nature of this auditing application is such that the gamma distributions entering a convolution are seldom if ever extremely different from each other. For example, the scale parameters, which are essentially functions of the intensity of the work performed on the participating components, will seldom differ by more than a factor of ten. And the shape parameters associated with the lowest-level components, which are largely a function of the most likely misstatement, also tend not to be extremely different or to be very high. This provides some further comfort that in this application one cannot go too far wrong by approximating highly complex and intractable convolutions by gamma distributions with appropriate shape and scale parameters.

The organization of the remainder of the paper is as follows. In Section 2, we compare skewness and kurtosis of the exact distribution and our approximation; and we

show that for $n = 2$, differences are caused mainly by very different scale parameters. We also argue that the approximation generally improves with increasing n . For this reason attention will be concentrated on the case $n = 2$ in Section 3, where Sim's exact expression is presented and compared with our approximation for both the pdf and cdf. Moschopoulos' method was used to check the outcomes. In Section 4 we concentrate on percentage points for general n , using simulation to establish the quality of approximation (2). Finally, Section 5 summarizes our findings.

2 Comparing skewness and kurtosis

Since our approximation (2) has by definition the correct mean and variance, it seems logical to consider first the next two central moments. Define skewness γ and kurtosis κ as

$$\gamma = E[(X - \mu)^3], \kappa = E[(X - \mu)^4]$$

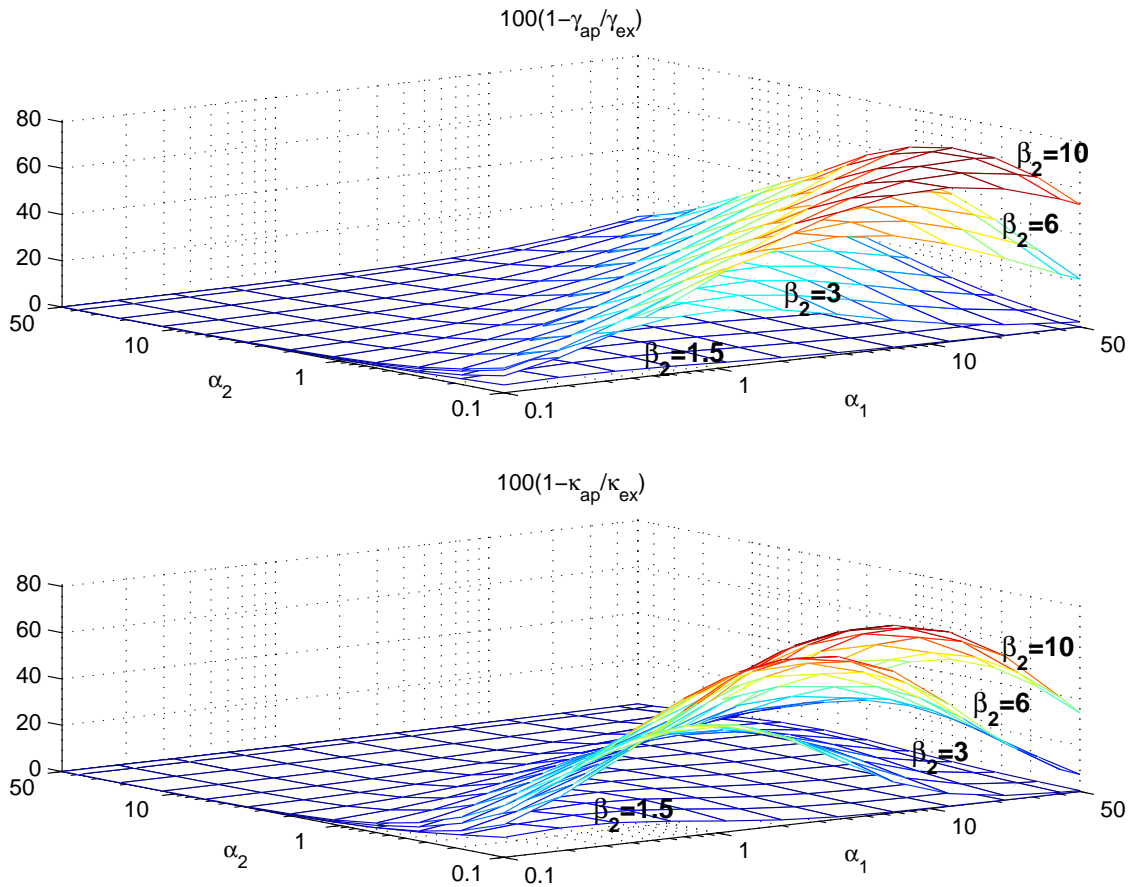
Consider the case $n = 2$, with $X_1 \sim \Gamma(\alpha_1, 1)$ and $X_2 \sim \Gamma(\alpha_2, \beta)$ and let γ_{ex} and κ_{ex} denote respectively the exact skewness and kurtosis of the convolution $X_1 + X_2$, while γ_{ap} and κ_{ap} denote the corresponding values of approximation (2). Appendix A then gives the following results ($d = \alpha_2/\alpha_1$):

$$\begin{cases} \gamma_{ex} = 2\alpha_1(1 + d\beta^3) \\ \gamma_{ex} - \gamma_{ap} = 2\alpha_1(1 - \beta)^2 d\beta / (1 + d\beta) \end{cases} \quad (3)$$

$$\begin{cases} \kappa_{ex} = 3\alpha_1^2(1 + d\beta^2)^2 + 6\alpha_1(1 + d\beta^4) \\ \kappa_{ex} - \kappa_{ap} = 6\alpha_1(1 - \beta)^2 d\beta [2 + \beta + d\beta(1 + 2\beta)] / (1 + d\beta)^2 \end{cases} \quad (4)$$

Note that γ_{ap} and κ_{ap} are always too low (for $\beta \neq 1$).

Figure 1 shows the relative differences as functions of α_1, α_2 and β . It is obvious that relative deviations are high for very low values of α_2 - in particular if α_1 is high - and for high values of the scale parameter β .

Figure 1: Relative differences $100(1 - \gamma_{ap}/\gamma_{ex})$ and $100(1 - \kappa_{ap}/\kappa_{ex})$.

Appendix A further shows that for any $\Gamma(\alpha, \beta)$ the relative skewness γ^* and kurtosis κ^* equal

$$\begin{cases} \gamma^* = \gamma/\sigma^3 = 2/\sqrt{\alpha} \\ \kappa^* = \kappa/\sigma^4 = 3 + 6/\alpha \end{cases} \quad (5)$$

So, for $\alpha \rightarrow \infty$, both parameters take the values that correspond to the normal distribution. Of course, this reflects the fact that any gamma distribution with high shape parameter α resembles a normal quite well. Now, since the convolution of normals is normal again, it may be concluded that our approximation improves whenever α in (2) increases.

This leads us to the question: In what circumstances will the addition of a new variable increase α and hence improve the approximation? If we denote the approximation (2) of the distribution of $Y = X_1 + X_2 + \dots + X_n$ now by $\Gamma(\alpha_{old}, \beta_{old})$, adding another term $X_{n+1} \sim \Gamma(\alpha_{n+1}, \beta_{n+1})$ to Y gives the new approximation

$$\Gamma(\alpha_{new}, \beta_{new})$$

where

$$\alpha_{new} = [\alpha_{old}\beta_{old} + \alpha_{n+1}\beta_{n+1}]^2 / [\alpha_{old}\beta_{old}^2 + \alpha_{n+1}\beta_{n+1}^2]$$

Simple algebra then shows that for example the condition

$$\beta_{n+1} < 2\beta_{old}$$

is sufficient for $\alpha_{new} > \alpha_{old}$ to hold. In words: if the scale parameter of the added variable is not too large, the shape parameter of the convolution increases, and the approximation will be more accurate. For this reason we concentrate on the case $n = 2$ in the next section.

3 Exact distribution; comparing pdf and cdf

The first exact expression for the convolution of Y was given by Mathai (1982). His expression for the cases where the α_i are equal or are integers leads to an easy computation. His expression for general $\alpha_i > 0$ is much more complex and was simplified by Moschopoulos (1985) into a gamma-series representation. A different infinite series expression was given by Sim (1992). (The latter two papers are referred to in Johnson et al. (1994), but note that the formula for $b_i(r)$ on p. 384 contains two typos.)

In the sequel we will use Sim's expression for the pdf of Y , which in our notation reads

$$p(y) = \frac{1}{\Gamma(\alpha_n)} \left(\prod_{i=1}^n \beta_i^{-\alpha_i} \right) y^{\alpha_n-1} \exp\left(-\frac{y}{\beta_n}\right) \sum_{r=0}^{\infty} \frac{b_n(r) \binom{\alpha_{n-1}^{(r)}}{(\alpha_n^{(r)})} \{(\beta_n^{-1} - \beta_{n-1}^{-1})y\}^r}{r!} \quad (6)$$

where

$$\alpha_n = \sum_{i=1}^n \alpha_i,$$

$$\alpha_k^{(r)} = \Gamma(\alpha_k + r) / \Gamma(\alpha_k),$$

and

$$b_i(r) = \begin{cases} 1, & i = 2, \\ \sum_{j=0}^r \frac{b_{i-1}(j) \binom{\alpha_{i-2}^{(j)}}{\alpha_{i-1}^{(j)}} (-r)^{(j)} c_i^j}{\alpha_{i-1}^{(j)} j!}, & i = 3, 4, \dots, n \end{cases}$$

for $r = 0, 1, 2, \dots$, and finally

$$c_i = \frac{\beta_{i-2}^{-1} - \beta_{i-1}^{-1}}{\beta_i^{-1} - \beta_{i-1}^{-1}}.$$

We programmed the Sim representation in MATLAB. The main body of the program is reproduced in Appendix B; the complete program, including plotting routines, is available from the second author). An important feature of this program is to prevent numerical difficulties by performing the necessary calculations in an order that prevents the production of very large (or small) numbers. The input parameters of the program are the α_i and β_i and k , where k defines the point $\mu + k\sigma$, denoting the largest value of y for which the pdf is determined. The summation stops when successive sums differ by less than ε (we used $\varepsilon = 10^{-6}$). When all β_i 's are equal, the summation in (6) must be replaced by 1. (In general, all parameters β_i have to be ordered such that no two successive β_i 's are equal. For $n \geq 3$, a minor problem occurs in calculating the c_i when more than half of the β_i 's are equal; a practical solution for that case appears to be to fractionally change sufficiently many β_i 's.)

An alternative to using this MATLAB program is the MATHEMATICA program, which implements the Moschopoulos representation for the pdf of Y , provided by Alouini et al. (2001). We programmed the expression of Moschopoulos in MATLAB for a comparison with Sim's expression. In all cases considered, both methods yielded the same pdf-values, but the Moschopoulos' method always took substantially more CPU-time. We have therefore adopted Sim's expression as our benchmark.

Sim's exact pdf will now be compared with approximation (2); using numerical integration Sim's cdf can be obtained and compared as well. For the reasons presented at the end of Section 2 and since the calculation of (6) is rather time-consuming, we restricted ourselves to the case $n = 2$ and consider only eight parameter value combinations: $\alpha_i \in \{2, 5\}$ ($i = 1, 2$), $\beta_2 \in \{2, 10\}$. Note that for our main application scale parameter values larger than 10 are unimportant because auditors will unlikely audit one part 10 times as intensely as an other part.

Figure 2 shows the eight exact pdf's (smooth curves) as well as the values corresponding with the approximation (2). The agreement is fairly good; as was to be expected, the largest differences occur around the mode. Figure 3 is quite similar, but relates to the cdf's. Now, deviations are much smaller: many dots are hardly visible because they coincide with the smooth curve.

To show that the approximation will probably deteriorate, in particular the pdf, we include Figure 4.

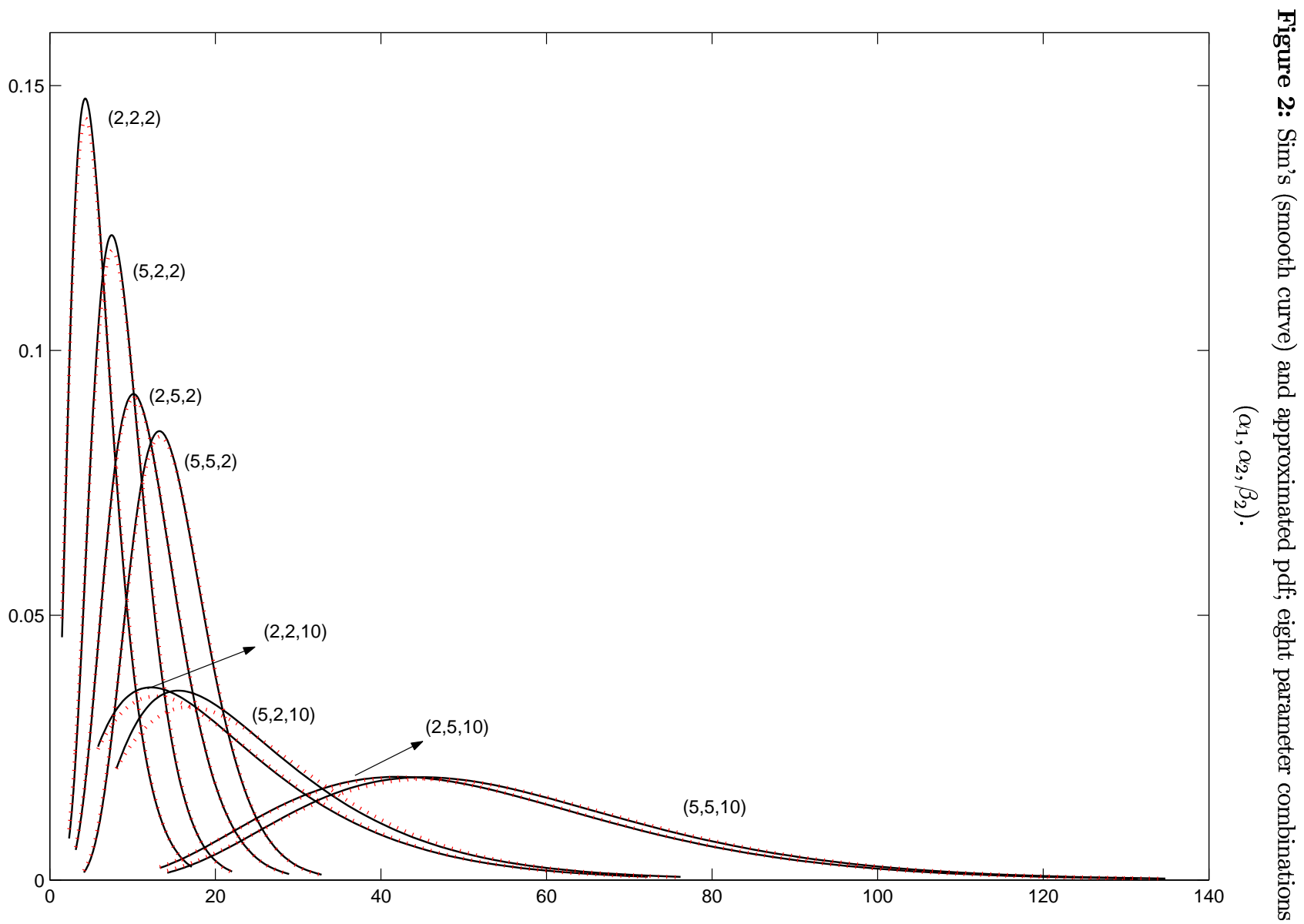


Figure 3: Sim's (smooth curve) and approximated cdf; eight parameter combinations (cf. Figure 2).

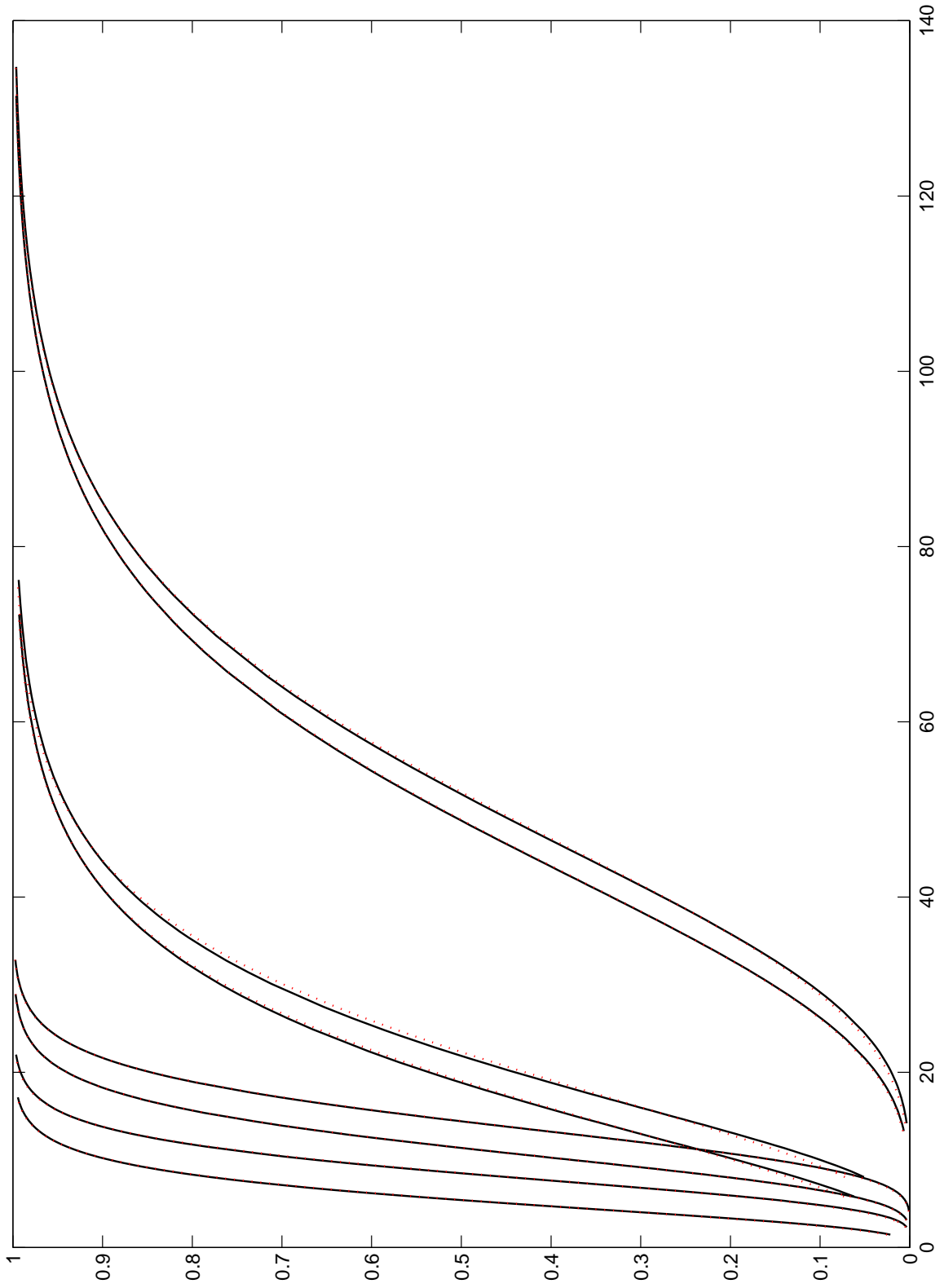
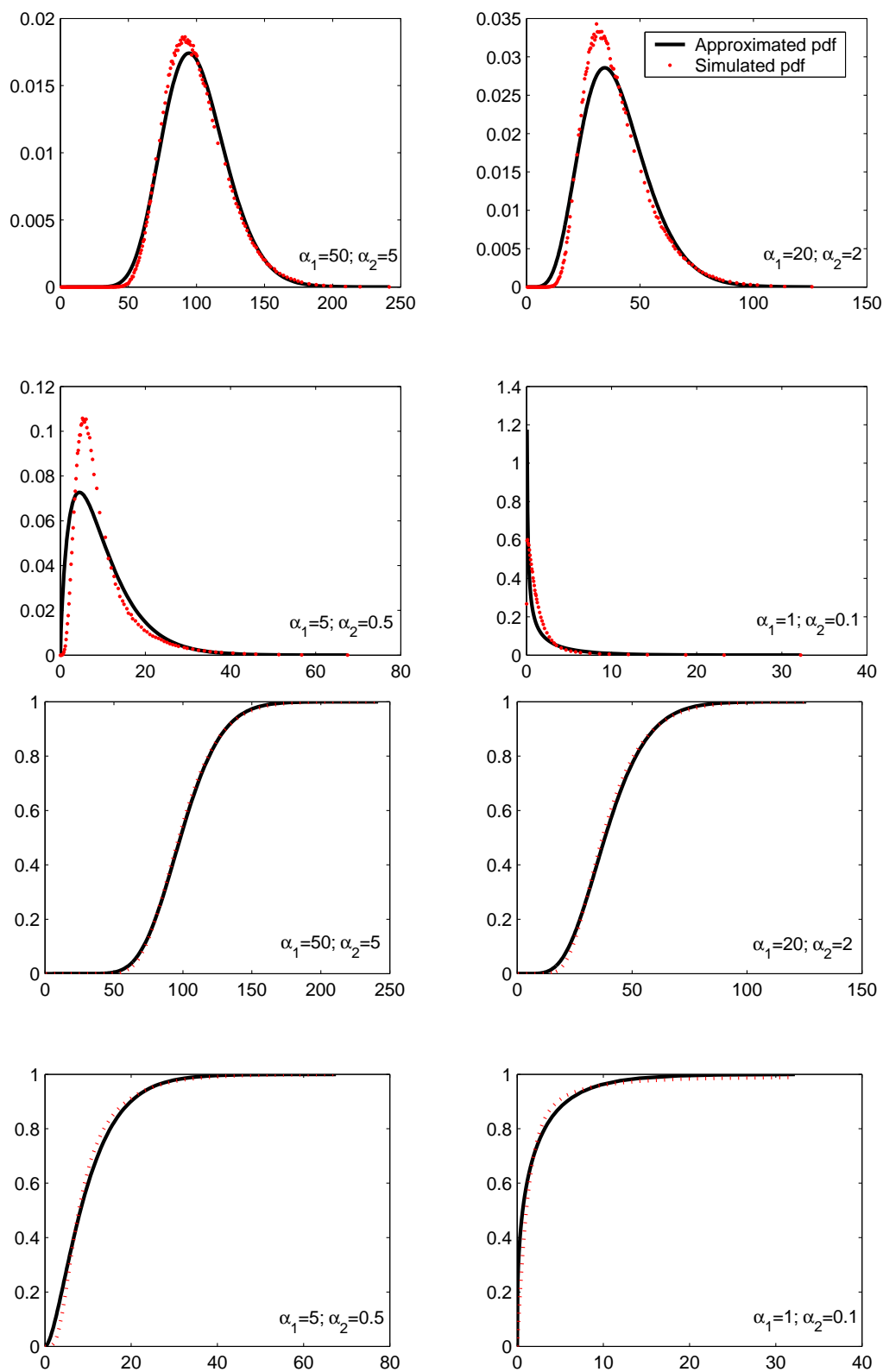


Figure 4: Pdf (top four) and cdf of approximate (solid) and simulated (dotted) distribution of the convolution when $\alpha_1/\alpha_2 = 10$, $\beta_1 = 1$, $\beta_2 = 10$.



4 Comparing percentage points

Since the calculation of Sim's expression (6) takes too much time to allow large-scale application, simulation will be used in this section. We will concentrate on the accuracy of our approximation for the 95%-point of the convolution, because this point is very frequently used in inferential statistics. Values of n up to 20 will be considered and broad combinations of parameter values.

First, we describe our simulation program and discuss its accuracy. To construct the convolution of twenty gamma distributions, we first chose twenty parameter combinations from uniform distributions. Let $U(a_1, a_2)$ denote the uniform distribution on the interval $[a_1, a_2]$. Then the parameter value combinations are given by (α_j, β_j) , $j = 1, 2, \dots, 20$ where

$$\begin{cases} \alpha_j \sim U(\alpha_L, \alpha_U) \\ \beta_1 = 1, \beta_j \sim U(1, 10), j = 2, 3, \dots, 20 \end{cases}$$

So, the β_j -values are allowed to vary with a factor 10. Further research will be necessary to study the properties of the approximated distribution for β_i differing by factors larger than 10. The intervals $[\alpha_L, \alpha_U]$ for α_j are chosen from the range $[0.1, 50]$; to be able to study a wide variety of combinations we took separate intervals for j even and odd. Figure 5 shows the seven intervals $[\alpha_L, \alpha_U]$ that were taken into consideration; the scale of the two axes is logarithmic. In this section we only consider a part of this area, indicated in the rectangles of Figure 5.

For a given parameter combination (α_j, β_j) , $j = 1, 2, \dots, 20$, the parameters (α, β) of the approximating convolution (2) are found; its 95%-point is indicated as q_{95} . For any particular parameter vector (α_j, β_j) , $j = 1, 2, \dots, 20$, $N = 10000$ drawings of $(X_j)_{j=1, \dots, 20}$ are obtained, where $X_j \sim \Gamma(\alpha_j, \beta_j)$; this yields a $N \times 20$ -sized matrix $\{x_{ij}\}_{i=1, \dots, N; j=1, \dots, 20}$ of random gamma drawings. From the row sums $\left\{ \sum_{j=1}^{20} x_{ij} \right\}_{i=1, \dots, N} = \{s_i\}_{i=1, \dots, N}$ of the matrix we can derive the simulated cumulative probability

$$\hat{p} = \frac{1}{N} \sum_{i=1}^N \delta(s_i < q_{95})$$

where $\delta(\text{true}) = 1$ and $\delta(\text{false}) = 0$. The same matrix can be used to obtain similar results for smaller (even) n by simply deleting columns from the matrix. We did this for $n \in \{20, 10, 6, 4, 2\}$.

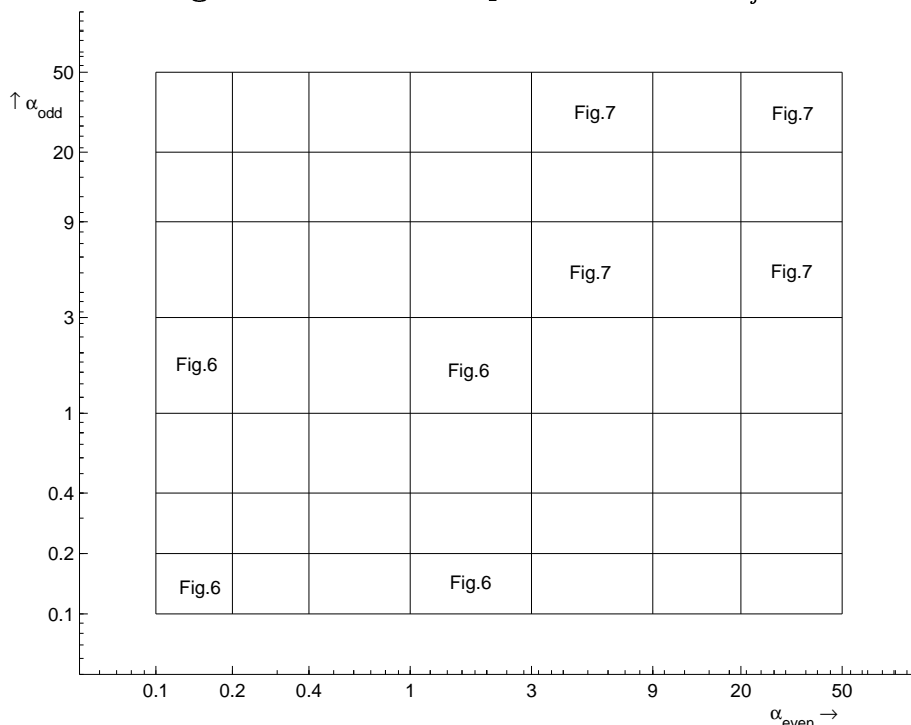
Let p denote the true probability that the convolution is smaller than q_{95} ; then \hat{p} has approximately the normal distribution with mean p and standard deviation $\sqrt{p(1-p)/N}$.

Since $p \approx 95\%$ and $N = 10000$, this standard deviation equals 0.218%. Consequently, the interquartile range of this normal distribution equals 0.294%, while the probability of an ‘outlier’ on each side of the interval

$$[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$$

is 0.35%.

Figure 5: Overview of parameter values α_j .



The complete procedure described above is repeated 1000 times (each of which is based on a new random vector (α_j, β_j) , $j = 1, \dots, 20$ of gamma distribution parameters), resulting in 1000 cdf-values, $\{\hat{p}_k\}_{k=1, \dots, 1000}$, at the 95-percentile of the corresponding pseudo distributions for $n \in \{20, 10, 6, 4, 2\}$. Although different parameter combinations are taken, the above reasoning remains valid: so we obtain 1000 observations from the approximate normal distribution with mean p and standard deviation 0.218%.

As a check of our simulations Figures 6 and 7 show boxplots of the 1000 replications. The medians M of the \hat{p}_k are in general quite close to 0.95; the largest deviations occur for small n . Note that $M > 0.95$ implies that the approximated value q_{95} is conservative. Further note that the agreement with the theoretical normal distribution is quite good; e.g. the expected number of outliers on both sides is 3.5. We observe more outliers only

for small n . Hence we conclude that the variability of the cumulative probabilities at q_{95} - due to the different parameter value combinations - is small. Hence, in the remainder we will concentrate on the medians M obtained from each set of 1000 repetitions - keeping in mind the above indicated accuracy. As a check we repeated this simulation for the median (50%-point) of the convolution; since the results are quite similar, they are not reproduced here.

To continue our discussion of the accuracy of our simulation study, we looked into the case $n = 2$ and $\beta_2 = 10$ in more detail. Figure 8 shows a three-dimensional picture of the cumulative probabilities at q_{95} according to the exact Sim's expression; Figure 9 shows the simulated values \hat{p}_k . The two figures are quite similar, apart from the irregularities caused by simulation. It is clear that (for $n = 2$) deviations are largest if in particular α_2 is small.

For all 49 rectangles in Figure 5 this simulation process was repeated. The medians M of the 1000 values \hat{p}_k (for the 1000 different parameter value combinations within this rectangle) are presented in Table 1; more precisely: the deviations $100(0.95 - M)$ are given. So, the value 0.06 in the upper left-hand corner means: $M = 0.9494$. Deviations exceeding 0.3 (in absolute value) are printed in bold type, deviations between 0.1 and 0.3 in italics. Note however, that for individual parameter combinations deviations may be somewhat larger. It is clear that most deviations are smaller than 0.1: in most cases M is between 0.949 and 0.951. Higher deviations can occur for small n and/or small α_j , the extreme value being -0.74 ($n = 2$, $\alpha_{odd} \in [1, 3]$, $\alpha_{even} \in [0.1, 0.2]$), giving $M = 0.9574$. This agrees with our results at the end of Section 2: in general, the approximation will be better for larger n .

In Appendix C we consider even lower values of α_j : between 0.01 and 0.1.

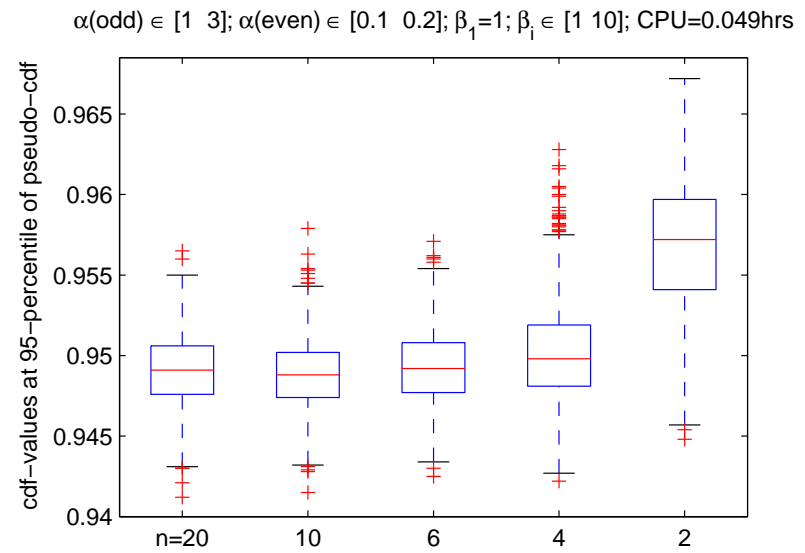
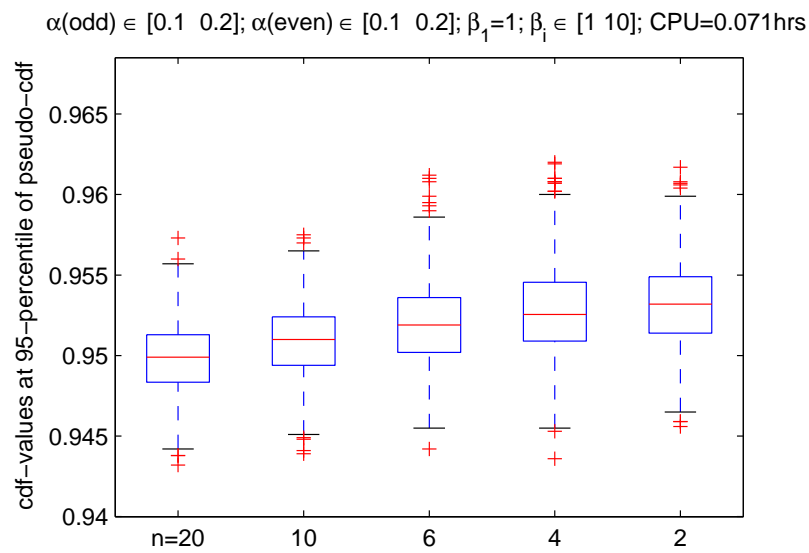
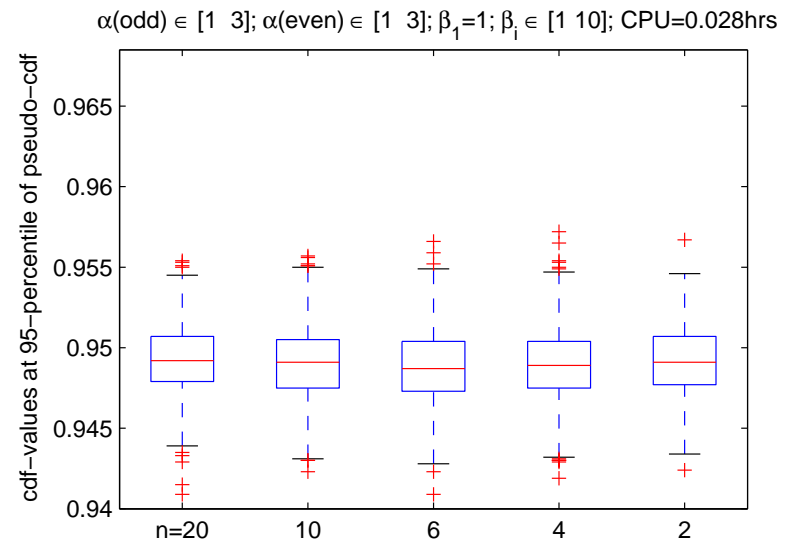
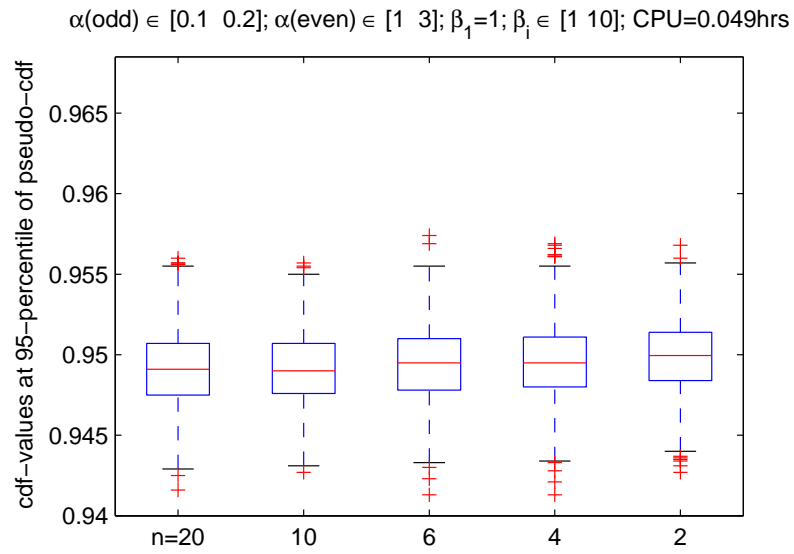


Figure 6: Boxplot $\hat{p}_k; 0.1 \leq \alpha_j \leq 3$.

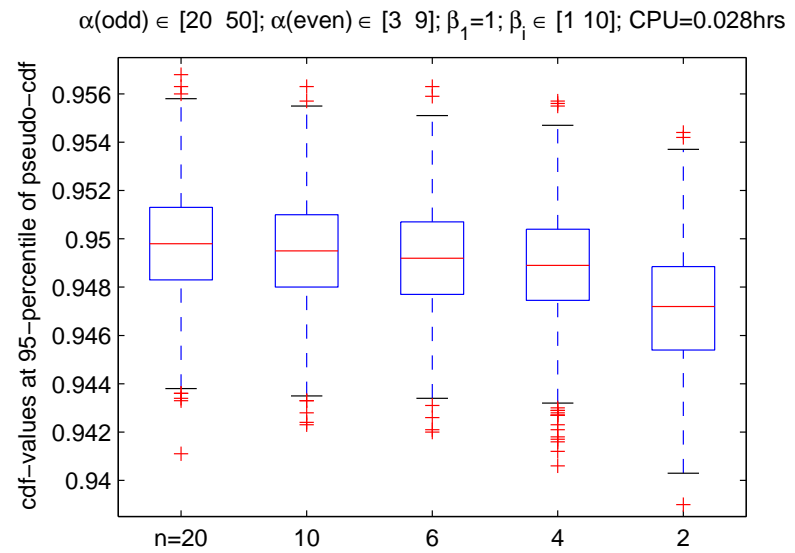
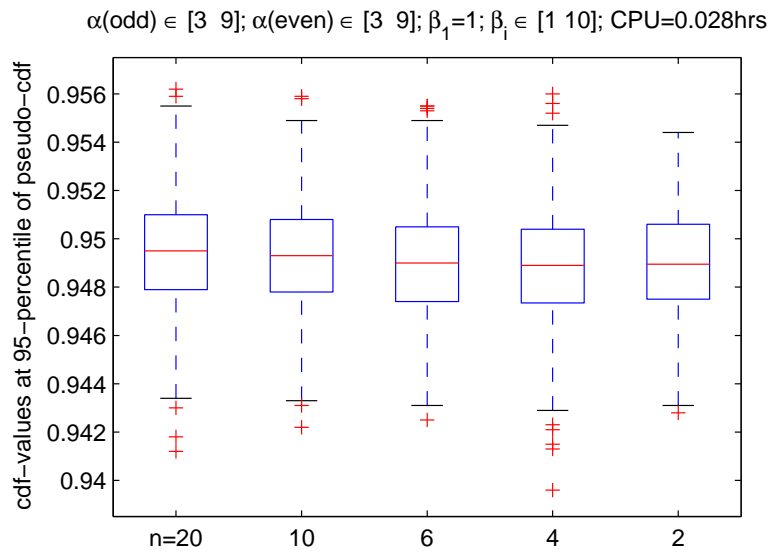
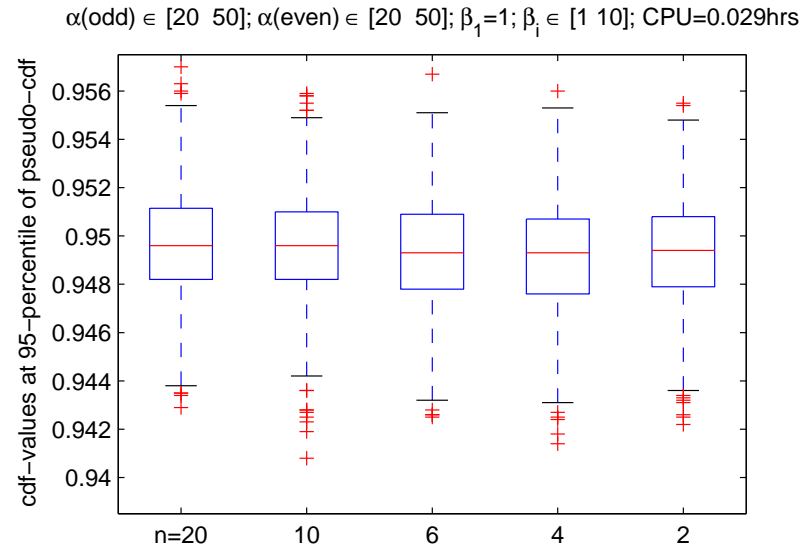
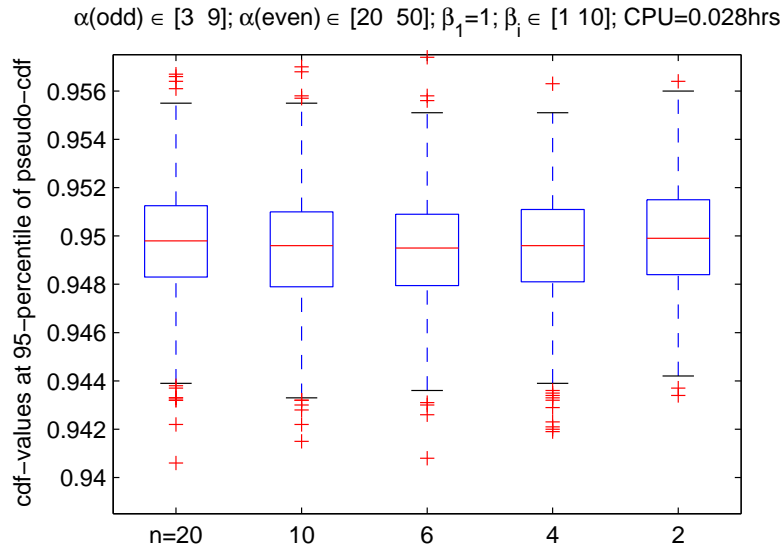
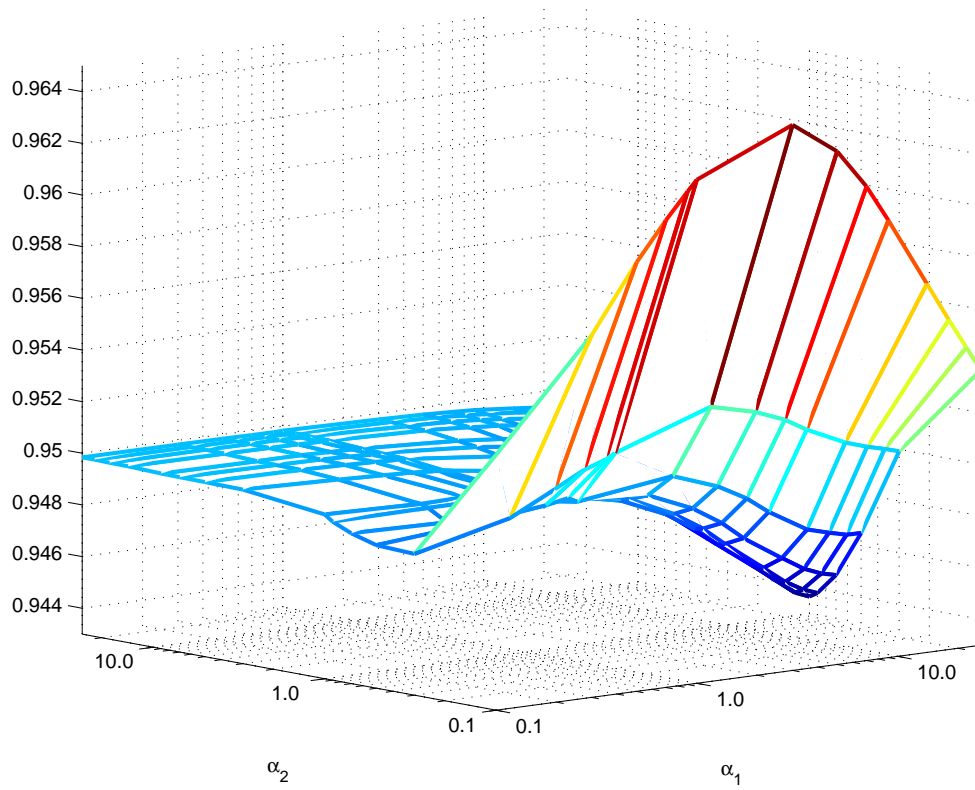
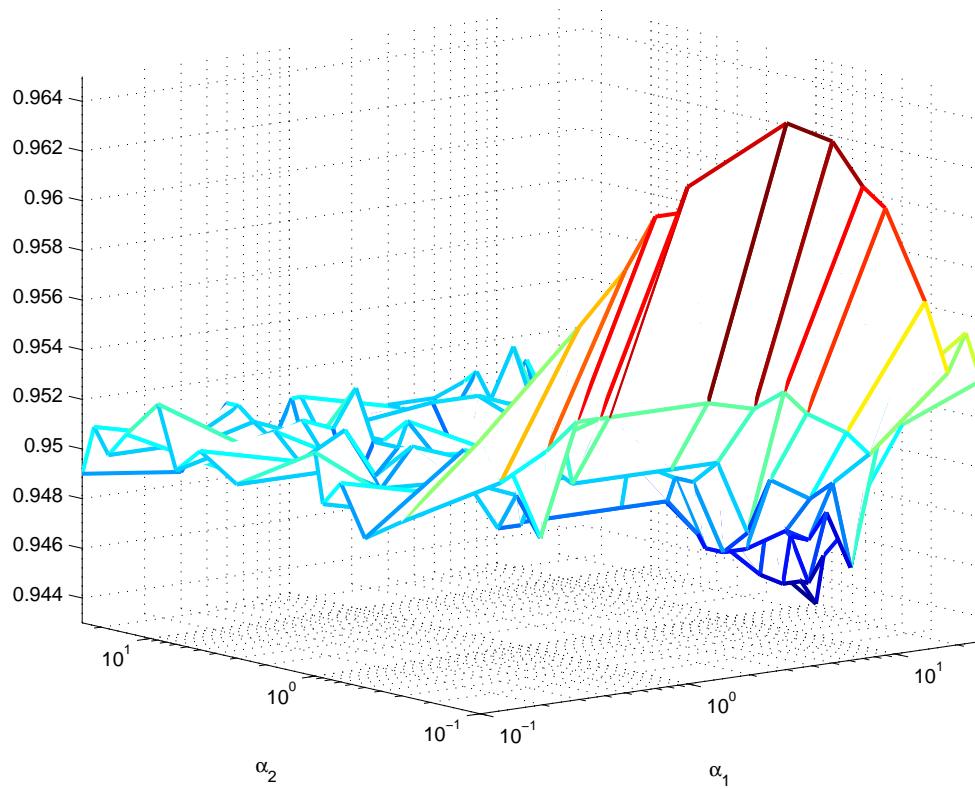


Figure 7: Boxplot \hat{p}_{k_3} ; $3 \leq \alpha_j \leq 50$.

Table 1: Values of $100(0.95 - M)$; $\beta_1 = 1$; $\beta_i \in [1, 10]$; $i = 2, \dots, n$.

20-50	20	0.06	0.04	0.03	0.04	0.05	0.04	0.04
	10	0.08	0.05	0.05	0.06	0.06	0.05	0.04
	$n : 6$	0.07	0.07	0.07	0.07	0.09	0.07	0.04
	4	0.07	0.07	0.10	<i>0.11</i>	<i>0.12</i>	0.10	0.06
	2	-0.05	0.03	<i>0.21</i>	0.36	<i>0.29</i>	<i>0.16</i>	0.07
9-20	20	0.05	0.06	0.05	0.05	0.04	0.05	0.04
	10	0.08	0.08	0.09	0.07	0.06	0.06	0.03
	$n : 6$	0.10	0.10	0.10	<i>0.11</i>	0.10	0.07	0.05
	4	0.10	0.10	<i>0.13</i>	<i>0.16</i>	<i>0.12</i>	0.08	0.05
	2	<i>-0.19</i>	-0.04	<i>0.18</i>	<i>0.29</i>	<i>0.19</i>	0.10	0.03
3-9	20	0.08	0.08	0.08	0.07	0.06	0.05	0.03
	10	0.10	<i>0.11</i>	0.10	0.08	0.08	0.07	0.04
	$n : 6$	<i>0.12</i>	<i>0.11</i>	<i>0.12</i>	<i>0.13</i>	0.10	0.08	0.05
	4	0.08	0.10	<i>0.13</i>	<i>0.13</i>	<i>0.12</i>	0.09	0.06
	2	-0.44	<i>-0.17</i>	0.10	<i>0.20</i>	0.09	0.05	0.02
1-3	20	0.10	0.10	0.07	0.09	0.07	0.05	0.03
	10	0.10	<i>0.12</i>	0.10	0.09	0.07	0.06	0.04
	$n : 6$	0.09	0.09	<i>0.11</i>	0.09	0.07	0.05	0.04
	4	0.01	0.05	0.09	0.10	0.08	0.07	0.04
	2	-0.74	-0.30	-0.01	0.07	0.05	0.01	0.00
0.4-1.0	20	0.08	0.09	0.10	0.08	0.07	0.06	0.04
	10	0.04	0.07	0.08	0.09	0.07	0.05	0.03
	$n : 6$	-0.02	0.00	0.06	0.06	0.07	0.05	0.04
	4	<i>-0.13</i>	-0.05	0.03	0.06	0.06	0.05	0.03
	2	-0.68	-0.27	-0.02	0.02	0.01	0.01	-0.01
0.2-0.4	20	0.07	0.08	0.09	0.08	0.07	0.04	0.04
	10	-0.02	0.02	0.06	0.08	0.07	0.06	0.03
	$n : 6$	<i>-0.12</i>	-0.05	0.01	0.05	0.06	0.06	0.04
	4	<i>-0.21</i>	<i>-0.14</i>	-0.03	0.02	0.07	0.04	0.04
	2	-0.48	<i>-0.19</i>	-0.02	0.02	0.00	-0.01	-0.01
0.1-0.2	20	0.01	0.07	0.10	0.07	0.05	0.04	0.02
	10	-0.09	-0.02	0.04	0.08	0.06	0.05	0.05
	$n : 6$	<i>-0.18</i>	-0.10	0.00	0.05	0.07	0.05	0.06
	4	<i>-0.27</i>	<i>-0.16</i>	-0.04	0.03	0.05	0.04	0.05
	2	-0.32	-0.09	-0.03	0.00	0.00	-0.00	0.01
$\alpha_{odd} \uparrow$	$\alpha_{even} \longrightarrow$	0.1-0.2	0.2-0.4	0.4-1.0	1-3	3-9	9-20	20-50

Figure 8: The probability that the convolution is below q_{95} according to Sim.**Figure 9:** Simulated probability that the convolution is below q_{95} .

5 Conclusions

A number of numerical methods exists for finding the near-exact convolution of gamma distributions. Here, we showed that this convolution can be described straightforwardly by another gamma, having the correct mean and variance. Table 1 shows that the degree of accuracy is quite high: in case the shape parameters are not below 0.1 and the scale parameters differ by no more than a factor of 10 at most, the cumulative probability at q_{95} generally is between 94 and 96%. Some evidence was given that the approximation is only slightly worse for shape values as low as 0.01. Furthermore, the approximation improves as n increases. For $n \geq 6$, for example, this cumulative probability is between 94.9 and 95.1%.

Our simple expression is in particular useful whenever further theoretical derivations are needed. The application in auditing, presented near the end of the Introduction, is just one example. But we think that our simple approximation may be useful in the excess water flow example - and many others! - given by Mathai (1982). The final outcomes in any specific application can of course easily be checked by means of simulation.

Note that the approximation can be extended immediately to all linear combinations of independent gamma type r.v.'s by considering the coefficients as scale parameters. An open question is whether our results hold as well for dependent r.v.'s; of course, then the variance formula has to be adapted.

6 Appendix A

It is easy to check that $X_1 \sim \Gamma(\alpha_1, 1)$ has the property

$$E(X_1^k) = \alpha_1(\alpha_1 + 1)(\alpha_1 + 2) \dots (\alpha_1 + k - 1), k = 1, 2, 3, \dots$$

from which we find

$$\begin{cases} \gamma_1 = E[(X_1 - \alpha_1)^3] = 2\alpha_1 \\ \kappa_1 = E[(X_1 - \alpha_1)^4] = 3\alpha_1(\alpha_1 + 2) \end{cases}$$

consequently, for $X_2 \sim \Gamma(\alpha_2, \beta)$ it follows

$$\gamma_2 = 2\alpha_2\beta^3, \kappa_2 = 3\alpha_2(\alpha_2 + 2)\beta^4$$

Hence, the exact skewness and kurtosis of the convolution of X_1 and X_2 equal

$$\begin{cases} \gamma_{ex} = E[X_1 + X_2 - \alpha_1 - \alpha_2\beta]^3 = \gamma_1 + \gamma_2 = 2\alpha_1 + 2\alpha_2\beta^3 \\ \kappa_{ex} = \kappa_1 + \kappa_2 + 6E[(X_1 - \alpha_1)^2] E[(X_2 - \alpha_2\beta)^2] \\ \quad = 3\alpha_1(\alpha_1 + 2) + 3\alpha_2(\alpha_2 + 2)\beta^4 + 6\alpha_1\alpha_2\beta^2 \end{cases}$$

which reduce to the first expressions in (3) and (4), respectively. The second expressions follow from approximation (2), which gives

$$\begin{cases} \gamma_{ap} = 2(\alpha_1 + \alpha_2\beta^2)^2/(\alpha_1 + \alpha_2\beta) \\ \kappa_{ap} = 3(\alpha_1 + \alpha_2\beta^2)^2 + 6(\alpha_1 + \alpha_2\beta^2)^3/(\alpha_1 + \alpha_2\beta)^2 \end{cases}$$

7 Appendix B

```
function pdf=SIM1992(y,ALPHA,BETA,n);
% Calculating the pdf of the sum of n independent gamma r.v.'s
% according to the representation of Sim (1992);
% see also Johnson et al. (1994), Section 8.4 (17.110).
epsilon = 1.e-6;
alfa = sum(ALPHA);
Part1 = prod(BETA.^(-ALPHA))*y^(alfa-1)*exp(-y/BETA(n))/gamma(alfa);
Part2 = 0;
if BETA/BETA(1) == ones(size(BETA)),
    Part2 = 1;
else
    for i = 3:n,
        c(i) = (1/BETA(i-2)-1/BETA(i-1))/(1/BETA(i)-1/BETA(i-1));
    end
    r=-1; ready = 0;
    while not(ready)|r<20,
        Part2Old = Part2;
        r = r+1; b(2,r+1) = 1;
        for i = 3:n,
            Salfim2 = sum(ALPHA(1:i-2));
            Salfim1 = sum(ALPHA(1:i-1));
            b(i,r+1) = 0;
            for j = 0:r,
```

```

b(i,r+1) = b(i,r+1) + b(i-1,j+1)*prod(c(i)*(-r+(0:j-1))...
.*(Salfim2+(0:j-1))./((1:j).*(Salfim1+(0:j-1))));
end
end
Part2 = Part2 + b(n,r+1)*prod(((1/BETA(n)-1/BETA(n-1))*y)...
*(alfa-ALPHA(n)+(0:r-1))./((alfa+(0:r-1)).*(1:r)));
ready = abs(Part2-Part2Old)<epsilon;
end
end
pdf = Part1*Part2;

```

8 Appendix C

In our auditing application, shape values smaller than 0.1 may occur; if so, they will be rather numerous. Hence, we here consider the case that many of the α_j are between 0.01 and 0.1. Their precise values will be taken at random from two intervals in \mathbb{R}^+ - as before; the m 'large' values will be indicated by $\alpha_{1:m}$, the 'small' ones by $\alpha_{m+1:n}$. The β_j will be chosen as before, while we now take $n = 50, 20,$ and 10 throughout.

Figure 10 shows which combinations of intervals from \mathbb{R}^+ we chose for further investigation.

Figure 10: Overview of parameter values α_j .

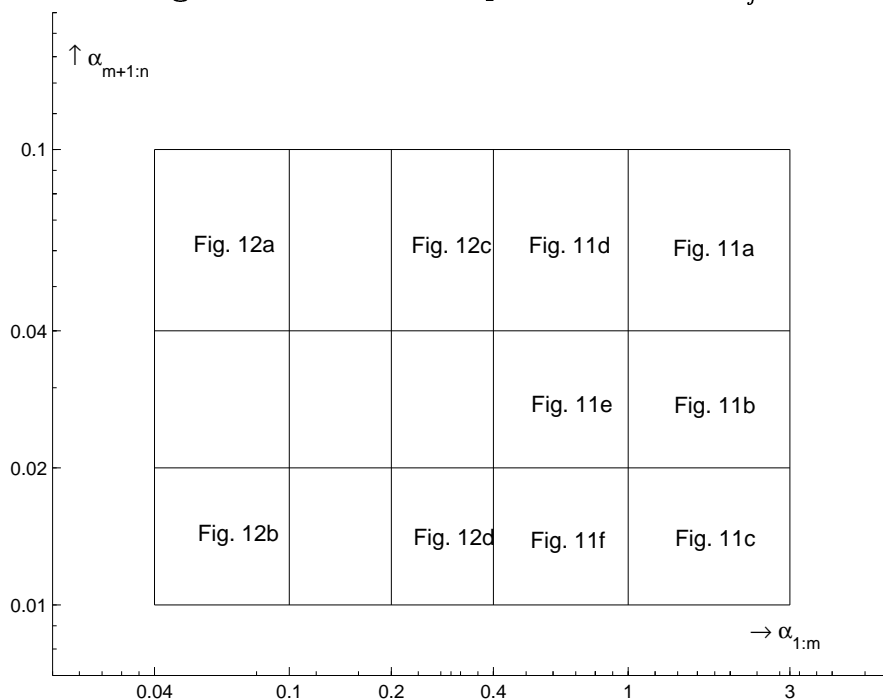
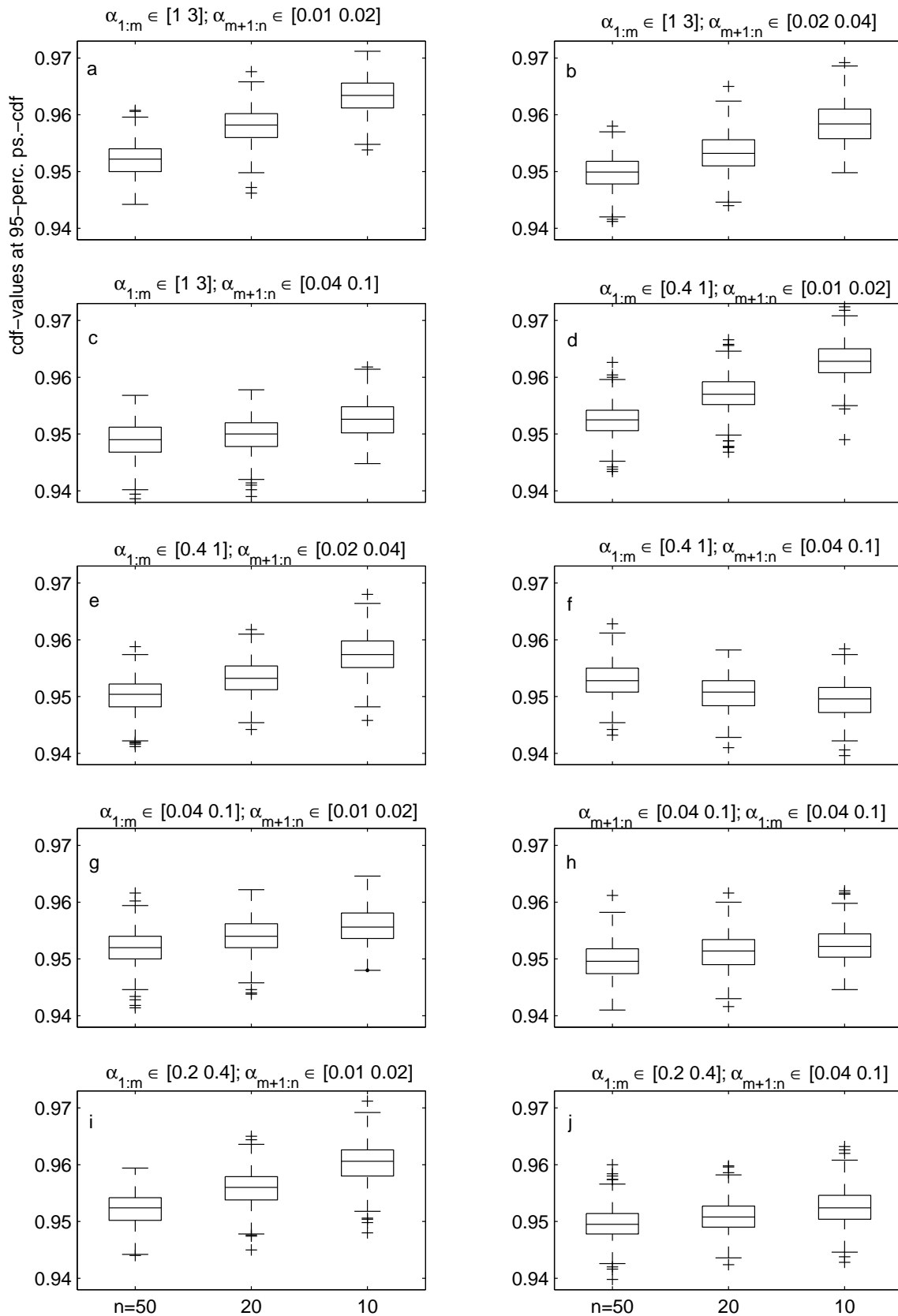


Figure 11: Boxplot \hat{p}_k ; $\alpha_{1:m} \in [0.04, 3]$, $\alpha_{m+1:n} \in [0.01, 0.1]$.

The values of m were varied in the different combinations, but in all situations considered the approximation was worst for $m = 1$. Only this worst combination is shown in Figure 11. The alternative values of m considered were (apart from 1):

Figure 11a-c: $m = 2$ and 3

Figure 11d-f: $m = 3$ and 5

Figure 11g-j: $m = n/5$ and $n/2$.

Among these worst cases, the largest difference between M and 0.95 occur in Figures 11a, d and i: here M equals 0.9634, 0.9628 and 0.9606, respectively. Note, however, that fortunately for our application the achieved assurance is higher than the prescribed 95%.

9 References

- Akkouchi, M., 2005. On the convolution of gamma distributions, *Soochow Journal of Mathematics*, 31(2), 205-211.
- Alouini, M.-S., Abdi, A. and Kaveh, M., 2001, Sum of gamma variates and performance of wireless communication systems over Nakagami-fading channels, *IEEE Transactions on Vehicular Technology*, 50(6), 1471-1480.
- Fleurant, C., Duchesne, J. and Raimbault P., 2004, An allometric model for trees, *J. for Theoretical Biology*, 227, 137-147.
- Johnson, N.L., Kotz S. and Balakrishnan, N., 1994, *Continuous univariate distributions*, Vol.1, 2nd ed., Wiley, New York.
- Mathai, A.M., 1982, Storage capacity of a dam with gamma type inputs, *Ann. Inst. Statist. Math.*, 34, Part A, 591-597.
- Mathai, A.M. and Moschopoulos, P.G., 1991, On a multivariate gamma, *Journal of Multivariate Analysis*, 39, 135-153.
- Moschopoulos, P.G., 1985, The distribution of the sum of independent gamma random variables, *Ann. Inst. Statist. Math.*, 37, Part A, 541-544.
- Sim, C.H., 1992, Point process with correlated gamma interarrival times, *Stat. & Prob. Letters*, 15, 135-141.

Stewart, T.R., Assurance and materiality in group and other multi-component audits, unpublished working paper (available from the author upon request). Mimeo. (Deloitte & Touche LLP, New York, 2007).

Thom, H.C.S., 1968, Approximate convolution of the gamma and mixed gamma distributions, *Monthly Weather Review*, 96, 883-886.

Tsionas, E.G., 2004, Bayesian inference for multivariate gamma distributions, *Statistics and Computing*, 14, 223-233.