THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

Conditional Monte Carlo for Sums, with Applications to Insurance and Finance

Søren Asmussen

Research Report

No. 01 | January 2017

Conditional Monte Carlo for Sums, with Applications to Insurance and Finance*

Søren Asmussen

Department of Mathematics, Aarhus University, asmus@math.au.dk

Abstract

Conditional Monte Carlo replaces a naive estimate Z of a number z by its conditional expectation given a suitable piece of information. It always reduces variance and its traditional applications are in that vein. We survey here other potential uses such as density estimation and Value-at-Risk calculations, going in part into the implementation in various copula structures. Also the interplay between these different aspects comes into play.

Keywords: Archimedean copula; Density estimation; Expected shortfall; Log-normal sums; Rare event simulation; Value-at-Risk

1 Introduction

Let z be a number represented as an expectation $z = \mathbb{E}Z$. The crude Monte Carlo (CrMC) method for estimating z then proceeds by simulating R replications $Z_1, \ldots, Z_R$ of Z and returning the average $\overline{z} = (Z_1 + \cdots + Z_R)/R$ as point estimate. The uncertainty is reported as an asymptotic confidence interval based on the CLT; for example, the two-sided 95% confidence interval is $\overline{z} \pm 1.96\, s/R^{1/2}$ where s^2 is the empirical variance of the sample $Z_1, \ldots, Z_R$.

The more refined conditional Monte Carlo (CdMC) method uses a piece of information collected in a σ-field $\mathcal{F}$ and is implemented by performing CrMC with Z replaced by $Z_{\text{Cond}} = \mathbb{E}[Z \,|\, \mathcal{F}|$. It is traditionally classified as a variance reduction method but it can also be used for smoothing, though this is much less appreciated.

Both aspects are well illustrated via the problem of estimating $\mathbb{P}(S_n \le x)$ where $S_n = X_1 + \cdots + X_n$ is a sum of r.v.'s. The apparent choice for CrMC is $Z = Z(x) = \mathbb{I}(S_n \le x)$. For CdMC, a simple possibility is to take $\mathcal{F} = \sigma(X_1, \ldots, X_{n-1})$. In the case where $X_1, X_2, \ldots$ are i.i.d. with common distribution F one then has

$$Z_{\text{Cond}} = \mathbb{P}(S_n \le x \,|\, X_1, \ldots, X_{n-1}) = F(x - S_{n-1})\,. \tag{1.1}$$

This estimator has two noteworthy properties:

*Figures in this version are in colour and do not all print well in black-white

- for a fixed x its variance is smaller than that of $\mathbb{I}(S_n \le x)$ used in the CrMC method;
- when averaged over the number R of replications, it leads to estimates of $\mathbb{P}(S_n \le x)$ which are smoother as function of $x \in (-\infty, \infty)$ than the more traditional empirical c.d.f. of R simulated replicates of S_n.

This last property is easily understood for a continuous F, where $Z_{\mathrm{Cond}}(x) = F(x - S_{n-1})$ is again continuous and therefore averages are so. In contrast, the empirical c.d.f. always has jumps. It also suggest that $f(x - S_{n-1})$ may be an interesting candidate for estimating the density $f_n(x)$ of S_n when F itself admits a density $f(x)$. In fact, density estimation is a delicate topic where traditional methods such as kernel smoothing or finite differences often involve tedious and ad hoc tuning of parameters like choice of kernel, window size etc.

The variance reduction property holds in complete generality by the general principle (known as Rao-Blackwellization in statistics) that conditioning reduces variance:

$$\mathbb{V}\mathrm{ar}\ Z = \mathbb{E}\big[\mathbb{V}\mathrm{ar}(Z \,|\, \mathcal{F})\big] + \mathbb{V}\mathrm{ar}\big[\mathbb{E}(Z \,|\, \mathcal{F})\big] \ge \mathbb{V}\mathrm{ar}\big[\mathbb{E}(Z \,|\, \mathcal{F})\big] = \mathbb{V}\mathrm{ar}\ Z_{\mathrm{Cond}}\,.$$

In view of the huge literature on variance reduction, this may appear appealing but it also has some caveats inherent in the choice of $\mathcal{F}$: $\mathbb{E}(Z \,|\, \mathcal{F})$ must be computable and have a variance that is *substantially* smaller than that of Z. Namely, if CdMC reduces the variance on Z of Z_{Cond} by a factor of $\tau < 1$, the same variance on the average could be obtained by taking $1/\tau$ as many replications in CrMC as in CdMC, see Asmussen & Glynn (2007) p. 126. This point is often somewhat swept under the carpet!

The present paper discusses such issues related to the CdMC method via the example of inference on the distribution of a sum $S_n = X_1 + \cdots + X_n$. Here the X_i are assumed i.i.d. in Sections 2–6, but we look into dependence in some detail in Section 7, whereas a few comments on different marginals are given in Section 8.

The motivation comes to a large extent from problems in insurance and finance such as assessing the form of the density of the loss distribution, estimating the tail of the aggregated claims in insurance, calculating the Value-at-Risk (VaR) or expected shortfall of a portfolio etc. In many such cases, the tail of the distribution of S_n is of particular interest, with the relevant tail probabilities being of order 10^{-2}–10^{-4} (but note that in other application areas, the relevant order is much lower, say 10^{-8}–10^{-12} in telecommunications). By "tail" we are not just thinking of the *right* tail, i.e. $\mathbb{P}(S_n > x)$ for large x, which is relevant for the aggregated claims and portfolios with short positions. Also the *left* tail $\mathbb{P}(S_n \le x)$ for small x comes up in a natural way, in particular for portfolios with long positions, but has received much less attention until the recent studies by Asmussen *et al.* (2016) and Gulisashvili & Tankov (2016).

The most noted use of CdMC in the insurance-finance-rare-event area appears to be the algorithm of Asmussen & Kroese (2006) for calculating the right tail of a heavy-tailed sum. A main application is ruin probabilities. We give references and put this is perspective to the more general problems of the present paper in Section 6. Otherwise, the use of CdMC in insurance and finance seem to be remarkably few compared to other MC based tools such as importance sampling, stratification, simulation-based estimation of sensitivities (greeks), just to name a few (see Glasserman, 2004, for these and other examples). Some exceptions are Fu *et al.* (2009) who study an CdMC estimator of a sensitivity of a quantile (not the quantile itself!) with respect to a model parameter, and Chu & Nakayama (2012) who give a detailed mathematical derivation of the CLT for quantiles estimated in a CdMC set-up, based on methodology from Bahadur (1966) and Ghosh (1971) (see also Nakayama, 2014).

Conventions

Throughout the paper, $\Phi(x)$ denotes the standard normal c.d.f., $\overline{\Phi}(x) = 1 - \Phi(x)$ its tail and $\varphi(x) = \mathrm{e}^{-x^2/2}/\sqrt{2\pi}$ the standard normal p.d.f. For the gamma(α, λ) distribution, α is the shape parameter and λ the rate so that the density is $x^{\alpha-1}\lambda^{\alpha}\mathrm{e}^{-\lambda x}$.

Because of the financial relevance, an example that will be used frequently is X to be lognormal$(0,1)$, i.e. of the form $X = \mathrm{e}^V$ with V normal$(0,1)$, and $n = 10$. Note that the mean of V is just a scaling factor and hence unimportant. In contrast, the variance (and the value of n) matters quite of lot for the shape of the distribution of S_n, but to be definite, we took it to be one. We refer to this set of parameters as our recurrent example, and many other examples are taken as smaller or larger modifications

2 Density Estimation

If F has a density f, then S_n has density

$$f_n(x) = f^{*n}(x) = \int_{x_1+\cdots+x_n=x} f(x_1)\cdots f(x_n)\,\mathrm{d}x_1\,\cdots\,\mathrm{d}x_n\,.$$

The convolution integral can only be evaluated numerically for rather small n, and we shall here consider the estimator $f(x - S_{n-1})$ of $f_n(x)$. Because of the analogy with (1.1), it seems reasonable to classify this estimator within the CdMC area, but it should be noted that there is no apparent natural unbiased estimator Z of $f_n(x)$ for which $\mathbb{E}\big[Z \,|\, X_1, \ldots, X_{n-1}\big] = f(x - S_{n-1})$. Of course, intuitively

$$\mathbb{P}(S_n \in \mathrm{d}x \,|\, X_1, \ldots, X_{n-1}) = f(x - S_{n-1})$$

but $\mathbb{I}(S_n \in \mathrm{d}x)$ is not a well-defined r.v.! Nevertheless:

Proposition 2.1. *The estimator $f(x - S_{n-1})$ of $f_n(x)$ is unbiased..*

Proof. $\mathbb{E}f(x - S_{n-1}) = \displaystyle\int f_{n-1}(y)f(x-y)\,\mathrm{d}y = f_n(x)$. □

Unbiasedness is in fact quite a virtue in itself, since the more traditional kernel- and finite difference estimators are not so! It also implies that the average over R replications converges to the correct value $f_n(x)$ as $R \to \infty$.

Because of the lack of an obvious CrMC comparison, we shall not go into detailed properties of $\mathbb{V}\mathrm{ar}\, f(x - S_{n-1})$; one expects such a study to be quite similar to the one in Section 3 dealing with $\mathbb{V}\mathrm{ar}\, F(x - S_{n-1})$. Instead, we shall give some numerical examples

Fig. 1 illustrates the influence on the number R of replications. For each of the four values $R = 2^8, 2^{10}, 2^{12}, 2^{14}$ we performed 3 sets of simulation, to assess the degree of randomness inherent in R being finite. Obviously, $R = 2^{14} \approx 16{,}000$ is almost perfect but the user may go for a substantially smaller value depending on how much the random variation and the smoothness is a concern.

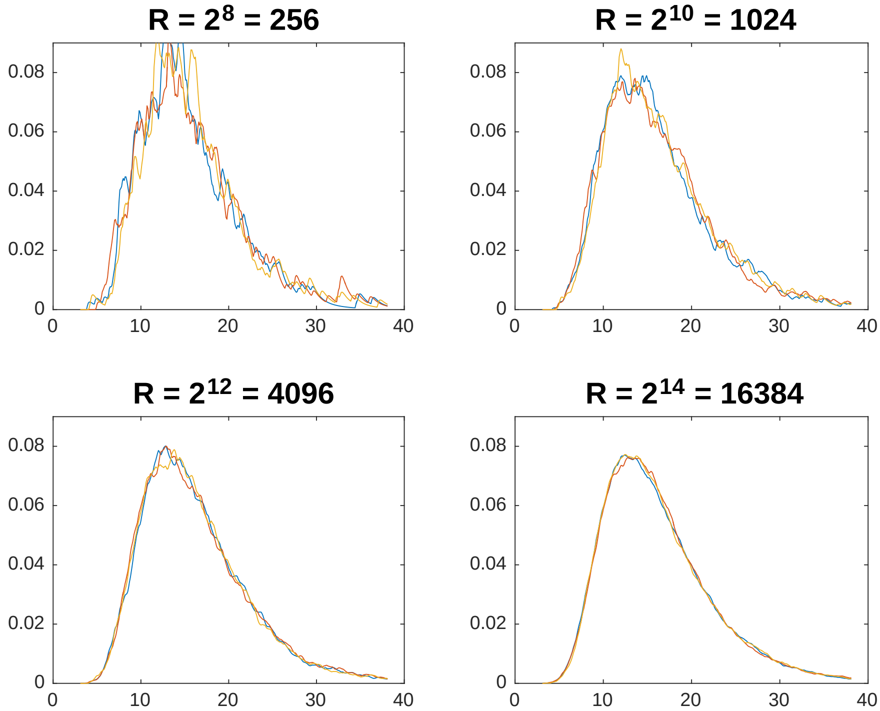


Figure 1: Estimated ensity of S_n as function of R

A reasonable question is the comparison of CdMC and a kernel estimate of the form $k(x - S_n)$ for small or moderate R. In Fig. 2, we considered our recurrent example of sum of lognormals, but took $R = 32$ for both of the estimators $f(x - S_{n-1})$ and $k(x - S_n)$, with k chosen as the normal $(0, \sigma^2)$ density. The upper right panel is a histogram of the 32 simulated values of S_n and the upper left the CdMC estimator. The two lower panels are the kernel estimates, with an extreme high value $\sigma^2 = 10^2$ to the left and an extreme low $\sigma^2 = 10^{-2}$ to the right. A high value will produce oversmoothed estimates and a low undersmoothed ones with a marked effect of single observations. However, for R as small as 32 it is hard to assess what is a reasonable value of σ^2. In fairness, we also admit that the single observation effect is clearly visible for the CdMC estimator and that it leads to estimates which are undersmoothed. By this we mean more precisely that if f is C^p for some $p = 0, 1, \ldots$, then $f_n(x)$ is C^{np} but $f(x - S_{n-1})$ only C^p. In contrast, a normal kernel estimate is even C^∞.

The first example of CdMC density estimation we know of is in Asmussen & Glynn (2007) p. 146, but in view of the simplicity of the idea, there may well have been earlier instances. We return to some further aspects of the methodology in Section 5.

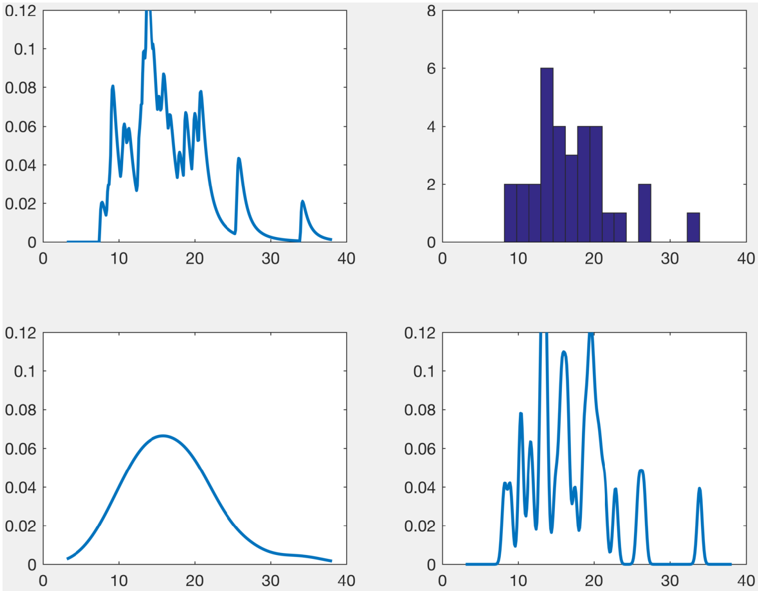


Figure 2: Comparison with kernel smoothing

3 Variance Reduction for the CDF

Conditional Monte Carlo always gives variance reduction. But as argued, it needs to be substantial for the procedure to be worthwhile. Further in many applications the right and/or left tail is of particular interest, so one may pay particular attention to the behaviour there.

Remark 3.1. That CdMC gives variance reduction in the tails can be seen intuitively by the following direct argument without reference to Rao-Blackwellization. The CrMC, resp. the CdMC, estimators of $\overline{F}(x)$ are $\mathbb{I}(S_n > x)$ and $\overline{F}(x - S_{n-1})$, with second moments

$$\mathbb{E}\mathbb{I}(S_n > x)^2 = \mathbb{E}\mathbb{I}(S_n > x) = \int f_{n-1}(y)\overline{F}(x-y)\,\mathrm{d}y\,, \tag{3.1}$$

$$\mathbb{E}\overline{F}(x - S_{n-1})^2 = \int f_{n-1}(y)\overline{F}(x-y)^2\,\mathrm{d}y\,. \tag{3.2}$$

In the right tail (say), these can be interpreted as the tails of the r.v.'s $S_{n-1} + X$, $S_{n-1} + X^*$ where X, X^* are independent of S_{n-1} and have tails $\overline{F}$ and $\overline{F}^2$. Since $\overline{F}^2(x)$ is of smaller order than $\overline{F}(x)$ in the right tail, the tail of $S_{n-1} + X^*$ should be of smaller order than that of $S_{n-1} + X$, implying the same ordering of the second moments. However, as n becomes large one also expects the tail of S_{n-1} to be more and more important compared to those of X, X^* so that the difference should be less and less marked. The analysis to follow will confirm these guesses. ◊

As measure of performance, we consider the ratio $r_n(x)$ of the CdMC variance to the CrMC variance,

$$r_n(x) = \frac{\mathbb{V}\mathrm{ar}\big[\overline{F}(x - S_{n-1})\big]}{F(x)\overline{F}(x)} = \frac{\mathbb{V}\mathrm{ar}\big[F(x - S_{n-1})\big]}{F(x)\overline{F}(x)} \tag{3.3}$$

(note that the two alternative expressions reflect that the variance reduction, is the same whether CdMC is performed for F itself or the tail $\overline{F}$).

To provide some first insight, we start by Fig. 3, giving $r_n(x_{n,z})$ as function of z where $x_{n,z}$ is the z-quantile of S_n. In Fig. 3a, the underlying F is Pareto with tail $\overline{F}(x) = 1/(1+x)^{3/2}$ and in Fig. 3b, it is standard normal. Both figures consider the cases of a sum of $n = 2, 5$ or 10 terms and use $R = 250,000$ replications of the vector $Y_1, \ldots, Y_{n-1}$ (variances are more difficult to estimate than means, therefore the high value of R). The dotted line for AK may be ignored for the moment. The argument z on the horizontal axis is in $\log_{10}$-scale, and $x_{n,z}$ was takes as the exact value for the normal case and the CdMC estimate for the Pareto case.

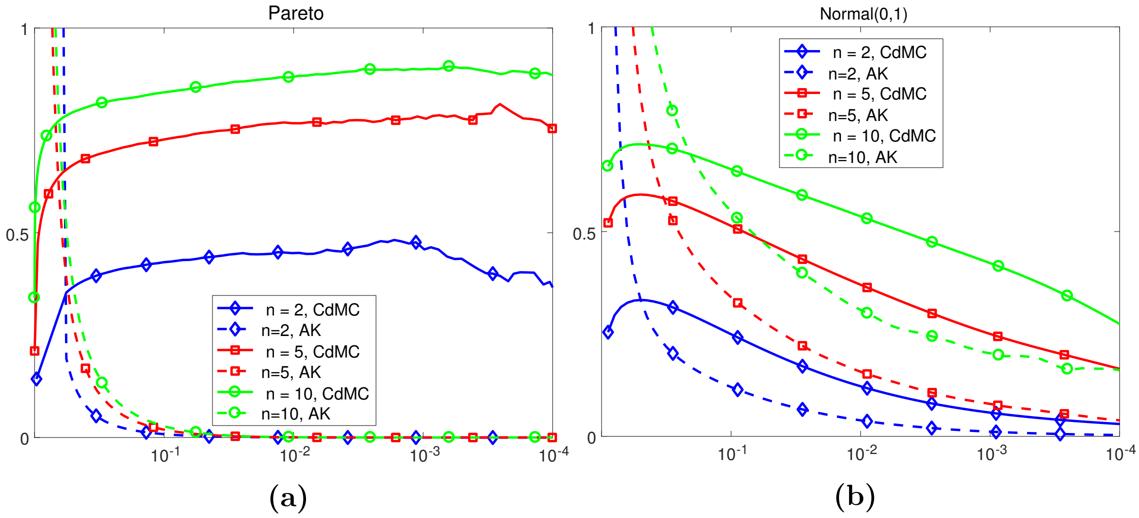


Figure 3: The ratio $r_n(z)$ in (3.3), with F Pareto in (a) and normal in (b)

For the Pareto case in Fig. 3a, it seems that the variance reduction is decreasing in both x and n, in fact it is only substantial in the left tail. For the normal case, note that there should be symmetry around $x = 0$, corresponding to $z(x) = 1/2$ with base-10 logarithm -0.30. This is confirmed by the figure (though the feature is of course somewhat disguised by the logarithmic scale). In contrast to the Pareto case, it seems that the variance reduction is very big in the right (and therefore also left) tail but also that it decreases as n increases.

We proceed to a number of theoretical results supporting these empirical findings. They all use formula (3.2) for the second moment of the CdMC estimator, which for $X > 0$ takes the form

$$\mathbb{E}\overline{F}(x - S_{n-1})^2 \overset{X \geq 0}{=} \overline{F}(x) + \int_0^x f_{n-1}(y)\overline{F}(x-y)^2 \, dy. \tag{3.4}$$

For the exponential distribution, the calculations are particularly simple:

Example 3.2. Assume $\overline{F}(x) = e^{-x}$, $n = 2$. Then $\mathbb{P}(X_1 + X_2 > x) = xe^{-x} + e^{-x}$ and (3.4) takes the form

$$\overline{F}(x) + \int_0^x e^{-y}e^{-2(x-y)} \, dy = e^{-x} + e^{-2x}(e^x - 1) = 2e^{-x} - e^{-2x}$$

and so for the right tail,

$$r_2(x) = \frac{2\mathrm{e}^{-x} - \mathrm{e}^{-2x} - (x\mathrm{e}^{-x} + \mathrm{e}^{-x})^2}{(x\mathrm{e}^{-x} + \mathrm{e}^{-x})(1 - x\mathrm{e}^{-x} - \mathrm{e}^{-x})}.$$

For $x \to \infty$, this gives

$$r_2(x) = \frac{2\mathrm{e}^{-x} + \mathrm{o}(\mathrm{e}^{-x})}{x\mathrm{e}^{-x} + \mathrm{o}(x\mathrm{e}^{-x})} = \frac{1}{x}\big(1 + \mathrm{o}(1)\big) \to 0.$$

In the left tail $x \to 0$, Taylor expansion give that up to the third order term

$$2\mathrm{e}^{-x} - \mathrm{e}^{-2x} \sim 1 - x^2 + x^3, \quad x\mathrm{e}^{-x} + \mathrm{e}^{-x} = 1 - x^2/2 + x^3/6,$$

and so

$$\begin{aligned} r_2(x) &\sim \frac{1 - x^2 + x^3 - (1 - x^2/2 + x^3/6)^2}{(1 - x^2 + x^3)(x^2/2 - x^3/6)} \\ &\sim \frac{1 - x^2 + x^3 - (1 - x^2 + 2x^3/6)}{x^2/2} = \frac{2x}{3} \to 0. \end{aligned}$$ ◊

The relation $r_n(x) \to 0$ in the left tail (i.e., as $x \to 0$) in the exponential example is in fact essentially a consequence of the support being bounded to the left:

Proposition 3.3. *Assume $X > 0$ and that the density $f(x)$ satisfies $f(x) \sim cx^p$ as $x \to 0$ for some $c > 0$ and some $p > -1$. Then $r_n(x) \sim dx^{p+1}$ as $x \to 0$ for some $0 < d < \infty$.*

The following result explains the right tail behaviour in the Pareto example and shows that this extends to other standard heavy-tailed distributions like the lognormal or DFR Weibull (for subexponential distributions, see e.g. Embrechts *et al.*, 1997):

Proposition 3.4. *Assume $X > 0$ is subexponential. Then $r_n(x) \to 1 - 1/n$ as $x \to \infty$.*

For light tails, Example 3.2 features a different behaviour in the right tail, namely $r_n(x) \to 0$. Here is one more such light-tailed example:

Proposition 3.5. *If X is standard normal, then $r_n(x) \to 0$ as $x \to \infty$. More precisely,*

$$r_n(x) \sim \frac{1}{x}\sqrt{\frac{2n-1}{n\pi}}\mathrm{e}^{-x^2/[2n(2n-1)]}.$$

The proofs of Propositions 3.3–3.5 are in the Appendix.

To formulate a result of type $r_n(x) \to 0$ as $x \to \infty$ in a sufficiently broad class of light-tailed F encounters the difficulty that the general results giving the asymptotics of $\mathbb{P}(S_n > x)$ as $x \to \infty$ with n fixed are somewhat involved (the standard light-tailed asymptotics is for $\mathbb{P}(S_n > bn)$ as $n \to \infty$ with b fixed, cf. e.g. Jensen, 1995). It is possible to obtain more general versions of Example 3.2 for close-to-exponential tails by using results of Cline (1986) and of Proposition 3.5 for

Table 1: Variance reduction for sum of 3 gammas

z	x_z	CdMC	IS-ECM	CdMC+IS-ECM
0.95	39.5	0.628	0.121	0.048
0.99	44.2	0.561	0.032	0.010

thinner tails by involving Balkema *et al.* (1993). The adaptation of in particular Balkema *et al.* (1993) is, however, technical and for ease of exposition of the present paper it will be presented elsewhere.

One may note that the variance reduction is so moderate in the range of z considered in Fig. 3b that CdMC may hardly be worthwhile for light tails except for possibly very small n. If variance reduction is a major concern, the obvious alternative is to use the standard importance sampling (IS) algorithm which uses exponential change of measure (ECM). The r.v.'s $X_1, \ldots, X_n$ are here generated from the exponentially twisted distribution with density $f_\theta(x) = \mathrm{e}^{\theta x} f(x)/\mathbb{E}\mathrm{e}^{\theta X}$ where θ should be chosen such that $\mathbb{E}_\theta S_n = x$. The estimator of $\mathbb{P}(S_n > x)$ is

$$\frac{\mathrm{e}^{\theta S_n}}{\big[\mathbb{E}\mathrm{e}^{\theta X}\big]^n}\mathbb{I}(S_n > x) \tag{3.5}$$

see Asmussen & Glynn (2007), pp. 167–169 for more detail. Further variance reduction would be obtained by applying CdMC to (3.5) as implemented in the following example.

Example 3.6. To illustrate the potential of the IS-ECM algorithm, we consider the sum of $n = 10$ gamma(3,1) r.v.'s at the $z = 0.95, 0.99$ quantiles x_z. The exponentially twisted distribution is gamma$(3, 1-\theta)$ and $\mathbb{E}_\theta S_n = x$ means $3/(1-\theta) = x$, i.e. $\theta = 1 - 3/(x/n)$. With $R = 100.000$ replications, we obtained the values of $r_n(x)$ at the z quantiles for $z = 0.95$, 0.99 given in Table 1. It is seen that IS-ECM indeed performs much better that CdMC, but that CdMC is also moderately useful for providing some further variance reduction. ♢

A further financially relevant implementation of the IS-ECM algorithm is in Asmussen *et al.* (2016) for lognormal sums. It is unconventional in the way of dealing with the left tail (which is light) rather than the right tail (which is heavy) and in that the ECM is not explicit but done in an approximately efficient way. Another IS algorithm for the left lognormal sum tail is in Gulisashvili & Tankov (2016), but the numerical evidence of Asmussen *et al.* (2016) makes it efficiency somewhat doubtful.

4 Value-at-Risk

The Value-at-Risk $\mathrm{VaR}_\alpha(S_n)$ of S_n at level α is intuitively defined as the number such that the probability of a loss larger than $\mathrm{VaR}_\alpha(S_n)$ is $1-\alpha$. Depending on whether small or large values of S_n mean a loss, there are two forms around, the actuarial $\mathrm{VaR}_\alpha(S_n)$ defined as the α-quantile $x_{\alpha,n}$ and the financial VaR defined as $-x_{1-\alpha,n}$. Typical values of α are 0.95 and 0.95 but smaller values occur in Basel II for certain types of business lines. We use here the actuarial definition.

The traditional simulation-based estimate is based on R simulated values $S_n^{(1)}$, $\ldots, S_n^{(R)}$ and taken as the α-quantile $\widehat{q}_\alpha = F_R^{-1}(\alpha\,;S_n)$ of the empirical c.d.f.

$$F_R(x\,;S_n) = \frac{1}{R}\sum_{r=1}^{R} \mathbb{I}(S_n^{(r)} \le x).$$

Thus $\widehat{q}_\alpha$ is more complicated than an average of i.i.d. r.v.'s but nevertheless, there is a CLT

$$\sqrt{R}(\widehat{q}_\alpha - q_\alpha) \to \mathcal{N}(0,\sigma_\alpha^2) \qquad \text{where } \sigma_\alpha^2 = \frac{\alpha(1-\alpha)}{f_n(q_\alpha)^2}. \tag{4.1}$$

Thus confidence interval requires an estimate of $f_n(q_\alpha)$, an issue about which Glynn (1996) writes that "the major challenge is finding a good way of estimating $f_n(q_\alpha)$, either explicitly or implicitly" (without providing a method for doing this!) and Glasserman *et al.* (2000) that "estimation of $f_n(q_\alpha)$ is difficult and beyond the scope of this paper".

When confidence bands for the VaR are given, a common practice is therefore bootstrap. However, in our sum setting, CdMC easily gives $f_n(q_\alpha)$, as outlined in Section 2. In addition, the method provides some variance reduction because of its improved estimates of the c.d.f. In fact, if $\widehat{q}_{\alpha;\mathrm{Cond}}$ is defined as the solution of $\widehat{F}_{\mathrm{Cond}}(\widehat{q}_{\alpha;\mathrm{Cond}}) = \alpha$ where $\widehat{F}_{\mathrm{Cond}}(x) = \sum_1^R F\bigl(x - S_{n-1}^{(r)}\bigr)/R$, then

$$\sqrt{R}(\widehat{q}_{\alpha;\mathrm{Cond}} - q_\alpha) \to \mathcal{N}(0,\sigma_{\alpha;\mathrm{Cond}}^2) \qquad \text{where } \sigma_{\alpha;\mathrm{Cond}}^2 = \frac{\mathbb{V}\mathrm{ar}(\widehat{F}_{\mathrm{Cond}}(q_\alpha)}{f_n(q_\alpha)^2}. \tag{4.2}$$

and $\mathbb{V}\mathrm{ar}\,\widehat{F}_{\mathrm{Cond}}(q_\alpha) < \alpha(1-\alpha)$. Details on how to arrive at (4.2) are in Bahadur (1966), Ghosh (1971), Serfling (1980), Asmussen & Glynn (2007) III.4, Chu & Nakayama (2012) and Nakayama (2014), and sketched in Section A in connection with similar formulas for the expected shortfall in (4.3) below.

Remark 4.1. At a first sight, the more obvious way to involve CdMC would have been to give the VaR estimates as the average over R replications of the α-quantile $\widetilde{q}_\alpha$ in the conditional distribution of S_n given S_{n-1}. However, this does not provide the correct answer and in fact introduces a bias that does not disappear for $R \to \infty$ as it does for $\widehat{q}_\alpha = F_R^{-1}(\alpha\,;S)$ and $\widehat{q}_{\alpha;\mathrm{Cond}}$. For a simple example illustrating this, consider the i.i.d. $\mathcal{N}(0,1)$-setting,. Here $\widetilde{q}_\alpha = S_{n-1} + z_\alpha$ where $z_\alpha = \Phi^{-1}(\alpha)$ with expectation z_α but the correct answer is $\sqrt{n}z_\alpha$! ♢

Example 4.2. As illustration, we used CdMC with $R = 50.000$ replications to compute VaR_α and the associated confidence interval for the sum of $n = 5, 10, 25, 50$ lognormal$(0,1)$ r.v.'s. The results are in Table 2. ♢

An alternative risk measure receiving much current attention is the expected shortfall

$$\mathrm{ES}_\alpha(S_n) = \mathrm{VaR}_\alpha(S_n) + \mathbb{E}\bigl[S_n - \mathrm{VaR}_\alpha(S_n)\bigr]^+. \tag{4.3}$$

The obvious CdMC algorithm for estimating $\mathrm{ES}_\alpha(S_n)$ is to first simulate $S_{n-1}^{(1)}, \ldots, S_{n-1}^{(R)}$, next compute $\widehat{q}_{\alpha;\mathrm{Cond}}$ as above, and finally to let $\widehat{e} = \widehat{q} + R^{-1}\sum_1^R m\bigl(\widehat{q} - S_{n-1}^{(r)}\bigr)$ where $m(x) = \mathbb{E}(X-x)^+$. Giving a confidence interval is somewhat more complicated that for the VaR, and since we are not aware of a sufficiently close reference, the details are in Section A of the Appendix.

Table 2: VaR estimates for lognormal example

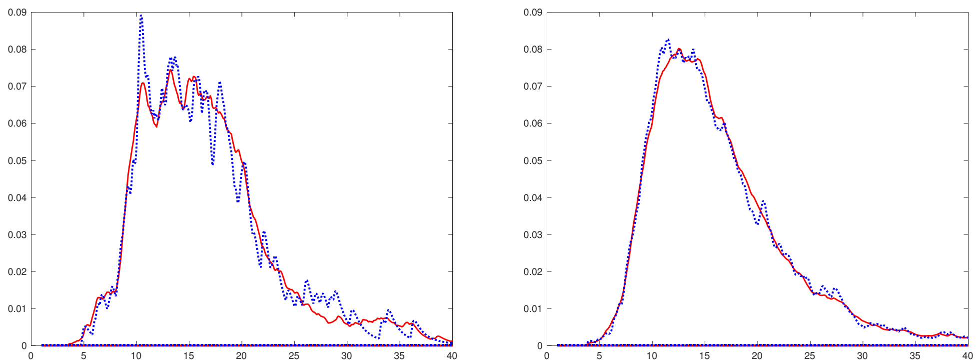
	$\alpha = 0.95$	$\alpha = 0.99$
$n = 5$	17.0 ± 0.2	25.4 ± 0.6
$n = 10$	29.0 ± 0.2	39.9 ± 0.7
$n = 25$	61.0 ± 0.3	74.6 ± 0.8
$n = 50$	109.8 ± 0.4	127.2 ± 1.0

5 Averaging

The idea of using $f(x - S_{n-1})$ and $F(x - S_{n-1})$ as estimators of f_n, resp. F_n, has an obvious asymmetry in that X_n has a special role among $X_1, \dots, X_n$ by being the one that is not simulated but handled via its conditional expectation given the rest. An obvious idea is therefore to repeat the procedure with X_n replace by X_k and average over $k = 1, \dots, n$. This leads to the alternative estimators

$$\frac{1}{n}\sum_{k=1}^{n} f(x - S_n + X_k)\,, \quad \text{resp.} \quad \frac{1}{n}\sum_{k=1}^{n} F(x - S_n + X_k)\,. \tag{5.1}$$

Fig. 4 illustrates the procedure for our recurrent example of estimating the density of the sum of $n = 10$ lognormals.



(a) **(b)**

Figure 4: $f(x - S_{n-1})$ solid, (5.1) dotted. (a) $R = 128$, (b) $R = 1024$

It is seen that the averaging procedure has the obvious advantage of producing a smoother estimate. This may be particularly worthwhile for small sample sizes R, as illustrated in Fig. 5. Here $R = 32$, the upper panel gives the histogram of the simulated 32 values of S_n and the lower panel the CdMC estimates, with the simple one in the first column and the averaged one in the second.

The overall performance of the idea involves two further aspects, computational effort and variance.

To asses the performance in terms of variance, consider estimation of the c.d.f. F and let

$$\begin{aligned} \omega^2 &= \mathbb{V}\mathrm{ar}\, F(x - S_{n-1}) = \mathbb{V}\mathrm{ar}\, F(x - S_n + S_k), \\ \rho &= \mathbb{C}\mathrm{orr}\big(F(x - S_n + S_k), F(x - S_n + S_\ell)\big), \qquad k \neq \ell. \end{aligned}$$

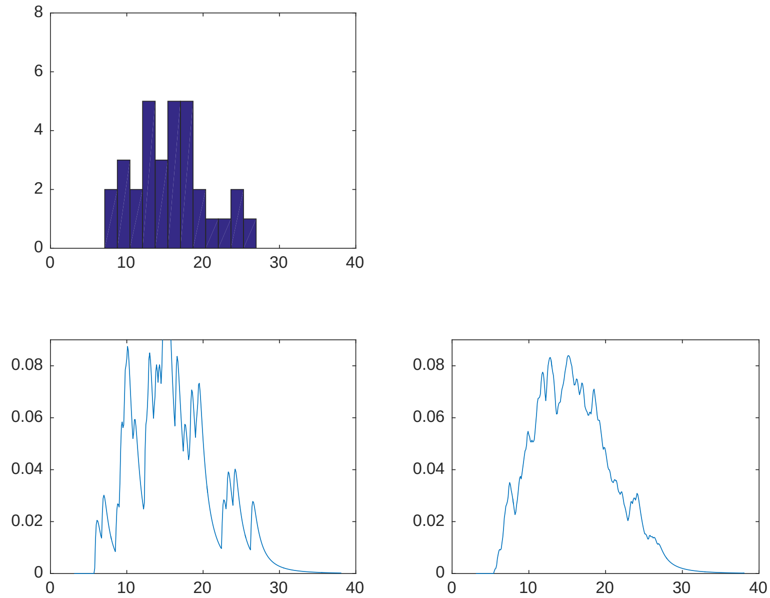


Figure 5: $R = 32$.
Upper panel simulated data, lower left $f(x - S_{n-1})$, lower right (5.1)

Then ω^2 is the variance of the simple CdMC estimator whereas that of the averaged one is $\omega^2[1/n + (1 - 1/n)\rho]$. Here $\rho = 0$ for $n = 2$, but one expects ρ to increase to 1 as n increases. The implication is that there is some variance reduction, but presumably it is only notable for small n. This is illustrated in Table 3, giving some numbers for the sum of lognormals$(0, 1)$. Within each column, the first entry vf_1 is the variance reduction factor $r_n(q_\alpha)$ for simple CdMC computed at the estimated α-quantile of S_n as given by Example 3.6, the second the same for averaged CdMC. For each entry, the two numbers correspond to the two values of α. The last column gives the empirical estimate of the correlation ρ as defined above.

Table 3: Comparison of simple and averaged CdMC

	$n = 5$	$n = 10$	$n = 25$	$n = 50$
vf_1	0.66 0.67	0.77 0.72	0.82 0.82	0.90 0.89
vf_n	0.49 0.50	0.64 0.60	0.74 0.74	0.84 0.82
ρ	0.67 0.67	0.81 0.82	0.90 0.89	0.93 0.93

When assessing the computational efficiency, it seems reasonable to compare with the alternative of using simple CdMC with a larger R than the one used for averaging. The choice between these two alternatives involves, however, features varying from case to case. Averaging has a plus if computation of densities is less costly than random variate generation, a minus the other way round.

6 The AK Estimator

The idea underlying the estimator Z_{AK} of Asmussen & Kroese (2006) is to combine an exchangeability argument with CdMC. More precisely (for convenience assuming existence of densities to exclude multiple maxima) one has $z = n\,\mathbb{P}(S_n > x, M_n = X_n)$ where $M_k = \max_{i \le k} X_i$. Applying CdMC with $\mathcal{F} = \sigma(X_1, \ldots, X_{n-1})$ to this expression the estimator comes out as

$$Z_{\mathrm{AK}}(x) = n\,\overline{F}\big(M_{n-1} \vee (x - S_{n-1})\big). \tag{6.1}$$

There has been a fair amount of follow-up work on Asmussen & Kroese (2006) and sharpenings, see in particular Hartinger and & Kortschak (2009), Chan & Kroese (2011), Asmussen *et al.* (2011), Asmussen & Kortschak (2012, 2015), Ghamami & Ross (2012) and Kortschak & Hashorva (2013). In summary, the state-of-the-art is that Z_{AK} not only has bounded relative error but in fact vanishing relative error in a wide class of heavy-tailed distributions. Here the relative (squared) error is the traditional measure of efficiency in the rare-event simulation literature, defined as the ratio $r_n^{(2)}(z)$ (say) between the variance and the square of the probability z in question (note that $r_n(x)$ is defined similarly in (3.3) but without the square in the denominator). Bounded relative error means $\limsup_{z\to 0} r_n^{(2)}(z) < \infty$ and is usually considered as the most one can hope for, cf. Asmussen & Glynn (2007) VI.1.

Theorem 6.1. *Assume that the distribution of X is either regularly varying, lognormal or Weibull with tail e^{-x^β} where $0 < \beta < \log(3/2)/\log 2 \approx 0.585$. Then there exists constants $\gamma > 0$ and $c < \infty$ depending on the distributional parameters such that*

$$\mathbb{V}\mathrm{ar}\, Z_{\mathrm{AK}}(x) \sim c x^{-\gamma} \mathbb{P}(S_n > x)^2 \qquad \text{as } x \to \infty.$$

The efficiency of the AK estimator for heavy-tailed F is apparent from Fig. 3a, where it outperforms simple CdMC. For light-tailed F it has been noted that Z_{AK} does not achieve bounded relative error, and presumably this is the reason it seems to have been discarded in this setting. For similar reasons as in Section 3, we shall not go into a general treatment of the efficiency of the AK estimator for light–tailed F, but only present the results for two basic examples when $n = 2$.

Example 6.2. Assume $n = 2$, $f(x) = \mathrm{e}^{-x}$. Then $M_{n-1} = X_{n-1} = X_1$ and $M_{n-1} > x - X_{n-1}$ precisely when $X_1 > x/2$. This gives

$$\begin{aligned}
\tfrac{1}{4}\mathbb{E} Z_{\mathrm{AK}}(x)^2 &= \int_0^{x/2} \mathrm{e}^{-2(x-y)} \mathrm{e}^{-y}\,\mathrm{d}y + \int_{x/2}^{\infty} \mathrm{e}^{-2y}\mathrm{e}^{-y}\,\mathrm{d}y \\
&= \mathrm{e}^{-2x}(\mathrm{e}^{x/2} - 1) + \tfrac{1}{3}\mathrm{e}^{-3x/2} \sim \tfrac{4}{3}\mathrm{e}^{-3x/2}, \qquad x \to \infty.
\end{aligned}$$

Compared to CrMC, this corresponds to an improvement of the error by a factor of order $\mathrm{e}^{-x/2}/x$.

Example 6.3. Let $n = 2$ and let F be normal$(0, 1)$. Calculations presented in the Appendix then give

$$\mathbb{V}\mathrm{ar}\, Z_{\mathrm{AK}}(x)^2 \sim \frac{64}{3x^3(2\pi)^{3/2}} \mathrm{e}^{-3x^2/8}, \qquad x \to \infty. \tag{6.2}$$

Compared to CrMC, this corresponds to an improvement of the error by a factor of order $\mathrm{e}^{-5x^2/8}$.

As discussed in Section 3, the variance reduction obtained via Z_{AK} reflects itself in improved estimates of the VaR. For the expected shortfall, Z_{AK} based algoritms are discussed in Hartinger & Kortschak (2009). They assume $\mathrm{VaR}_\alpha(S_n)$ to be known, but the discussion of Section A of the Appendix covers how to give confidence intervals if it is estimated.

Remark 6.4. For rare-event problems similar or related to that of estimating $\mathbb{P}(S_n > x)$, a number of alternative algorithms with similar efficiency as ZAK have later been developed, see for example Dupuis *et al.* (2007), Juneja (2007) and Blanchet & Glynn (2008). Some of these have the advantage of a potentially broader applicability, though Z_{AK} remains the one having the greater simplicity. $\lozenge$

7 Dependence

The current trend in dependence modeling is to use copulas and we shall here show some implementations of CdMC to this point of view. Among the many references in the area, Whelan (2004), McNeil (2006), Cherubini *et al.* (2007), Wu *et al.* (2007) and Mai & Scherer (2012) are of particular relevance for the following.

In the sum setting, we consider again the same marginal distribution F of $X_1, \ldots, X_n$. Let $(U_1, \ldots, U_n)$ be a random vector distributed according to the copula in question. Then S_n can be simulated by taking $X_i = F^{-1}(U_i)$, i.e. we can write

$$S_n = F^{-1}(U_1) + \cdots + F^{-1}(U_n)\,. \tag{7.1}$$

To estimate the c.d.f. or p.d.f. of S_n we then need the conditional c.d.f. and p.d.f. $F_{n|1:n-1}$, $f_{n\,|\,1:n-1}$ of X_n given a suitable σ-field $\mathcal{F}$ w.r.t. which $U_1, \ldots, U_{n-1}$ are measurable. Indeed, then

$$F_{n|1:n-1}(x - S_{n-1})\,, \qquad \text{resp. } f_{n\,|\,1:n-1}(x - S_{n-1}) \tag{7.2}$$

are unbiased estimates.

For a first example where $F_{n|1:n-1}$, $f_{n\,|\,1:n-1}$ are available, we consider Gaussian copulas. This is the case $U_i = \Phi(Y_i)$ where $\boldsymbol{Y}_n = (Y_1 \ldots Y_n)^\top$ is a multivariate normal vector with standard normal marginals and a general correlation matrix $\boldsymbol{C}$. In block-partitioned notation, we can write

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{n-1} & \boldsymbol{C}_{n-1,n} \\ \boldsymbol{C}_{n,n-1} & 1 \end{pmatrix}$$

where $\boldsymbol{C}_{n-1}$ is $(n-1)\times(n-1)$, $\boldsymbol{C}_{n-1,n}$ is $(n-1)\times 1$ and $\boldsymbol{C}_{n,n-1} = \boldsymbol{C}_{n-1,n}^\top$.

Proposition 7.1. *Consider the Gaussian copula and define*

$$\begin{aligned} c_{n|1:n-1} &= \boldsymbol{C}_{n,n-1}\boldsymbol{C}_{n-1}^{-1}\boldsymbol{C}_{n-1,n}\,, \\ \mu_{n|1:n-1} &= \boldsymbol{C}_{n,n-1}\boldsymbol{C}_{n-1}^{-1}(Y_1 \ldots Y_{n-1})^\top, \end{aligned}$$

$\mathcal{F} = \sigma(U_1, \ldots, U_n) = \sigma(Y_1, \ldots, Y_n)$. *Then*

$$F_{n|1:n-1}(y) = \Phi\Big(\big[\Phi^{-1}(F(y)) - \mu_{n|1:n-1}\big]/\sqrt{c_{n|1:n-1}}\Big) \tag{7.3}$$

$$f_{n|1:n-1}(y) = \frac{\varphi\Big(\big[\Phi^{-1}(F(y)) - \mu_{n|1:n-1}\big]/\sqrt{c_{n|1:n-1}}\Big)}{\varphi\big(\Phi^{-1}(F(y))\big)\sqrt{c_{n|1:n-1}}}\, f(y) \tag{7.4}$$

Proof. We have

$$\begin{aligned} F_{n|1:n-1}(y) &= \mathbb{P}(X_n \le y \,|\, \mathcal{F}) = \mathbb{P}\big(F^{-1}(\Phi(Y_n)) \le y \,|\, \mathcal{F}\big) \\ &= \mathbb{P}\big(Y_n \le \Phi^{-1}\big(F(y)\big) \,|\, \mathcal{F}\big) \end{aligned}$$

which reduces to (7.3) since the conditional distribution of Y_n given $\mathcal{F}$ is normal with mean $\mu_{n|1:n-1}$ and variance $c_{n|1:n-1}$. (7.4) then follows by differentiation. □

Example 7.2. The density of the sum S_n of $n = 10$ lognormals is in Fig. 6 for the Gaussian copula. The matrix $\boldsymbol{C}$ is taken as exchangeable, meaning that all off-diagonal elements are the same ρ, and various values of ρ are considered. ◊

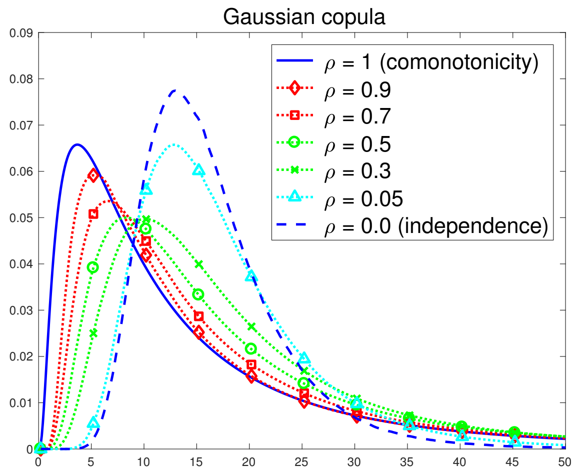


Figure 6: Density of a lognormal sum with an exchangeable Gaussian copula

As second example, we shall consider Archimedean copulas

$$\mathbb{P}(U_1 \le u_1, \ldots, U_n \le u_n) = \psi\big(\phi(u_1) + \cdots + \phi(u_n)\big) \tag{7.5}$$

where ψ is called the generator and ϕ is its functional inverse. Under the additional condition that the r.h.s. of (7.5) defines a copula for any n, it is known that ψ is the Laplace transform $\psi(s) = \mathbb{E}\mathrm{e}^{-sZ}$ of some r.v. $Z > 0$, and we shall consider only that case. A convenient representation is then

$$(U_1, \ldots, U_n) = \big(\psi(V_1/Z), \ldots, \psi(V_n/Z)\big) \tag{7.6}$$

where $V_1, \ldots, V_n$ are i.i.d. standard exponential and independent of Z. See, e.g., Marshall & Olkin (1988).

Proposition 7.3. *Define* $\mathcal{F} = \sigma(V_1, ..., V_{n-1}, Z)$. *Then*

$$F_{n|1:n-1}(y) = \exp\big\{-Z\phi\big(F(y)\big)\big\}\,, \tag{7.7}$$

$$f_{n|1:n-1}(y) = -Z\phi'\big(F(y)\big)\exp\big\{-Z\phi\big(F(y)\big)\big\}f(y)\,. \tag{7.8}$$

For the survival copula $\big(1 - \psi(V_1/Z), \ldots, 1 - \psi(V_n/Z)\big)$,

$$F_{n|1:n-1}(y) = 1 - \exp\big\{-Z\phi\big(\overline{F}(y)\big)\big\}\,, \tag{7.9}$$

$$f_{n|1:n-1}(y) = -Z\phi'\big(\overline{F}(y)\big)\exp\big\{-Z\phi\big(\overline{F}(y)\big)\big\}f(y)\,. \tag{7.10}$$

Proof. Formulas (7.8), (7.10) follow by straightforward differentiation of (7.7), (7.9), and (7.7) from

$$\begin{aligned} F_{n|1:n-1}(y) &= \mathbb{P}(X_n \le y \,|\, \mathcal{F}) = \mathbb{P}\bigl(F^{-1}(\psi(V_n/Z)) \le y \,|\, \mathcal{F}\bigr) \\ &= \mathbb{P}\bigl(V_n/Z \ge \phi(F(y)) \,|\, \mathcal{F}\bigr) = \exp\bigl\{-Z\phi\bigl(F(y)\bigr)\bigr\}. \end{aligned}$$

Similarly for the survival copula,

$$\begin{aligned} F_{n|1:n-1}(y) &= \mathbb{P}\bigl(F^{-1}(1-\psi(V_n/Z)) \le y \,|\, \mathcal{F}\bigr) = \mathbb{P}\bigl(\psi(V_n/Z)) \ge \overline{F}(y) \,|\, \mathcal{F}\bigr) \\ &= \mathbb{P}\bigl(V_n/Z \le \phi(\overline{F}(y)) \,|\, \mathcal{F}\bigr) = 1 - \exp\bigl\{-Z\phi\bigl(\overline{F}(y)\bigr)\bigr\}\,. \end{aligned}$$

□

Some numerical results follow for lognormal sums with the two most common Archimedean copulas, Clayton and Gumbel.

Example 7.4. The Clayton copula corresponds to Z being Gamma with shape parameter α. Traditionally, the parameter is taken as $\theta = 1/\alpha$ and the scale (which is unimportant for the copula) chosen such a that $\mathbb{E}Z = 1$. This means that the generator is $\psi(t) = 1/(1+t\theta)^{1/\theta}$ with inverse $\phi(y) = (y^{-\theta}-1)/\theta$,

The Clayton copula approaches independence as $\theta \downarrow 0$, i.e. $\alpha \uparrow \infty$, and approaches comonotonicity as $\theta \uparrow \infty$, i.e. $\alpha \downarrow 0$. The density of the sum S_n of $n = 10$ lognormals is in Fig. 7a for the Clayton copula itself and in Fig. 7b for the survival copula. The Clayton copula has tail independence in the right tail but tail dependence in the left, implying the opposite behaviour for the survival copula. Therefore the survival copula may sometimes be the more interesting one for risk management purposes in the Clayton case. ♢

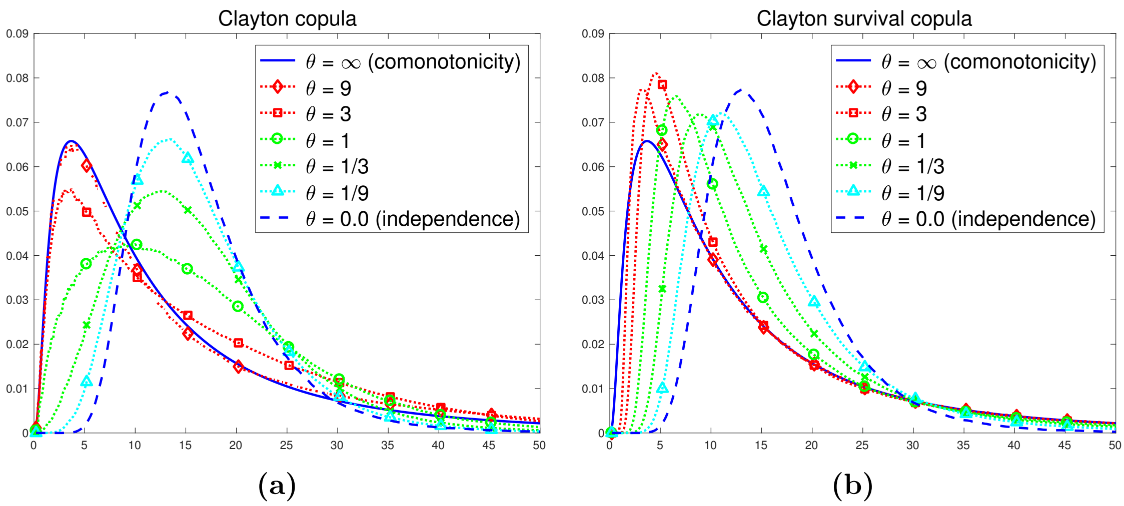
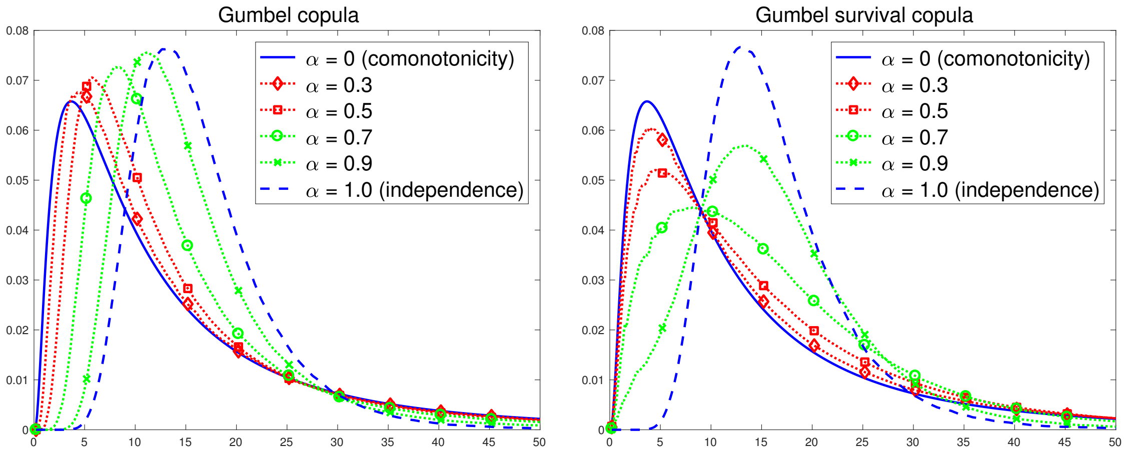


Figure 7: Density of lognormal sum with a Clayton copula

Example 7.5. The Gumbel copula corresponds to Z being strictly α-stable with support $(0,\infty)$. Traditionally, the parameter is taken as $\theta = 1/\alpha$ and the scale chosen such that the generator is $\psi(t) = \mathrm{e}^{-t^\alpha} = \mathrm{e}^{-t^{1/\theta}}$, with inverse $\phi(y) = (-\log y)^\theta$. The Gumbel copula approaches comonotonicity as $\theta \to \infty$, i.e. $\alpha \downarrow 0$, whereas independence corresonds to $\theta = \alpha = 1$. It has tail dependence in the right tail but tail independence in the left.



(a) **(b)**

Figure 8: Density of lognormal sum with a Gumbel copula

The density of the sum S_n of $n = 10$ lognormals is in Fig. 8a for the Gumbel copula itself and in Fig. 8b for the survival copula. $\Diamond$

Remark 7.6. Despite the simplicity of the Marshall-Olkin representation, much of the literature on conditional simulation in the Clayton (and other Archimedean) copulas concentrates on describing the conditional distribution of U_n given $U_1, \ldots, U_{n-1}$, see e.g. Cherubini *et al.* (2004). Even with this conditioning, we point out that it may be simpler to just consider the conditional distribution of Z given $U_1 = u_1, \ldots, U_{n-1} = u_{n-1}$. Namely, given $Z = z$ the r.v. $W_i = \phi(U_i) = V_i/z$ has density $z\mathrm{e}^{-zw_i}$ so that the conditional density must be proportional to the joint density

$$f_Z(z)z\mathrm{e}^{-zw_1} \cdots z\mathrm{e}^{-zw_{n-1}} \;\propto\; z^{1/\theta+n-2} \exp\bigl\{-z\bigl(1/\theta + w_1 + \cdots + w_{n-1}\bigr)\bigr\}$$

where the last expression uses $Z \sim \mathrm{gamma}(1/\theta, 1/\theta)$. This gives in particular that for the Clayton copula the conditional distribution of Z given $U_1 = u_1, \ldots, U_{n-1} = u_{n-1}$ is $\mathrm{gamma}(\alpha_n, \lambda_n)$ where

$$\alpha_n = 1/\theta + n - 1\,, \qquad \lambda_n = 1/\theta + \phi(u_1) + \cdots + \phi(u_{n-1})\,. \qquad \Diamond$$

Remark 7.7. Calculation of the VaR follows just the same pattern in the copula context as in the i.i.d. case, cf. the discussion around (4.2). One then needs to replace F with the conditional distribution in (7.2). Also the expected shortfall could be in principle be calculated by replacing the $\mathbb{E}(X - x)^+$ from the i.i.d. case with the similar conditional expectation. However, in examples one encounters the difficulty that the form of (7.2) is not readily ameneable to such computations. For example, in the Clayton copula with standard exponential marginals the conditional density is

$$\frac{Z}{(1-\mathrm{e}^{-x})^{\theta+1}} \exp\Bigl\{-Z\Bigl(\frac{1}{(1-\mathrm{e}^{-x})^{\theta}} - 1\Bigr)\Bigr\}\,\mathrm{e}^{-x}\,.$$

The expressions for say a lognormal marginal F are even less inviting! $\Diamond$

8 Concluding remarks

The purpose of the present paper has not been to promote the use of CdMC in all the problems looked into, but rather to present some discussion of both the potential and the limitations of the method. Two aspects were argued from the outset to be potentially attractive, variance reduction and smoothing.

As mentioned in Section 6, the traditional measure of efficiency in the rare-event simulation literature is the relative squared error $r_n^{(2)}(z)$, and bounded relative error (BdRelErr is usually consider as the most one can hope for. This and even more is obtained for Z_{AK}. Simple CdMC for the c.d.f. does not achieve bounded relative error, but nevertheless it was found to be worthwhile at least in the right tail of light-tailed sums and in the left tail when the increments are non-negative. For a quantitative illustration, consider estimating $\mathbb{P}(S_2 > x)$ in the normal$(0,1)$ case. The variances of the different estimators were all found to be of the form $Cx^\gamma \mathrm{e}^{-\beta x^2}$; note that for $\mathbb{P}(S_2 > x)$ itself (B.1) gives $C = 1/\sqrt{\pi}$, $\gamma = -1$, $\beta = 1/4$. A good algorithm thus corresponds to a large value of β and the values found in the respective results are given in Table 4. Note that estimates similar to those of Asmussen & Glynn (2007) VI.1 show that BdRelErr is in fact obtained by the IS-ECM algorithm sketched at the end of Section 3.

Table 4: δ in normal right tail, $n = 2$

CrMC	CdMC	AK	BdRelErr
$\frac{1}{4} = \frac{6}{24}$	$\frac{1}{3} = \frac{8}{24}$	$\frac{3}{8} = \frac{9}{24}$	$\frac{1}{2} = \frac{12}{24}$

Also the smoothing performance of CdMC came out favourably in the examples considered. Averaging as in Section 5 seemed to often be worthwhile. We found hat the ease with which CdMC produces plots of densities even in quite complicated models like the Clayton or Gumbel copulas in Figs. 7, 8 is a quite noteworthy property of the method.

In general, one could argue that when CdMC applies to either variance reduction or density estimation, it is at worst harmless and at best improves upon naive methods without involving more than a minor amount of extra computational effort. Some further comments:

1. When moving away from i.i.d. assumptions, we concentrated on dependence. Different marginals $F_1, \ldots, F_n$ can, however, also be treated by CdMC. For example, an obvious estimator for $\mathbb{P}(S_n > x)$ in this case is
$$\frac{1}{n}\sum_{i=1}^{n} \overline{F}_i(x - S_n + X_i)\,.$$
This generalizes in an obvious way to Z_{AK}. For discussion and extensions of these ideas, see e.g. Chan & Kroese (2011) and Kortschak & Harshova (2013).
2. The example we have treated is sums but the CdMC method is not restricted to this case. In general, it is of course a necessary condition to have enough structure that conditional distributions are computable in a simple form.

Acknowledgement

The MSc dissertation work of Søren Høg provided a substantial impetus for the present study. I am grateful to Patrick Laub for patiently looking into a number of MatLab queries.

References

Asmussen, S., Blanchet, J., Juneja, S. & Rojas-Nandayapa, L. (2011). Efficient simulation of tail probabilities of sums of correlated lognormals. *Annals of Operations Research*, **189**, 5–23.

Asmussen, S. & Glynn, P.W. (2007). *Stochastic Simulation. Algorithms and Analysis.* Springer-Verlag.

Asmussen, S., Jensen, J.L. & Rojas-Nandayapa, L. (2016). Exponential family techniques in the lognomal left tail. *Scand. J. Statist.*, **43**, 774–787.

Asmussen, S. & Kortschak, D. (2012). On error rates in rare event simulation with heavy tails, *Proceedings of the Winter Simulation Conference 2012.*

Asmussen, S. & Kortschak, D. (2015). Error rates and improved algorithms for rare event simulation with heavy Weibull tails. *Methodology and Computing in Applied Probability*, **17**, 441–461

Asmussen, S. & Kroese, D.P. (2006). Improved algorithms for rare event simulation with heavy tails *Adv. Appl. Probab.*, 38, 545–558.

Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.*, **37**, 577–580.

Balkema, A.A., Klüppelberg, C. & Resnick, S.I. (1993). Densities with Gaussian tails. *Proc. London Math. Soc.* **66**, 568–588.

Blanchet, J. & Glynn, P.W. (2008). Efficient rare-event simulation for the maximum of heavy-tailed random walks *Ann. Appl. Probab.*, 18, 1351–1378.

Chan, J. & Kroese, D. (2011). Rare-event probability estimation with conditional Monte Carlo. *Annals of Operations Research*, **189**, 43-61.

Cherubini, U., Luciano, E. & Vecchiato, W. (2004). *Copula Methods in Finance.* John Wiley & Sons.

Chu, F. & Nakayama, M.K. (2012). Confidence intervals for quantiles when applying variance-reduction techniques. *ACM TOMACS*, **22**.

Cline, D.B.H. (1986). Convolution tails, product tails and domains of attraction. *Probab. Th. Rel. Fields*, **72**, 529–557.

Dupuis, P. , Leder, K. & Wang, H. (2007). Large deviations and importance sampling for a tandem network with slow-down, *QUESTA*, **57**, 71–83.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events for Finance and Insurance.* Springer-Verlag.

Fu, M.C. Hong, L.J. & Hu, J.-Q. (2009). Conditional Monte Carlo estimation of quantile sensitivities. *Management Science*, **55**, 2019–2027.

Ghamami, S. and Ross, S.M. (2012). Improving the Asmussen–Kroese-type simulation estimators *J. Appl. Probab.*, **49**, 1188–1193.

Ghosh, J.K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.*, **42**, 1957–1961.

Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering.* Springer-Verlag.

Glasserman, P., Heidelberger, P. & Shahabuddin, P. (2000). Variance reduction techniques for estimating value-at-risk. *Management Science*, **46**, 1349–1364.

Glynn, P.W. (1996). Importance sampling for Monte Carlo estimation of quantiles. In *Proceedings of the 2nd St. Petersburg Workshop on Simulation.* Publishing House of St. Petersburg University, St. Petersburg, Russia, pp. 1800–185.

Gulisashvili, A. & Tankov, P. (2016). Tail behavior of sums and differences of log-normal random variables. *Bernoulli*, **22**, 444–493.

Hartinger, J. & Kortschak, D. (2009). On the efficiency of the Asmussen-Kroese-estimators and its application to stop-loss transforms. *Blätter DGVFM*, **30**, 363–377.

Jensen, J.L. (1995). *Saddlepoint Approximations.* Clarendon Press, Oxford

Juneja, S. (2007). Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error. *QUESTA*, **57**, 115–127.

Kortschak, D. & Hashorva, E. (2013). Efficient simulation of tail probabilities for sums of log-elliptical risks. *Journal of Computational and Applied Mathematics*, **47**, 53–67.

Mai, J.-F. & Scherer, M. (2012). *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications.* Imperial College Press.

Marshall, A.W. & Olkin, I. (1988). Families of multivariate distributions. *JASA*, **83**, 834–841.

McNeil, A.J. (2006). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, **78**, 567–581.

Nakayama, M.K. (2014). Quantile estimation when applying conditional Monte Carlo. In *Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH), 2014 International Conference on*, pp. 280–285. IEEE.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley.

Wu, F., Valdez , E. & Sherris, M. (2007). Simulating from exchangeable Archimedean copulas. *Comm. Statist. Simulation Comput.*, **36**, 1019–1034.

Whelan, N. (2004). Sampling from Archimedean copulas. *Quantitative Finance*, **4**, 339–352.

A Confidence intervals for expected shortfall

Consider estimation of a quantile q of the distribution F_n of S_n and the corresponding expexted shortfall $q + \int_q^\infty \overline{F}_n(x)\,\mathrm{d}x$, cf. (4.3). Let $F_R(x)$ be estimators of $F_n(x)$ such that

$$F_R(x) \sim F(x) - Z(x)/\sqrt{R}, \quad \overline{F}_R(x) \sim \overline{F}_n(x) + Z(x)/\sqrt{R}$$

as $R \to \infty$ for a suitable Gaussian process Z. For example, $\mathbb{V}\mathrm{ar}\, Z(x) = F_n(x)\overline{F}_n(x)$,

$$\mathbb{C}\mathrm{ov}\big(Z(x), Z(y)\big) = F_n(x \wedge y) - F(x)F(y) = \overline{F}_n(x \vee y) - \overline{F}(x)\overline{F}(y)$$

for the empirical c.d.f. For other examples, in particular CdMC, these quantities are typically not explicit but must be estimated from the simulation output and vary from case-to-case.

Let $\widehat{q}$ be the solution of $\alpha = F_R(\widehat{q})$. Since also $\alpha = F(q)$, it follows that

$$F(q) - F_R(q) = F_R(\widehat{q}) - F_R(q) \approx f(q)(\widehat{q} - q)$$

which gives the classical CLT

$$\widehat{q} - q \approx \frac{Z(q)}{f(q)\sqrt{R}} \tag{A.1}$$

see the remarks after (4.2) for references.

The obvious choice of estimator of the expexted shortfall e is then $\widehat{e} = \widehat{q} + \int_{\widehat{q}}^{\infty} \overline{F}_R(x)\,\mathrm{d}x$, and we get

$$\begin{aligned}
\widehat{e} &= \widehat{q} + \int_{\widehat{q}}^{\infty} \overline{F}_R(x)\,\mathrm{d}x = \widehat{q} + \int_{\widehat{q}}^{q} \overline{F}_R(x)\,\mathrm{d}x + \int_{q}^{\infty} \overline{F}_R(x)\,\mathrm{d}x \\
&= \widehat{q} + (q - \widehat{q})(1-\alpha) + \int_{q}^{\infty} \overline{F}_R(x)\,\mathrm{d}x \\
&= \widehat{q} - q + (\widehat{q} - q)(\alpha - 1) + q + \int_{q}^{\infty} \big\{\overline{F}(x) + Z(x)/\sqrt{R}\big\}\,\mathrm{d}x \\
&= \alpha(\widehat{q} - q) + e + \int_{0}^{\infty} Z(x)/\sqrt{R}\,\mathrm{d}x.
\end{aligned}$$

Hence

$$\widehat{e} - e \approx \frac{1}{\sqrt{R}}\Big(\frac{\alpha}{f(q)} Z(q) + \int_{q}^{\infty} Z(x)\,\mathrm{d}x\Big). \tag{A.2}$$

In the case of CdMC for an i.i.d. sum , the algorithm all together becomes:

1. Simulate $S_{n-1}^{(1)}, \ldots, S_{n-1}^{(R)}$.
2. Compute $\widehat{q}$ as solution of $\displaystyle\frac{1}{R}\sum_{r=1}^{R} \overline{F}\big(q - S_{n-1}^{(r)}\big) = 1 - \alpha$.
3. Let $\displaystyle\widehat{e} = \widehat{q} + \frac{1}{R}\sum_{r=1}^{R} m(\widehat{q} - S_{r;n-1})$ where $m(x) = \mathbb{E}(X - x)^+$.
4. Let $\displaystyle\xi_r = \frac{\alpha}{f(\widehat{q})}\overline{F}(\widehat{q} - S_{r;n-1}) + m(\widehat{q} - S_{r;n-1})$,
   $\displaystyle\overline{\xi} = \frac{1}{R}\sum_{r=1}^{R}\xi_r$, $\displaystyle s^2 = \frac{1}{R-1}\sum_{r=1}^{R}(\xi_r - \overline{\xi})^2$.
5. Return the 95% confidence interval $\widehat{e} \pm 1.96\, s/\sqrt{R}$.

B Technical proofs

Proof of Proposition 3.3. Using

$$\int_0^x y^a (x-y)^b\,\mathrm{d}y = B(a+1, b+1) x^{a+b+1}$$

we get

$$f^{*2}(x) \sim c_1^2 \int_0^x y^p (x-y)^p\,\mathrm{d}y = c_2\, x^{2p+1}$$

as $x \to 0$ and, by induction.

$$f^{*n}(x) \sim c_3 x^{np+n-1}, \quad F^{*n}(x) = \int_0^x f^{*n}(y)\,\mathrm{d}y \sim c_4 x^{np+n}\,.$$

Hence

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[F(x-S_{n-1}] &\sim \int_0^x c_5 y^{np+n-p-2} F(x-y)^2\,\mathrm{d}y - c_4^2 x^{2np+2n} \\
&= \int_0^x c_5 y^{np+n-p-2} c_1^2 (x-y)^{2p+2}\,\mathrm{d}y - c_4^2 x^{2np+2n} \\
&\sim c_6 x^{np+n+p+1}\,, \\
r_n(x) &\sim \frac{c_6 x^{np+n+p+1}}{c_4 x^{np+n}(1-c_4 x^{np+n})} \sim c_7 x^{p+1}\,. \qquad \square
\end{aligned}$$

Proof of Proposition 3.4. Let Z be a r.v. with tail $\overline{F}(x)^2$. By general subexponential theory, $\mathbb{P}(S_k > x) \sim k\overline{F}(x)$ for any fixed k and $\mathbb{P}(Y+Z > x) \sim \overline{F}(x)$ since Z therefore has lighter tail than S_{n-1}. Hence

$$\begin{aligned}
\mathbb{E}F(x-S_{n-1})^2 &= \overline{F}_{n-1}(x) + \mathbb{P}(S_{n-1}+Z > x, X \le x) = \overline{F}_{n-1}(x) + \mathrm{o}\big(\overline{F}(x)\big)\,, \\
\mathbb{V}\mathrm{ar}\,F(x-S_{n-1}) &= \overline{F}_{n-1}(x) + \mathrm{o}\big(\overline{F}(x)\big) - \mathrm{O}\big(\overline{F}(x)^2\big) \sim (n-1)\overline{F}(x)\,. \qquad \square
\end{aligned}$$

In the last two proofs, we shall need the Mill's ratio estimate of the normal tail, stating that if V is standard normal, then

$$\mathbb{P}(\sigma V > x) \begin{cases} \le \dfrac{\sigma}{x\sqrt{2\pi}}\mathrm{e}^{-x^2/2\sigma^2} & \text{for } x > 0, \\ \sim \dfrac{\sigma}{x\sqrt{2\pi}}\mathrm{e}^{-x^2/2\sigma^2} & \text{as } x \to \infty. \end{cases} \tag{B.1}$$

A slightly more general version, proved in the same way via L'Hospital, is

$$\int_{bx}^{\infty} \frac{1}{(y+cx)^k}\mathrm{e}^{-ay^2/2}\,\mathrm{d}y \sim \frac{1}{ab(b+c)^k x^{k+1}}\mathrm{e}^{-ab^2x^2/2}\,. \tag{B.2}$$

Proof of Proposition 3.5. We have

$$\mathbb{E}\overline{F}(x-s_{n-1})^2 = \int_{-\infty}^{\infty} \overline{\Phi}(x-y)^2 \frac{1}{\big(2\pi(n-1)\big)^{1/2}}\mathrm{e}^{-y^2/2(n-1)}\,\mathrm{d}y. \tag{B.3}$$

Let $\overline{H}(x) = \mathbb{P}(V_{n-1}+V_{1/2} > x)$ where $V_{n-1}, V_{1/2}$ are independent mean zero normals with variances $n-1$ and $1/2$, and note that

$$\mathbb{P}(V_{n-1} > x - A) = \mathrm{o}\big(\overline{H}(x)\big) \tag{B.4}$$

according to (B.1). The $y > x - A$ part of the integral in (B.3) is bounded by (B.4). Noting that

$$\frac{\overline{\Phi}(x)^2}{\mathbb{P}(V_{1/2} > x)} \sim \frac{\mathrm{e}^{-x^2}/x^2 2\pi}{1/2\cdot \mathrm{e}^{-x^2}/x\sqrt{2\pi}} = \frac{1}{x}\sqrt{2/\pi},$$

the $y \le x - A$ part asymptotically becomes

$$\begin{aligned}
&\int_{-\infty}^{x-A} \frac{\sqrt{2/\pi}}{x-y} \mathbb{P}(V_{1/2} > x - y) \frac{1}{\big(2\pi(n-1)\big)^{1/2}} \mathrm{e}^{-y^2/2(n-1)}\,\mathrm{d}y \\
&\qquad \sim \frac{1}{x}\sqrt{2/\pi}\,\mathbb{P}(V_{n-1} + V_{1/2} > x, V_{n-1} \le x - A) \\
&\qquad = \frac{1}{x}\sqrt{2/\pi}\,\overline{H}(x) - \mathrm{o}\big(\overline{H}(x)\big).
\end{aligned}$$

Thus

$$\begin{aligned}
r_n(x) &\sim \frac{\overline{H}(x)\sqrt{2/\pi}/x}{\mathbb{P}(S_n > x)} \sim \frac{\sqrt{n-1/2}\,\mathrm{e}^{-x^2/(2n-1)}/\sqrt{2\pi}\sqrt{2/\pi}/x}{\sqrt{n}\,\mathrm{e}^{-x^2/2n}/x\sqrt{2\pi}} \\
&\frac{1}{x}\sqrt{\frac{2n-1}{n\pi}}\mathrm{e}^{-x^2/[2n(2n-1)]}\,,
\end{aligned}$$

where we used (B.1) two times with $\sigma^2 = n - 1/2$, resp. $\sigma^2 = n$. □

Proof of (6.2). In the same way as in Example 6.2, $\max(M_{n-1}, x - S_{n-1}) = \max(X_1, x - X_1)$ splits up into $X_1 \le x/2$ and $X_1 > x/2$ parts. Using (B.3) to estimate $\overline{\Phi}(y)$, the $X_1 > x/2$ part of $\mathbb{E}Z^2_{\mathrm{AK}}(x)$ becomes

$$\frac{4}{\sqrt{2\pi}} \int_{x/2}^{\infty} \overline{\Phi}(y)^2 \mathrm{e}^{-y^2/2}\,\mathrm{d}y \sim \frac{4}{(2\pi)^{3/2}} \int_{x/2}^{\infty} \frac{1}{y^2} \mathrm{e}^{-3y^2/2}\,\mathrm{d}y.$$

The $X \le x/2$ part is

$$\begin{aligned}
&\frac{4}{\sqrt{2\pi}} \int_{-\infty}^{x/2} \overline{\Phi}(x-y)^2 \mathrm{e}^{-y^2/2}\,\mathrm{d}y = \frac{4}{\sqrt{2\pi}} \int_{x/2}^{\infty} \overline{\Phi}(y)^2 \mathrm{e}^{-(x-y)^2/2}\,\mathrm{d}y \\
&\quad \sim \frac{4}{(2\pi)^{3/2}} \int_{x/2}^{\infty} \frac{1}{y^2} \mathrm{e}^{-y^2-(x-y)^2/2}\,\mathrm{d}y \;= \frac{4}{(2\pi)^{3/2}} \mathrm{e}^{-x^2/3} \int_{x/2}^{\infty} \frac{1}{y^2} \mathrm{e}^{-3(y-x/3)^2/2}\,\mathrm{d}y \\
&\quad = \frac{4}{(2\pi)^{3/2}} \mathrm{e}^{-x^2/3} \int_{x/6}^{\infty} \frac{1}{(y+x/3)^2} \mathrm{e}^{-3y^2/2}\,\mathrm{d}y \\
&\quad = \frac{4}{(2\pi)^{3/2}} \mathrm{e}^{-x^2/3} \frac{1}{3 \cdot 1/6 \cdot (1/2)^2 x^3} \mathrm{e}^{-3(x/6)^2} = \frac{32}{3x^3(2\pi)^{3/2}} \mathrm{e}^{-3x^2/8}\,.
\end{aligned}$$

Adding up, the results follows. □