Supplement to "Compound multivariate Hawkes processes: Large deviations and rare event simulation"

RAVIAR S. KARIM[1,a], ROGER J. A. LAEVEN[2,b] and MICHEL MANDJES[3,c]

This text serves as an appendix to the paper "Compound multivariate Hawkes processes: Large deviations and rare event simulation". For context, notation and definitions, see the paper. We provide the proofs of some results and all simulation experiments.

Appendix A: Steepness in the univariate case

In this appendix, we prove steepness of the limiting cumulant in the univariate case. The appendix has two objectives. First, in this single-dimensional setting, elements that look intricate in the proof of Theorem 1 now simplify and become significantly more transparent; indeed, this univariate proof helps the reader navigating the proof that we gave for the multivariate case. Second, our proof is of a generic nature, in that it does not rely on the fact that in this univariate setting the distribution of the cluster size is explicitly known. This distinguishes our approach from the one followed in [39], which explicitly uses that the cluster size has a Borel distribution. In the multivariate case, the explicit distribution of the joint cluster size is unknown, thus prohibiting the approach of [39].

Consider the case of $d = d^\star = 1$; we leave out all indices. Let S denote the total number of events in a cluster. To guarantee stability of the Hawkes process, we assume that the univariate counterpart of Assumption 1 is in place, that is,

$$\mu = \mathbb{E}[B]c = \mathbb{E}[B]\int_0^\infty g(v)\,\mathrm{d}v \in (0,1). \tag{A.1}$$

We know that, under the univariate counterparts of Assumptions 2–3,

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}\big[e^{\theta Z(t)}\big] = \overline{\lambda}(\mathbb{E}\big[m_U(\theta)^S\big]-1) =: \Lambda(\theta), \tag{A.2}$$

where $m_U(\theta) = \mathbb{E}[e^{\theta U}]$ is the moment generating function of U and $\overline{\lambda}$ is the base rate. In this univariate case, the probability generating function $f(\cdot)$ of the cluster size S satisfies the fixed-point representation

$$f(z) = z\mathbb{E}\big[e^{Bc(f(z)-1)}\big]. \tag{A.3}$$

First, we take derivatives on both sides of Eqn. (A.3) to obtain

$$f'(z) = \mathbb{E}\big[e^{Bc(f(z)-1)}\big] + zf'(z)\mathbb{E}\big[Bc\,e^{Bc(f(z)-1)}\big],$$

which, due to (A.3), can be rewritten as

$$f'(z) = \frac{f(z)/z}{1-b(z)},$$

where $b(z) := z\,\mathbb{E}[Bc\,e^{Bc(f(z)-1)}]$. The next step is to show that there exists $\widehat{z} \geqslant 1$ such that $b(\widehat{z}) = 1$. The domain $\mathscr{D}_f$ is obtained from Proposition 1 when $d = 1$ as $\mathscr{D}_f = [0, \widehat{z}]$, where $\widehat{z} > 1$ is given by

$$\widehat{z} = \mathbb{E}[Bc\,e^{Bc(\widehat{x}-1)}]^{-1}, \tag{A.4}$$

and where $\widehat{x} > 1$ solves the equation

$$x\mathbb{E}[Bc\,e^{Bc(x-1)}] = \mathbb{E}[e^{Bc(x-1)}], \tag{A.5}$$

see also [28, Theorem 3.1.1]. Since $f'(z)$ is well-defined for $0 < z < \widehat{z}$, we consider what happens when we approach the boundary $\widehat{z}$. We compute, using Eqns. (A.4) and (A.5) and the fact $f(\widehat{z}) = \widehat{x}$, that

$$b(\widehat{z}) = \widehat{z}\,\mathbb{E}[Bc\,e^{Bc(\widehat{x}-1)}] = 1. \tag{A.6}$$

Hence, we obtain

$$\lim_{z \uparrow \widehat{z}} f'(z) = \lim_{z \uparrow \widehat{z}} \frac{f(z)/z}{1 - b(z)} = \infty, \tag{A.7}$$

also noting that for the numerator $f(\widehat{z})/\widehat{z} = \widehat{x}/\widehat{z} > 0$.

Concerning the steepness of $\Lambda(\cdot)$, let there be a $\widehat{\theta} > 0$ such that $m_U(\widehat{\theta}) = \widehat{z}$, which is the univariate counterpart of Assumption 3. We then have that

$$\liminf_{\theta \uparrow \widehat{\theta}} \Lambda'(\theta) \geqslant \overline{\lambda}\,\mathbb{E}[U]\,\mathbb{E}\big[S m_U(\widehat{\theta})^{S-1}\big] \geqslant \overline{\lambda}\,\mathbb{E}[U]\,f'(\widehat{z}) = \infty, \tag{A.8}$$

where the first inequality is due to Fatou's lemma, and the second inequality is due to $m'_U(\theta) \geqslant \mathbb{E}[U]$ for any $\theta \geqslant 0$.

Appendix B: Proof of Proposition 2

Proof. First note that $\theta^\star > 0$ is unique by combining the observations $\Psi_i(0) = 0$, $\Psi'_i(0) < 0$ as a consequence of (53), and the convexity of $\Psi_i(\cdot)$. The rest of the proof consists of two main steps: we first show that it suffices to identify the decay rate corresponding to the discrete-time counterpart of the process, and then we use [34, Theorem 3.1] to obtain the claim.

As a side remark, we mention that a *direct* proof, using essentially the same elements as those underlying [34, Theorem 3.1], can be devised. The respective lower bound then focuses on the contribution due to the most likely time scale of exceeding u, whereas in the upper bound it is shown that the contributions of other time scales vanish in the limit; cf. also the proof of similar results in [15, Theorems 2.1 & 2.2] and [10, Theorems 2.1 & 2.2].

Observe that, for $r < u$,

$$\mathbb{P}(\inf_{n \in \mathbb{N}} Y_i(n) \geqslant u) \leqslant p(u) \leqslant \mathbb{P}(\inf_{n \in \mathbb{N}} Y_i(n) \geqslant u - r),$$

from which we conclude that, if existing,

$$\lim_{u \to \infty} \frac{1}{u} \log \mathbb{P}(\inf_{t \geqslant 0} Y_i(t) \geqslant u) = \lim_{u \to \infty} \frac{1}{u} \log \mathbb{P}(\inf_{n \in \mathbb{N}} Y_i(n) \geqslant u).$$

Now we are in a position to invoke the discrete-time result of [34, Theorem 3.1]; notice that our decay rate $\theta^\star$ is the decay rate $\tilde{w}$ in [34]. The conditions of [34, Theorem 3.1] are fulfilled due to our net profit condition (53) and the assumed existence of the Lundberg root (56). □

Appendix C: Proof of Theorem 2

Proof. In order to eventually prove that our estimator is asymptotically efficient, the main idea is to find an upper bound on L_{τ_u}. By definition of the conditional intensities $\lambda_j(s)$ and $\lambda_j^{\mathbb{Q}}(s)$ in (59), we have

$$\exp\left(-\int_0^{\tau_u}\sum_{j=1}^d \big(1-f_j(\boldsymbol{m_U}(\boldsymbol{\theta}^\star))\big)\lambda_j(s)\mathrm{d}s\right)=\exp\left(-\int_0^{\tau_u}\sum_{j=1}^d \overline{\lambda}_j(1-f_j(\boldsymbol{m_U}(\boldsymbol{\theta}^\star)))\,\mathrm{d}s\right)$$
$$\times\exp\left(-\int_0^{\tau_u}\sum_{l=1}^d\sum_{j=1}^d\sum_{r=1}^{N_j(s)} B_{lj,r}\,g_{lj}(s-T_{j,r})(1-f_l(\boldsymbol{m_U}(\boldsymbol{\theta}^\star)))\,\mathrm{d}s\right),$$

where we switched the order of the summations over j and l. Then observe that, recalling that $\theta^\star$ solves the equation $\Psi_i(\theta^\star)=0$,

$$\exp\left(-\int_0^{\tau_u}\sum_{j=1}^d \overline{\lambda}_j(1-f_j(\boldsymbol{m_U}(\boldsymbol{\theta}^\star)))\,\mathrm{d}s\right)=\exp\left(-\tau_u\sum_{j=1}^d \overline{\lambda}_j(1-f_j(\boldsymbol{m_U}(\boldsymbol{\theta}^\star)))\right)=e^{r\theta^\star\tau_u}.$$

In addition, note that $Z_i(\tau_u)-r\tau_u>u$ due to the definition of τ_u, which implies that

$$e^{r\theta^\star\tau_u}e^{-\theta^\star Z_i(\tau_u)}\leqslant e^{-\theta^\star u}.$$

Also observe that since $c_{lj}=\|g_{lj}\|_{L^1(\mathbb{R}_+)}$, we have the bound

$$\exp\left(\int_0^{\tau_u}\sum_{j=1}^d\sum_{l=1}^d\sum_{r=1}^{N_j(s)} B_{lj,r}\,g_{lj}(s-T_{j,r})(f_j(\boldsymbol{m_U}(\boldsymbol{\theta}^\star))-1)\mathrm{d}s\right)$$
$$\leqslant\exp\left(\sum_{j=1}^d\sum_{l=1}^d\sum_{r=1}^{N_j(\tau_u)} B_{lj,r}\,c_{lj}(f_l(\boldsymbol{m_U}(\boldsymbol{\theta}^\star))-1)\right). \tag{C.1}$$

Finally, the contribution to the likelihood ratio L_{τ_u} due to the random marks is given by

$$\ell_j(\boldsymbol{v})=\exp(-\boldsymbol{v}^\top\bar{\boldsymbol{c}}^{\mathbb{Q}})\,m_{\boldsymbol{B}_j}(\bar{\boldsymbol{c}}^{\mathbb{Q}}), \tag{C.2}$$

where $\bar{\boldsymbol{c}}^{\mathbb{Q}}$ is the twist parameter for $\boldsymbol{B}_j$ as defined in Eqn. (49). As a consequence,

$$\exp\left(\sum_{j=1}^d\sum_{r=1}^{N_j(\tau_u)}\log\ell_j(\boldsymbol{B}_{j,r})\right)=\exp\left(-\sum_{j=1}^d\sum_{r=1}^{N_j(\tau_u)}\boldsymbol{B}_{j,r}^\top\bar{\boldsymbol{c}}^{\mathbb{Q}}\right)\prod_{j=1}^d m_{\boldsymbol{B}_j}(\bar{\boldsymbol{c}}^{\mathbb{Q}})^{N_j(\tau_u)}, \tag{C.3}$$

which implies that the expression given in (C.1) cancels against the exponential term in (C.3), noting that $\bar{\boldsymbol{c}}^{\mathbb{Q}}=\big(c_{1j}(f_1(\boldsymbol{m_U}(\boldsymbol{\theta}^\star))-1),\dots,c_{dj}(f_d(\boldsymbol{m_U}(\boldsymbol{\theta}^\star))-1)\big)^\top$.

Upon combining the above, and after some rewriting, we obtain that

$$p(u)\leqslant e^{-\theta^\star u}\,\mathbb{E}_{\mathbb{Q}}\left[\prod_{j=1}^d\exp\Big(-\log f_j(\boldsymbol{m_U}(\boldsymbol{\theta}^\star))+\log m_{\boldsymbol{U}_j}(\boldsymbol{\theta}^\star)+\log m_{\boldsymbol{B}_j}(\bar{\boldsymbol{c}}^{\mathbb{Q}})\Big)^{N_j(\tau_u)}\right].$$

As it turns out, the expression in the previous display simplifies considerably, as can be seen as follows. By (18), for any z,

$$\log f_j(z) = \log z_j + m_{\boldsymbol{B}_j}\big(c_{1j}(f_1(z)-1),\ldots,c_{dj}(f_d(z)-1)\big).$$

Now plugging in $\boldsymbol{z} = \boldsymbol{m_U}(\boldsymbol{\theta}^\star)$, we conclude that the expectation under the new measure $\mathbb{Q}$ fully reduces to unity, again by definition of $\bar{\boldsymbol{c}}^{\mathbb{Q}}$. This means that we have arrived at the upper bound $p(u) = \mathbb{E}_{\mathbb{Q}} L_{\tau_u} \leqslant e^{-\theta^\star u}$ and we are now in a position to conclude the statement. It follows directly from the observation

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}} L_{\tau_u} = \mathbb{E}_{\mathbb{Q}} L_{\tau_u}^2 - (\mathbb{E}_{\mathbb{Q}} L_{\tau_u})^2 \leqslant \mathbb{E}_{\mathbb{Q}} L_{\tau_u}^2 \leqslant e^{-2\theta^\star u},$$

in combination with Proposition 2. □

Appendix D: Proof of Theorem 3

Proof. We provide the proof for $d^\star = 1$, followed by a proof by example for $d^\star = 2$, which is easily extended for $d^\star > 2$. For $d^\star = 1$, we first observe that Theorem 1 yields that, using that we assumed $a_1 > \mu_1$,

$$\lim_{t\to\infty} \frac{1}{t} \log q_t(a_1) = -\inf_{x \geqslant a_1} \Lambda^\star(x) = -\Lambda^\star(a_1).$$

Define

$$\theta(a_1) := \arg\sup_\theta (\theta a_1 - \Lambda(\theta));$$

it is straightforward to verify that $\theta(a_1)$ is positive for $a_1 > \mu_1$. Letting I be the indicator function of the event $\{Z_1(t) \geqslant a_1 t\}$, $\mathbb{Q}$ the probability measure corresponding to exponentially twisting the original measure by $\theta(a_1)$, and L_t the appropriate likelihood, we now have that

$$q_t(a_1) = \mathbb{E}_{\mathbb{Q}}[L_t I].$$

As an aside, observe that in this setting, unlike the one discussed in Section 5.2, we do not have that $I = 1$ almost surely under $\mathbb{Q}$. This is an immediate consequence of the fact that we constructed $\mathbb{Q}$ such that $\lim_{t\to\infty} \mathbb{E}_{\mathbb{Q}} Z_1(t)/t = a_1$, so that the central limit theorem implies that in roughly half of the runs, we have that $I = 1$. The likelihood ratio can be evaluated by mimicking the calculations in Section 5.2. We thus obtain, leaving out indices in this single-dimensional case,

$$L_t = \exp\left(-\big(1 - f(m_U(\theta(a_1)))\big)\int_0^t \lambda(s)\,\mathrm{d}s\right) e^{-\theta(a_1) Z_1(t)}$$
$$\times \exp\left(\sum_{r=1}^{N(t)} \log \ell(B_r)\right)\left(\frac{m_U(\theta(a_1))}{f(m_U(\theta(a_1)))}\right)^{N(t)}.$$

Using that $\Lambda(\theta) = \overline{\lambda}(f(m_U(\theta)) - 1)$ and $m_B(c\,(f(z)-1)) - \log f(z) + \log z = 0$ (the latter identity being a consequence of (18)), and applying essentially the same majorizations as the ones used in Section 5.2, we readily obtain that

$$q_t(a_1) = \mathbb{E}_{\mathbb{Q}}[L_t I] \leqslant e^{\Lambda(\theta(a_1))\,t}\, e^{-\theta(a_1)\,Z_1(t)} I.$$

The event $\{I = 1\}$ is equivalent to $\{Z_1(t) \geqslant a_1 t\}$, so that

$$q_t(a_1) \leqslant e^{\Lambda(\theta(a_1))\,t}\, e^{-\theta(a_1)\,a_1 t} = e^{-\Lambda^\star(a_1)\,t}.$$

We have thus obtained that

$$q_t(a_1) \leqslant e^{-\Lambda^\star(a_1)\,t},$$

which can be seen as a variant of the classical Chernoff bound. The asymptotic efficiency for $d^\star = 1$ now follows directly. To this end, first note that using the very same reasoning we also find that

$$\mathbb{E}_{\mathbb{Q}}[L_t^2 I] \leqslant e^{-2\Lambda^\star(a_1)\,t},$$

so that also $\mathbb{V}\mathrm{ar}_{\mathbb{Q}}[L_t I] \leqslant e^{-2\Lambda^\star(a_1)\,t}$. Combining this with Theorem 1, we conclude that in this single-dimensional case, we have asymptotic efficiency under the measure $\mathbb{Q}$ defined above.

We now move to the case $d^\star = 2$, for which Theorem 1 gives

$$\lim_{t\to\infty} \frac{1}{t} \log q_t(\boldsymbol{a}) = -\inf_{(x_1,x_2)\in A} \Lambda^\star(\boldsymbol{x}).$$

Due to the convexity of the contour lines of $\Lambda^\star(\boldsymbol{a})$, with $\boldsymbol{a}^\star$ the optimizing $\boldsymbol{a} \in A$, three situations can occur: (i) $\boldsymbol{a}^\star = \boldsymbol{a}$, (ii) $a_1^\star = a_1$ and $a_2^\star > a_2$, and (iii) $a_1^\star > a_1$ and $a_2^\star = a_2$. As, by symmetry, cases (ii) and (iii) are conceptually the same and can therefore be treated identically, we restrict ourselves to discussing cases (i) and (ii) only. Let, as before, $\boldsymbol{\theta}(\boldsymbol{a}^\star)$ be the optimizing argument in the definition of $\Lambda^\star(\boldsymbol{a})$.

In case (i), using standard properties of the Legendre transform, we have that

$$\theta_1(\boldsymbol{a}^\star) = \frac{\partial}{\partial a_1}\Lambda^\star(\boldsymbol{a}^\star) > 0, \quad \theta_2(\boldsymbol{a}^\star) = \frac{\partial}{\partial a_2}\Lambda^\star(\boldsymbol{a}^\star) > 0.$$

We let $\mathbb{Q}$ correspond to the $\boldsymbol{\theta}(\boldsymbol{a}^\star)$-twisted version of the original probability measure. Going through the same steps as in the case $d^\star = 1$, we obtain that

$$q_t(\boldsymbol{a}) = \mathbb{E}_{\mathbb{Q}}[L_t I] \leqslant e^{\Lambda(\boldsymbol{\theta}(\boldsymbol{a}^\star))\,t}\, e^{-\theta_1(\boldsymbol{a}^\star)\,Z_1(t) - \theta_2(\boldsymbol{a}^\star)\,Z_2(t)}.$$

Then note that the right-hand side of the expression in the previous display is, on the set $\{I = 1\} = \{Z_1(t) \geqslant a_1^\star\, t, Z_2(t) \geqslant a_2^\star\, t\}$, bounded from above by

$$e^{\Lambda(\boldsymbol{\theta}(\boldsymbol{a}^\star))\,t}\, e^{-\theta_1(\boldsymbol{a}^\star)\,a_1^\star t - \theta_2(\boldsymbol{a}^\star)\,a_2^\star t} = e^{-\Lambda^\star(\boldsymbol{a}^\star)\,t}.$$

This implies $q_t(\boldsymbol{a}) \leqslant e^{-\Lambda^\star(\boldsymbol{a}^\star)\,t}$, but in addition that $\mathbb{V}\mathrm{ar}_{\mathbb{Q}}[L_t I] \leqslant e^{-2\Lambda^\star(\boldsymbol{a}^\star)\,t}$. We conclude, using the same reasoning as before, that in this case twisting by $\boldsymbol{\theta}(\boldsymbol{a}^\star)$ yields asymptotic efficiency.

Case (ii) works similarly. Observe that now (using that the line $x = a_1$ is a tangent of the contour lines of the Legendre transform)

$$\theta_1(\boldsymbol{a}^\star) = \frac{\partial}{\partial a_1}\Lambda^\star(\boldsymbol{a}^\star) > 0, \quad \theta_2(\boldsymbol{a}^\star) = \frac{\partial}{\partial a_2}\Lambda^\star(\boldsymbol{a}^\star) = 0.$$

The intuition is that in this case, if $Z_1(t) \geqslant a_1 t$, then with high probability also $Z_2(t) \geqslant a_2 t$, as reflected by the fact that

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\frac{Z_1(t)}{t} \geqslant a_1, \frac{Z_2(t)}{t} \geqslant a_2\right) = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\frac{Z_1(t)}{t} \geqslant a_1\right).$$

Let $\mathbb{Q}$ be the $\boldsymbol{\theta}(\boldsymbol{a}^\star)$-twisted version of the original probability measure. After straightforward algebra, we now obtain that

$$q_t(\boldsymbol{a}) = \mathbb{E}_{\mathbb{Q}}[L_t I] \leqslant e^{\Lambda(\boldsymbol{\theta}(\boldsymbol{a}^\star))\,t}\, e^{-\theta_1(\boldsymbol{a}^\star)\,Z_1(t)} = e^{-\Lambda^\star(\boldsymbol{a}^\star)\,t},$$

and $\mathbb{V}\mathrm{ar}_{\mathbb{Q}}[L_t I] \leqslant e^{-2\Lambda^\star(\boldsymbol{a}^\star)\,t}$. Hence, also in this case twisting by $\boldsymbol{\theta}(\boldsymbol{a}^\star)$ yields asymptotic efficiency.

It can be seen in a direct manner that the same procedure (i.e., working with a twist $\boldsymbol{\theta}(\boldsymbol{a}^\star)$ with non-negative entries) can be followed for any $d^\star$ larger than 2. We have thus established the stated result. □

Appendix E: Examples and numerical illustrations

In this section, we provide a set of simulation experiments that illustrate the proposed rare event simulation algorithms and assess the achievable efficiency gains relative to conventional simulation methods. All simulations have been conducted in `Python`; the computer code is available from the authors upon request. Throughout this section, we consider the bivariate setting for both the Hawkes process $\boldsymbol{N}(\cdot) = (N_1(\cdot), N_2(\cdot))^\top$ and the compound process $\boldsymbol{Z}(\cdot) = (Z_1(\cdot), Z_2(\cdot))^\top$, i.e., we set $d = d^\star = 2$.

E.1. Ruin probability

We focus in this subsection on the net cumulative claim process corresponding to the first component, i.e., $Y_1(t) = Z_1(t) - rt$. Our objective is to compute the ruin probability

$$p(u) = \mathbb{P}(\exists\, t > 0 : Y_1(t) > 0) = \mathbb{P}(\tau_u < \infty) = \mathbb{E}_{\mathbb{Q}}[L_{\tau_u}],$$

where we use the notations of Section 5. As before, it is assumed that the net profit condition (53) is in place. Let $p_n(u)$ denote our importance sampling estimator, see Eqn. (61).

We distinguish between the cases in which the marks are deterministic and random. It is anticipated that, due to the increased variability of the driving Hawkes process, the ruin probabilities will be larger under random marks than under deterministic marks (obviously assuming that the mean mark sizes in the random mark model equal their counterparts in the deterministic mark model). By applying our efficient simulation approach we can quantify this effect.

In the case of deterministic marks, the model primitives are assumed to take the following form. For $i, j = 1, 2$,

$$g_{ij}(t) \equiv g_i(t) = e^{-\alpha_i t}, \quad B_{ij} = \beta_{ij}, \quad U_{ij} \sim \mathrm{Exp}(u_{ij}),$$

where $\alpha_i, \beta_{ij}, u_{ij} > 0$. We note that under deterministic marks, the likelihood ratio L_{τ_u} given in Eqn. (60) simplifies considerably. Also note that in this case $m_{U_{ij}}(\theta) = u_{ij}(u_{ij} - \theta)^{-1}$ for any $\theta < u_{ij}$ and $i, j = 1, 2$. The specific parameters used in the simulation experiments are provided in the captions of the figures and tables that follow.

In order to be able to evaluate the likelihood ratio L_{τ_u}, we first calculate the 'twist vector' $(\theta^\star, 0)$, where $\theta^\star$ is found by solving $\Psi_1(\theta^\star) = 0$. We then exponentially twist $\boldsymbol{Y}(\cdot)$ by $(\theta^\star, 0)$, using the change of measure introduced in Section 5.1, enabling us to sample the likelihood ratio L_{τ_u} under the measure $\mathbb{Q}$, after which we can compute the importance sampling estimator $p_n(u)$. Recall that under the measure $\mathbb{Q}$, we have that the event $\{\tau_u < \infty\}$ happens with probability one, since the twisted process has positive drift and will hit any level $u > 0$. Figure 1 (left panel) illustrates the validity of Proposition 2 by showing

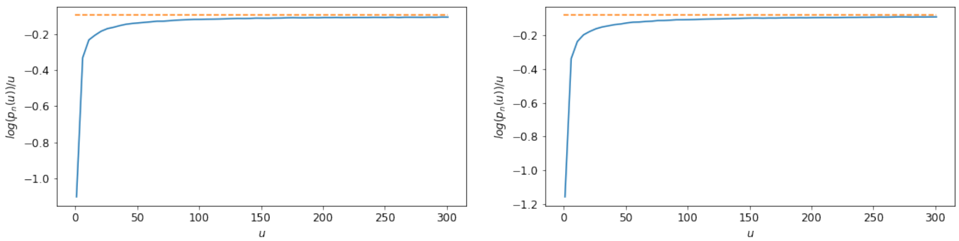

Figure 1. Left panel: Convergence of $u^{-1}\log p_n(u)$ to the logarithmic decay rate $-\theta^\star$ for the marginal process $Y_1(\cdot)$ in the bivariate model with deterministic marks. Chosen parameters are $\overline{\lambda}_1 = \overline{\lambda}_2 = 0.5$, $\alpha_1 = 2$, $\alpha_2 = 1.5$, $\beta_{11} = 0.5$, $\beta_{12} = 0.25$, $\beta_{21} = 0.3$, $\beta_{22} = 0.4$, $\mathbb{E}U_{11} = 2$, $\mathbb{E}U_{12} = \mathbb{E}U_{21} = 2.5$, $\mathbb{E}U_{22} = 3$ and $r = 8$. In this setting, solving $\Psi_1(\theta^\star) = 0$ yields $\theta^\star = 0.097$. Right panel: Convergence of $u^{-1}\log p_n(u)$ to the logarithmic decay rate $-\theta^\star$ for the marginal process $Y_1(\cdot)$ in the bivariate model with random marks. Chosen parameters are $\overline{\lambda}_1 = \overline{\lambda}_2 = 0.5$, $\alpha_1 = 2$, $\alpha_2 = 1.5$, $\mathbb{E}B_{11} = 0.5$, $\mathbb{E}B_{12} = 0.25$, $\mathbb{E}B_{21} = 0.3$, $\mathbb{E}B_{22} = 0.4$, $\mathbb{E}U_{11} = 2$, $\mathbb{E}U_{12} = \mathbb{E}U_{21} = 2.5$, $\mathbb{E}U_{22} = 3$ and $r = 8$. In this setting, solving $\Psi_1(\theta^\star) = 0$ yields $\theta^\star = 0.082$.

the convergence of $u^{-1}\log p_n(u)$ to the logarithmic decay rate $-\theta^\star$ as u grows large; it also provides insight into the speed of convergence for this specific instance.

In the case of random marks, we take

$$g_{ij}(t) \equiv g_i(t) = e^{-\alpha_i t}, \quad B_{ij} \sim \mathrm{Exp}(\gamma_{ij}), \quad U_{ij} \sim \mathrm{Exp}(u_{ij}),$$

where $\alpha_i, \gamma_{ij}, u_{ij} > 0$. To study the effect of the marks being random rather than deterministic, we take $\mathbb{E}B_{ij} = 1/\gamma_{ij} = \beta_{ij}$ (recalling that the β_{ij} were the deterministic marks that we have used in the first experiment). As before, we need to evaluate the likelihood ratio L_{τ_u}, for which we first solve the equation $\Psi_1(\theta^\star) = 0$. This requires solving the fixed-point equation (18) with random marks, where the i.i.d. assumption that was imposed on the marks B_{ij} implies that, for $j = 1, 2$, we have

$$f_j(z) = z_j \frac{\gamma_{1j}}{\gamma_{1j} - c_1(f_1(z) - 1)} \frac{\gamma_{2j}}{\gamma_{2j} - c_2(f_2(z) - 1)}. \tag{E.1}$$

Then we again exponentially twist $\boldsymbol{Y}(\cdot)$ by $(\theta^\star, 0)$ to sample under the measure $\mathbb{Q}$ and compute L_{τ_u}. Figure 1 (right panel) confirms convergence of $u^{-1}\log p_n(u)$ to $-\theta^\star$. We note that the decay rate $-\theta^\star$ is larger than in the case with deterministic marks, reflecting that the increased variability due to the random marks leads to a larger ruin probability.

Next, we study the efficiency of the proposed estimators in terms of the number of runs needed to reach a predefined level of precision. We continue generating runs until the relative standard error of the importance sampling estimators become smaller than a precision parameter $\epsilon > 0$. More precisely, given $\epsilon > 0$, we denote the relative standard error (after n runs, that is) and the number of required runs by

$$\epsilon_n := \frac{\sqrt{v_{n,\mathrm{IS}}(u)}}{p_n(u)\sqrt{n}}, \quad \widehat{n} := \inf\{n \in \mathbb{N} : \epsilon_n < \epsilon\}, \tag{E.2}$$

respectively, where

$$v_{n,\mathrm{IS}}(u) := \frac{1}{n}\sum_{m=1}^{n}\left(L_{\tau_u}^{(m)} - p_n(u)\right)^2.$$

u	$e^{-\theta_d^\star u}$	$p_{\widehat{n}_d}(u)$	$\widehat{n}_d$	$e^{-\theta_r^\star u}$	$p_{\widehat{n}_r}(u)$	$\widehat{n}_r$
1	$9.07 \cdot 10^{-1}$	$3.15 \cdot 10^{-1}$	109	$9.21 \cdot 10^{-1}$	$3.32 \cdot 10^{-1}$	150
2	$8.23 \cdot 10^{-1}$	$2.65 \cdot 10^{-1}$	122	$8.48 \cdot 10^{-1}$	$2.69 \cdot 10^{-1}$	173
5	$6.15 \cdot 10^{-1}$	$1.52 \cdot 10^{-2}$	128	$6.62 \cdot 10^{-1}$	$1.75 \cdot 10^{-1}$	199
10	$3.79 \cdot 10^{-1}$	$7.89 \cdot 10^{-2}$	136	$4.39 \cdot 10^{-1}$	$8.45 \cdot 10^{-2}$	232
20	$1.43 \cdot 10^{-1}$	$2.28 \cdot 10^{-2}$	142	$1.92 \cdot 10^{-1}$	$2.68 \cdot 10^{-2}$	267
50	$7.78 \cdot 10^{-3}$	$8.89 \cdot 10^{-4}$	204	$1.62 \cdot 10^{-2}$	$1.64 \cdot 10^{-3}$	386
100	$6.05 \cdot 10^{-5}$	$5.83 \cdot 10^{-6}$	245	$2.64 \cdot 10^{-4}$	$2.18 \cdot 10^{-5}$	533
200	$3.66 \cdot 10^{-9}$	$3.49 \cdot 10^{-10}$	302	$6.95 \cdot 10^{-8}$	$4.70 \cdot 10^{-9}$	595

Table 1. Lundberg bounds, ruin probabilities, and number of runs needed to reach a precision of $\epsilon = 0.05$, for deterministic (left panel) and random marks (right panel). Chosen parameters are: $\overline{\lambda}_1 = \overline{\lambda}_2 = 0.5$, $\alpha_1 = 2$, $\alpha_2 = 1.5$, $\mathbb{E}B_{11} = 0.5$, $\mathbb{E}B_{12} = 0.25$, $\mathbb{E}B_{21} = 0.3$, $\mathbb{E}B_{22} = 0.4$, $\mathbb{E}U_{11} = 2$, $\mathbb{E}U_{12} = \mathbb{E}U_{21} = 2.5$, $\mathbb{E}U_{22} = 3$ and $r = 8$. The twist parameters are $\theta_d^\star = 0.097$ and $\theta_r^\star = 0.082$.

We denote the number of runs required under deterministic marks by $\widehat{n}_d$ and under random marks by $\widehat{n}_r$. Clearly, $\widehat{n}_d$ and $\widehat{n}_r$ vary per experiment; we remedy this by performing the entire procedure multiple times and taking the average. We also display the associated Lundberg bounds, see Eqn. (62). The twist parameter corresponding to deterministic marks is denoted by $\theta_d^\star$, its counterpart for random marks is denoted by $\theta_r^\star$. The numbers in Table 1 confirm that random marks consistently lead to larger ruin probabilities. In addition, as expected, the number of runs needed grows in u at a very modest pace (despite the fact that $p(u)$ decays essentially exponentially in u).

In the remainder of this subsection, we consider the setting of random marks. In the next experiment we assess the computational advantage of the importance sampling estimator (using the measure $\mathbb{Q}$; indicated by subscript IS) when compared to the conventional Monte Carlo estimator (using the measure $\mathbb{P}$; indicated by subscript MC). Our goal is to compare the time it takes for both estimators to generate a sufficiently precise estimate of $p(u)$. For the IS estimator, we use Eqn. (E.2). For the conventional MC estimator based on n runs, denoted by $p_{n,\mathrm{MC}}(u)$, we have that

$$\epsilon_n = \frac{\sqrt{v_{n,\mathrm{MC}}(u)}}{p_{n,\mathrm{MC}}(u)\sqrt{n}} \approx \frac{1}{\sqrt{p_{n,\mathrm{MC}}(u)\, n}}, \tag{E.3}$$

since the variance $v_{n,\mathrm{MC}}(u) := p_{n,\mathrm{MC}}(u)(1 - p_{n,\mathrm{MC}}(u)) \approx p_{n,\mathrm{MC}}(u)$ for small probabilities.

In Table 2, we display the estimates of the ruin probabilities using MC and IS, including the average number of runs needed (in the table denoted by $\widehat{n}_{\mathrm{MC}}$ and $\widehat{n}_{\mathrm{IS}}$). As the absolute duration of each run highly depends on the specific hardware used, the programming language, the number of cores, etc., we decided to work with the *speedup ratio*, denoted by κ, which is the ratio of the simulation time needed under MC (to obtain the desired precision, that is) and its counterpart under IS.

Whereas MC is somewhat more efficient for (very) low values of u, IS dominates already for moderate values of u. For u larger than 60, it turned out to be even infeasible to obtain an MC estimate within a reasonable amount of time, whereas IS estimates can still be efficiently obtained. For instance, by extrapolating the results we found for smaller values of u, MC would take approximately 18 hours for $u = 70$ in our simulation environment; it would take even 1200 hours for $u = 100$. In these cases, we *estimated* κ by extrapolation of the running times under MC (growing effectively exponentially in u) and those under IS (growing effectively linearly in u); in the table these estimated values are given in italics. We conclude from the table that the speedup achieved by applying IS can be huge, particularly in the domain that ruin is rare; for $u = 300$ the speedup is expected to be as high as $9.00 \cdot 10^{15}$.

u	$p_{\widehat{n},\mathrm{MC}}(u)$	$\widehat{n}_{\mathrm{MC}}$	$p_{\widehat{n},\mathrm{IS}}(u)$	$v_{\widehat{n},\mathrm{IS}}(u)$	$\widehat{n}_{\mathrm{IS}}$	κ
1	$3.31 \cdot 10^{-1}$	807	$3.32 \cdot 10^{-1}$	$4.92 \cdot 10^{-2}$	150	$4.62 \cdot 10^{-1}$
2	$2.71 \cdot 10^{-1}$	1095	$2.69 \cdot 10^{-1}$	$3.47 \cdot 10^{-2}$	173	$3.60 \cdot 10^{-1}$
3	$2.39 \cdot 10^{-1}$	1357	$2.37 \cdot 10^{-1}$	$2.41 \cdot 10^{-2}$	187	$3.22 \cdot 10^{-1}$
5	$1.65 \cdot 10^{-1}$	2013	$1.75 \cdot 10^{-1}$	$1.61 \cdot 10^{-2}$	199	$3.85 \cdot 10^{-1}$
10	$8.40 \cdot 10^{-2}$	4368	$8.45 \cdot 10^{-2}$	$4.19 \cdot 10^{-3}$	232	$5.89 \cdot 10^{-1}$
20	$2.71 \cdot 10^{-2}$	14012	$2.68 \cdot 10^{-2}$	$4.92 \cdot 10^{-4}$	267	$1.38 \cdot 10^{0}$
30	$9.71 \cdot 10^{-3}$	38639	$1.02 \cdot 10^{-2}$	$7.29 \cdot 10^{-5}$	306	$3.84 \cdot 10^{0}$
40	$4.08 \cdot 10^{-3}$	96074	$4.01 \cdot 10^{-3}$	$1.16 \cdot 10^{-5}$	355	$1.12 \cdot 10^{1}$
50	$1.62 \cdot 10^{-3}$	263106	$1.64 \cdot 10^{-3}$	$2.44 \cdot 10^{-6}$	386	$3.96 \cdot 10^{1}$
60	$6.55 \cdot 10^{-4}$	747083	$6.46 \cdot 10^{-4}$	$4.36 \cdot 10^{-7}$	401	$1.43 \cdot 10^{2}$
70	n/a	n/a	$2.77 \cdot 10^{-4}$	$8.45 \cdot 10^{-8}$	410	*$4.98 \cdot 10^{2}$*
80	n/a	n/a	$1.12 \cdot 10^{-4}$	$1.64 \cdot 10^{-8}$	460	*$1.67 \cdot 10^{3}$*
100	n/a	n/a	$2.18 \cdot 10^{-5}$	$5.98 \cdot 10^{-10}$	533	*$2.04 \cdot 10^{4}$*
200	n/a	n/a	$4.70 \cdot 10^{-9}$	$3.51 \cdot 10^{-17}$	595	*$1.18 \cdot 10^{10}$*
300	n/a	n/a	$1.25 \cdot 10^{-12}$	$2.71 \cdot 10^{-24}$	683	*$9.00 \cdot 10^{15}$*

Table 2. Ruin probabilities, number of runs needed to reach a precision of $\epsilon = 0.05$, and speedup ratio κ, using Monte Carlo (MC) and Importance Sampling (IS). Chosen parameters are as in the caption of Table 1.

E.2. Exceedance probability

In this subsection, we numerically illustrate the rare event simulation procedure proposed in Section 5.3. We consider the simulation-based computation of the bivariate exceedance probability

$$q_t(a_1, a_2) = \mathbb{P}\left(\frac{Z_1(t)}{t} \geqslant a_1, \frac{Z_2(t)}{t} \geqslant a_2\right),$$

where we assume that we do not have that $a_1 \leqslant \mu_1 := \lim_{t\to\infty} Z_1(t)/t$ and $a_2 \leqslant \mu_2 := \lim_{t\to\infty} Z_2(t)/t$ to ensure that we are dealing with an event that becomes increasingly rare as $t \to \infty$. By Theorem 1,

$$\lim_{t\to\infty} \frac{1}{t} \log q_t(a_1, a_2) = -\inf_{(x_1,x_2)\in A} \Lambda^\star(x_1, x_2), \tag{E.4}$$

where $A = [a_1, \infty) \times [a_2, \infty)$. Denote the minimizer of the RHS of (E.4) by $\boldsymbol{a}^\star = (a_1^\star, a_2^\star)$, which can be obtained using standard optimization techniques since $\Lambda^\star(\cdot)$ is a convex function.

Consider estimating $q_t(a_1, a_2)$ by importance sampling. More precisely, with $I_{(a_1,a_2)}$ denoting the indicator of the event we are interested in, and L_t the likelihood ratio given in Eqn. (60), we have

$$q_t(a_1, a_2) = \mathbb{E}_{\mathbb{Q}}[L_t I_{(a_1,a_2)}]. \tag{E.5}$$

Let $q_{t,n}(a_1, a_2)$ be the importance sampling estimator for $q_t(a_1, a_2)$ as defined in Eqn. (63). To find the twist parameter, we solve the optimization problem

$$\boldsymbol{\theta}(\boldsymbol{a}^\star) = (\theta_1(a_1^\star, a_2^\star), \theta_2(a_1^\star, a_2^\star)) = \arg\sup_{\boldsymbol{\theta}} (\boldsymbol{\theta}^\top \boldsymbol{a}^\star - \Lambda(\boldsymbol{\theta})).$$

Figure 2 illustrates the behavior of $q_t(a_1, a_2)$ as t grows, converging to $-\inf_{\boldsymbol{x}\in A} \Lambda^\star(\boldsymbol{x}) = -\Lambda^\star(\boldsymbol{a})$, as stated in Eqn. (E.4). We choose $a_1 = 10 > 3.90 = \mu_1$ and $a_2 = 12 > 4.76 = \mu_2$, such that $\{Z_1(t) \geqslant a_1 t, Z_2(t) \geqslant a_2 t\}$ is an increasingly rare event as t grows. For our specific parameters $\boldsymbol{\theta}(\boldsymbol{a}^\star) > \mathbf{0}$ (componentwise, that is), corresponding to $\boldsymbol{a}^\star = \boldsymbol{a} = (a_1, a_2)$. Intuitively, this means that the most probable

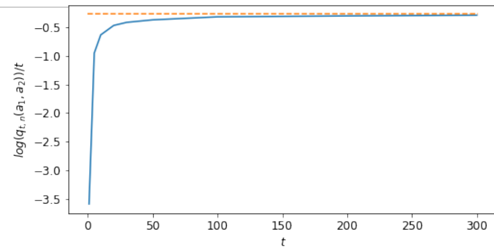

Figure 2. Convergence of $t^{-1}\log q_{t,n}(a_1,a_2)$ to the logarithmic decay rate $-\Lambda^\star(\boldsymbol{a}^\star)$ for the bivariate process $(Z_1(t)/t, Z_2(t)/t)_{t\in\mathbb{R}_+}$. Chosen parameters are: $a_1 = 10$, $a_2 = 12$, $\overline{\lambda}_1 = \overline{\lambda}_2 = 0.5$, $\alpha_1 = 2$, $\alpha_2 = 1.5$, $\mathbb{E}B_{11} = 0.5$, $\mathbb{E}B_{12} = 0.25$, $\mathbb{E}B_{21} = 0.3$, $\mathbb{E}B_{22} = 0.4$, $\mathbb{E}U_{11} = 2$, $\mathbb{E}U_{12} = \mathbb{E}U_{21} = 2.5$, $\mathbb{E}U_{22} = 3$. Since $\boldsymbol{a}^\star = \boldsymbol{a}$, the decay rate is $-\Lambda^\star(\boldsymbol{a}) = -0.276$ and the twist parameter is $\boldsymbol{\theta}^\star(\boldsymbol{a}) = (0.0376, 0.0256)$.

way in which the process $(Z_1(t)/t, Z_2(t)/t)_{t\in\mathbb{R}_+}$ reaches the region $A = [a_1,\infty)\times[a_2,\infty)$ is a straight line from the origin to (a_1, a_2).

Next, we quantify the computational advantage of the IS estimator over the conventional MC estimator, using the same approach as the one underlying Table 2. As before, we run simulations for both methods until the relative standard error ϵ_n is under the desired level of precision ϵ. Table 3 displays the comparison between MC and IS, for different values of t. As t increases, MC becomes infeasible due to the very steeply increasing number of runs needed as well as the simulation time needed per run. Already at $t = 15$, it would take approximately 69 hours in our simulation environment. IS, however, remains feasible, even in the domain of extremely small probabilities. Note that the number of runs needed for IS does initially not increase in a monotone fashion, which is due to the fact that for small t the process is not yet in the regime where the exceedance event is rare. We also note the speedup ratio κ of the exceedance probabilities increases more steeply (in t) than that of the ruin probabilities (in u).

t	$q_{t,\widehat{n},\text{MC}}(\boldsymbol{a})$	$\widehat{n}_{\text{MC}}$	$q_{t,\widehat{n},\text{IS}}(\boldsymbol{a})$	$v_{t,\widehat{n},\text{IS}}(\boldsymbol{a})$	$\widehat{n}_{\text{IS}}$	κ
1	$2.58\cdot10^{-2}$	16620	$2.61\cdot10^{-2}$	$8.27\cdot10^{-3}$	4879	$6.56\cdot10^{-1}$
2	$1.83\cdot10^{-2}$	21983	$1.78\cdot10^{-2}$	$3.47\cdot10^{-3}$	4378	$9.98\cdot10^{-1}$
3	$1.45\cdot10^{-2}$	29076	$1.42\cdot10^{-2}$	$1.92\cdot10^{-3}$	3800	$1.63\cdot10^{0}$
5	$7.88\cdot10^{-3}$	50869	$7.95\cdot10^{-3}$	$5.67\cdot10^{-4}$	3589	$3.66\cdot10^{0}$
10	$1.75\cdot10^{-3}$	205934	$1.69\cdot10^{-3}$	$2.73\cdot10^{-5}$	3818	$3.04\cdot10^{1}$
15	n/a	n/a	$3.64\cdot10^{-4}$	$1.40\cdot10^{-6}$	4223	*$3.24\cdot10^{2}$*
20	n/a	n/a	$7.83\cdot10^{-5}$	$7.10\cdot10^{-8}$	4631	*$3.92\cdot10^{3}$*
25	n/a	n/a	$1.87\cdot10^{-5}$	$4.48\cdot10^{-9}$	5123	*$5.04\cdot10^{4}$*
30	n/a	n/a	$3.84\cdot10^{-6}$	$2.23\cdot10^{-10}$	6043	*$6.25\cdot10^{5}$*
40	n/a	n/a	$2.18\cdot10^{-7}$	$8.05\cdot10^{-13}$	6749	*$1.28\cdot10^{8}$*
50	n/a	n/a	$1.15\cdot10^{-8}$	$2.69\cdot10^{-15}$	8209	*$3.43\cdot10^{10}$*
75	n/a	n/a	$9.25\cdot10^{-12}$	$2.16\cdot10^{-24}$	10106	*$3.93\cdot10^{16}$*
100	n/a	n/a	$6.31\cdot10^{-15}$	$1.44\cdot10^{-27}$	11926	*$5.56\cdot10^{22}$*

Table 3. Estimation of the exceedance probability $q_t(\boldsymbol{a})$, number of runs needed to reach a precision of $\epsilon = 0.05$, and speedup ratio κ, using Monte Carlo (MC) and Importance Sampling (IS). Chosen parameters are as in the caption of Figure 2.