THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

# Asymptotic Behavior of Total Times For Jobs That Must Start Over If a Failure Occurs

Søren Asmussen, Pierre Fiorini, Lester Lipsky, Tomasz Rolski and Robert Sheahan

Research Report

No. 09 | June 2007

# Asymptotic Behavior of Total Times For Jobs That Must Start Over If a Failure Occurs

This Thiele Research Report is also Research Report number 496 in the Stochastics Series at Department of Mathematical Sciences, University of Aarhus, Denmark.

# Asymptotic Behavior of Total Times For Jobs That Must Start Over If a Failure Occurs.

Søren Asmussen[1], Pierre Fiorini[2], Lester Lipsky[3]
Tomasz Rolski[4], & Robert Sheahan[3]

June 20, 2007

**Abstract**

Many processes must complete in the presence of failures. Different systems respond to task failure in different ways. The system may resume a failed task from the failure point (or a saved checkpoint shortly before the failure point), it may give up on the task and select a replacement task from the ready queue, or it may restart the task. The behavior of systems under the first two scenarios is well documented, but the third (*RESTART*) has resisted detailed analysis. In this paper we derive tight asymptotic relations between the distribution of *task times* without failures to the *total time* when including failures, for any failure distribution. In particular, we show that if the task time distribution has an unbounded support then the total time distribution $H$ is always heavy-tailed. Asymptotic expressions are given for the tail of $H$ in various scenarios. The key ingredients of the analysis are the Cramér–Lundberg asymptotics for geometric sums and integral asymptotics, that in some cases are obtained via Tauberian theorems and in some cases by bare-hand calculations.

**Key words** Cramér-Lundberg approximation, failure recovery, geometric sums, heavy tails, logarithmic asymptotics, mixture distribution, power tail, RESTART, Tauberian theorem

---

[1]Department of Mathematical Sciences, Aarhus University, Ny Munkegade, DK-8000 Aarhus C, Denmark

[2]Department of Computer Science, University of Southern Maine, Portland, Maine, USA

[3]Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06269-2155, USA

[4]Mathematical Institute, Wroclaw University, 50-384 Wroclaw, Poland

Partially upported by a Marie Curie Transfer of Knowledge Fellowship: Programme HAHAP MTKO-CT-2004-13389

# 1 Introduction

For many systems failure is rare enough that it can be ignored, or dealt with as an afterthought. For other systems, failure is common enough that the design choice of how to deal with it may have a significant impact on the performance of the system. Consider a job that ordinarily would take a time $T$ to be executed on some system (e.g., CPU). If at some time $U < T$ the processor fails, the job may take a *total time* $X \geq T$ to complete. We let $F, G$ be the distributions of $T, U$ and $H = H_{F,G}$ the distribution of $X$ which in addition to $F, G$ depends on the failure recovery scheme.

Many papers discuss methods of failure recovery and analyze their complexity in one or more metrics, like *restartable processors* in Chlebus *et al.* [7], or *stage checkpointing* in De Prisco *et al.* [8], etc. There are many specific and distinct failure recovery schemes, but they can be grouped into three broad classes:

*RESUME*, also referred to as preemptive resume (prs);
*REPLACE*, also referred to as preemptive repeat different (prd);
*RESTART*, also referred to as preemptive repeat identical (pri).

The analysis of the distribution function $H(x) = \mathbb{P}(X \leq x)$ when the policy is *RESUME* or *REPLACE* was carried out by Kulkarni *et al.* [14], [15]. In the *RESUME* scenario, if there is a processor failure while a job is being executed, after repair is implemented the job can continue where it left off. All that is required mathematically is to remember the state of the system when failure occurred. If repair time is an issue then the number of failures before final completion must also be considered. In what follows, we ignore the time for repairs, with the knowledge that this can be properly handled separately. In the *REPLACE* situation, if a job fails, it is replaced by a different job from the same distribution. Here, no details concerning the previous job are necessary in order to continue.

The work by Kulkarni *et al.* [14], [15], and Bobbio & Trivedi [4] clearly suggests that if $F$ is phase-type or, more generally, matrix-exponential ([16], [1], [2]), and $\overline{G}(u) = \mathbb{P}(U > u) = \mathrm{e}^{-\beta u}$, then $H$ for the *RESUME* and *REPLACE* policies can also be represented by matrix-exponential distributions. This means that they could be analyzed entirely within a Markov chain framework.

However, the RESTART policy has resisted detailed analysis. The total time distribution $H$ under this policy was defined and examined through its Laplace transform in Kulkarni *et al.* [14], [15]. They were able to show that it definitely was *not* matrix-exponential, i.e., the Laplace transform cannot be rational, and therefore it cannot be solved in the Markov Chain framework. However, by numerically taking the inverse Laplace transform (see Jagerman [10]), Chimento & Trivedi [6] (following a model proposed by Castillo [5]) were able to find the RESTART time distribution for a few cases, for a limited range of the total time ($x \leq 3\mathbb{E}T$). The method seems to be unstable for larger $x$. It is this problem that interests us here.

There are many examples of where the RESTART scenario is relevant. The obvious one alluded to above involves execution of a program on some computer. If the computer fails, and the intermediate results are not saved externally (e.g., by *checkpointing*), then the job must restart from the beginning. As another example, one might wish to copy a file from a remote system using some standard protocol as FTP or HTTP. The time it takes to copy a file is proportional to its length. A transmission error immediately aborts the copy and discards the partially received data, forcing the user to restart the copy from the beginning. Yet another example would be receiving 'customer service' by telephone. Often, while dealing with a particular service agent, the connection is broken. Then the customer must redial the service center, and invariably (after waiting in a queue) end up talking to a different agent, and have to explain everything from the beginning.

In our previous paper (Sheahan *et al.* [18]), we derived an expression for the Laplace transform of the total time distribution $H$ for the RESTART policy with exponential failure rate, $\beta$. We used it to get an expression for the moments $\mathbb{E}X^\ell$ of the total time. From this we were able to argue that if the task-time distribution has an exponential tail, then $X$ has infinite moments for $\ell \geq \alpha = \lambda/\beta$, where $\lambda$ is the rate of the exponential tail (i.e., $\overline{F}(t) \sim c\mathrm{e}^{-\lambda t}$). This in turn implies that roughly $\overline{H}(x) \approx c/x^\alpha$, i.e., $X$ is *power-tailed*.

This can have important implications, particularly in applications where the time to finish a task is bounded by necessity. If a task takes too long to complete it must be aborted, and an alternate solution provided. In such applications it may be important to know $\overline{H}(x)$, for that is the probability that a job will be aborted. Power tails and *heavy tails*, generally, have a small but non-negligible probability of lasting for many, many times the mean, and thus $\overline{H}(x)$ for large $x$ can be important.

In this paper we derive the asymptotic behavior of $\overline{H}(x)$ as $x \to \infty$ under more general assumptions than in [18] and in sharper form in a number of important cases. As a first guess, one could believe that the heaviness of $\overline{H}$ is determined by the heaviness of $\overline{F}$ and/or $\overline{G}$. However, it turns out that the important feature is rather how close are $\overline{F}$ and $\overline{G}$. This is demonstrated in a striking way by the following result for the diagonal case:

**Proposition 1.1** *If $F = G$, then $\overline{H}(x) \sim \dfrac{1}{\mu x}$.*

Here $\mu = 1/\mathbb{E}U$; we assume throughout in the paper that $\mu > 0$ and, for convenience, that $F, G$ have densities $f, g$ (this assumption can be relaxed at many places but we will not give the details). It is notable that no other conditions are required for Proposition 1.1, in particular no precise information on how heavy the common tail $\overline{F} = \overline{G}$ is!

The assumption that the task time distribution $F$ and the failure time distribution $G$ be identical of course lacks interpretation in the RESTART setting. Thus, Proposition 1.1 is more of a curiosity, which is further illustrated by the fact that a proof can be given which is far simpler than the our proofs for more general situations (see Section 6). Nevertheless, the result indicates that the tail behaviour of $H$ depends on a delicate balance between the tails of $F$ and $G$. We will also see that making $\overline{F}$ heavier makes $\overline{H}$ heavier, making $\overline{G}$ heavier makes $\overline{H}$ lighter. However, except for the case when $F$ has a finite support, $\overline{H}$ is always heavy-tailed:

**Proposition 1.2** *Assume that the support of $F$ is unbounded. Then $\mathrm{e}^{\epsilon x}\overline{H}(x) \to \infty$ for any $\epsilon > 0$.*

In general, we will be able to obtain sharp asymptotics for $\overline{H}(x)$ when $\overline{F}$ and $\overline{G}$ are not too far away. The form of the result (Theorem 2.2) is regular variation of $\overline{H}$. For example, the following result covers Gamma distributions:

**Corollary 1.1** *Assume $f, g$ belong to the class of densities of asymptotic form $c\mathrm{e}^{-\lambda t}t^{\alpha-1}$, with parameters $\lambda_F, \alpha_F, c_F$ for $f$ and $\lambda_G, \alpha_G, c_G$ for $g$. Then $\overline{H}(x) \sim c_H \log^{\alpha_F - \alpha_H\alpha_G} x / x^{\alpha_H}$, where $\alpha_H = \lambda_F/\lambda_G$ and*

$$c_H \;=\; \frac{c_F\Gamma(\alpha_H)\lambda_G^{\alpha_H-1-\alpha_F+\alpha_H\alpha_G}}{\mu^{\alpha_H}c_G^{\alpha_H}}\,.$$

Numerical illustrations are given in [18] for $\alpha_G = 1$ (i.e., $G$ exponential) and show an excellent fit.

When $\overline{F}$ and $\overline{G}$ are more different (say $F$ has a power tail and $G$ is exponential), we will derive logarithmic asymptotics for $\overline{H}$. We will see forms varying from extremely heavy tails like $1/\log^\alpha x$ over power tails $1/x^\alpha$ to moderately heavy tails like the Weibull tail $\mathrm{e}^{-x^\beta}$ with $\beta < 1$.

The proofs of the paper are based on the representation

$$X \;=\; T + S \;\text{ where }\; N \;=\; \inf\{n : U_{n+1} > T\}\;,\quad S \;=\; \sum_{i=1}^N U_i\,, \tag{1}$$

and $U_1, U_2, \ldots$ are the succesive failure times (assumed i.i.d. with distribution $G$ and independent of $T$). More precisely, we will use that given $T = t$, $S(t) = \sum_1^N U_i$ is a compound geometric sum for which exponential Cramér-Lundberg tail asymptotics is available, and uncondition to get our final results. In Section 2 we state our main results, except for the case of a bounded task time $T$ which is treated in Section 3. The analysis there departs from a careful study of the case $T \equiv t$. Section 4 is devoted to the proof of the following lemma, which is the key to the unbounded case:

**Lemma 1.1** *Let $\mu = 1/\mathbb{E}U$ and define*

$$I_\pm(x,\epsilon) \;=\; \int_0^\infty \exp\left\{-\mu\overline{G}(t)x(1\pm\epsilon)\right\} f(t)\,\mathrm{d}t$$

*Then for each $\epsilon > 0$,*

$$1-\epsilon \;\le\; \liminf_{x\to\infty}\frac{\overline{H}(x)}{I_+(x,\epsilon)} \;\le\; \limsup_{x\to\infty}\frac{\overline{H}(x)}{I_-(x,\epsilon)} \;\le\; 1+\epsilon\,.$$

This lemma essentially reduces the investigation of the asymptotics of $\overline{H}(x)$ to the (not always straightforward!) purely analytical study of the asymptotics of $I_+(x,\epsilon)$ and $I_-(x,\epsilon)$. Indeed, we will see in Section 5 that once this is done, one is most often able to obtain the logarithmic asymptotics of $\overline{H}(x)$ by letting $\epsilon \downarrow 0$, and in some cases even the sharp asymptotics. Finally, Section 6 contains some concluding remarks.

# 2 Statement of Main Results

Except for Proposition 3.2, we will assume throughout the paper that the support of $F$ is infinite.

We shall use the concept of *logarithmic asymptotics* familiar from large deviations theory and write $f(t) \approx_{\log} g(t)$ for two functions $f, g > 0$ with limits 0 at $t = \infty$ if $\log f(t)/\log g(t) \to 1$ as $t \to \infty$. We then consider the following distribution classes:

$$\mathscr{F}_1 : \ f(t) \approx_{\log} \mathrm{e}^{-\alpha t^{\eta}}, \qquad \mathscr{F}_2 : \ f(t) \approx_{\log} \frac{1}{t^{\alpha+1}},$$

$$\mathscr{G}_1 : \ \overline{G}(t) \approx_{\log} \mathrm{e}^{-\beta t^{\gamma}}, \qquad \mathscr{G}_2 : \ \overline{G}(t) \approx_{\log} \frac{1}{t^{\beta}}$$

Note that these definitions do not completely identify the tail behaviour of $F, G$. For example, if $\overline{G}(t) \sim ct^{\alpha}\mathrm{e}^{-\beta t^{\gamma}}$, then $G \in \mathscr{G}_1$, but one cannot identify $c, \alpha$, and if $f(t) \sim c\log^{\beta} t/t^{\alpha+1}$, then $F \in \mathscr{F}_1$, but one cannot identify $c, \beta$.

Note also that $f \in \mathscr{F}_2$ implies that $\overline{F}(t) \approx_{\log} 1/t^{\alpha}$, and that a sufficient (but not necessary) condition for $G \in \mathscr{G}_2$ is that $g(t)$ is regularly varying with index $-\beta - 1$.

Similarly to the definition of $f \approx_{\log} g$, we will write $f \approx_{\log\log} g$ if

$$\frac{\log\bigl(-\log f(t)\bigr)}{\log\bigl(-\log g(t)\bigr)} \ \to \ 1\,.$$

See further part (1:2) of Theorem 2.1 and Remark E) in Section 6.

With these distribution classes, we obtain a complete description of the logarithmic asymptotics of $\overline{H}(x)$ except for the case $f \in \mathscr{F}_1$, $G \in \mathscr{G}_2$ where we only obtain $\approx_{\log\log}$ asymptotics.:

**Theorem 2.1**
(1:1) *Assume* $F \in \mathscr{F}_1, G \in \mathscr{G}_1$. *Then* $\overline{H}(x) \approx_{\log} \exp\bigl\{-c_{11}\log^{\theta_{11}} x\bigr\}$ *where* $\theta_{11} = \eta/\gamma$, $c_{11} = \alpha/\beta^{\theta_{11}}$*;*
(2:2) *Assume* $F \in \mathscr{F}_2, G \in \mathscr{G}_2$. *Then* $\overline{H}(x) \approx_{\log} \dfrac{1}{x^{\theta_{22}}} = \exp\{-\theta_{22}\log x\}$ *where* $\theta_{22} = \alpha/\beta$*;*
(2:1) *Assume* $F \in \mathscr{F}_2, G \in \mathscr{G}_1$. *Then* $\overline{H}(x) \approx_{\log} \dfrac{1}{\log^{\theta_{21}} x} = \exp\{-\theta_{21}\log\log x\}$ *where* $\theta_{21} = \alpha/\gamma$*;*
(1:2) *Assume* $F \in \mathscr{F}_1, G \in \mathscr{G}_2$. *Then* $\overline{H}(x) \approx_{\log\log} \exp\bigl\{-x^{\theta_{12}}\bigr\}$ *where* $\theta_{12} = \eta/(\beta+\eta) \in (0,1)$.

Note that the asymptotic expressions are in agreement with $\overline{H}(x)$ being necessarily heavy-tailed, cf. Proposition 1.2. E.g. the asymptotics in part (1:2) is as for the heavy-tailed Weibull distribution, and the one in part (1:1) as for regular variation if $\theta_{11} = 1$ and as for the lognormal distribution if $\theta_{11} = 2$.

Generalizing [18], we will also show:

**Proposition 2.1** *Assume* $g(t) \geq cf(t)^{1/\alpha-\epsilon}$ *for all large* $t$*, where* $\alpha, \epsilon > 0$*,* $c < \infty$*. Then* $\int_0^{\infty} x^{\alpha} H(\mathrm{d}x) \ < \ \infty$*. If* $g(t) \leq cf(t)^{1/\alpha}$ *for all large* $t$*, where* $\alpha > 0$*,* $c > 0$*, then* $\int_0^{\infty} x^{\alpha} H(\mathrm{d}x) \ = \ \infty$*.*

For example, the mean of $H$ is finite when the tail of $F$ is slightly lighter than the tail of $G$ and infinite when it is equal or or heavier. Similar, checking finite variance amounts to a comparison of $\overline{F}$ and $\overline{G}^2$.

Our main results on sharp asymptotics is as follows (here and in the following, slowly varying functions are assumed to have the additional property of being bounded on compact subsets of $(0, \infty)$):

**Theorem 2.2** *Assume*

$$f(t) \;=\; g(t)\overline{G}(t)^{\beta-1}L_0\big(\overline{G}(t)\big) \tag{2}$$

*where $L_0(s)$ is slowly varying at $s=0$. Then*

$$\overline{H}(x) \;\sim\; \frac{\Gamma(\beta)}{\mu^\beta}\frac{L_0(1/x)}{x^\beta}, \quad x\to\infty\,. \tag{3}$$

Here $f(x)\sim g(x)$ means $f(x)/g(x)\to 1$. For example:

**Corollary 2.1** *Assume $f,g$ belong to the class of regularly varying densities of the form $L(t)/t^{1+\alpha}$ where $L$ is slowly varying, with parameters $\alpha_F, L_F$ for $f$ and $\alpha_G, L_G$ for $g$. Then $\overline{H}(x) \;=\; L_H(x)/x^{\alpha_H}$, where $\alpha_H=\alpha_F/\alpha_G$ and $L_H$ is slowly varying with*

$$L_H(x) \;\sim\; \frac{\Gamma(\alpha_H)\alpha_G^{\alpha_H-1}}{\mu^{\alpha_H}}\frac{L_F\big(x^{1/\alpha_G}\big)}{L_G^{\alpha_H}\big(x^{1/\alpha_G}\big)}\,.$$

**Corollary 2.2** *Assume $f,g$ belong to the class of densities of the form $\mathrm{e}^{-\lambda t^\eta}t^\alpha L(t)$ where $L$ is slowly varying at $t=\infty$, with parameters $\lambda_F,\alpha_F,L_F$ for $f$ and $\lambda_G,\alpha_G,L_G$ for $g$, and the same $\eta=\eta_F=\eta_G$. Then $\overline{H}(x) \;=\; L_H(x)/x^{\alpha_H}$, where $\alpha_H=\lambda_F/\lambda_G$ and $L_H$ is slowly varying with*

$$L_H(x) \;\sim\; \frac{\Gamma(\alpha_H)\lambda_G^{\alpha_H-1-\omega}\eta^{\alpha_H-1}}{\mu^{\alpha_H}\lambda_G^{\alpha_F/\eta-\alpha_G\alpha_H/\eta+\alpha_H-1}}\log^\omega x\;\frac{L_F\big(\log^{1/\eta}x\big)}{L_G^{\alpha_H}\big(\log^{1/\eta}x\big)}\,,$$

*where $\omega \;=\; \alpha_F/\eta+\alpha_H(\eta-\alpha_G-1)/\eta+1/\eta-1$.*

Of course Corollary 2.1 is close in spirit to Theorem 2.1(2:2); the conditions are slightly stronger, but so are also the conclusions. The difference between Corollary 2.2 and Theorem 2.1(1:1) is somewhat more marked, since Corollary 2.2 only applies when $\eta_F=\eta_G$ (i.e., $\eta=\gamma$ in the notation of Theorem 2.1, where $\eta=\gamma$ is not required).

Finally consider ordering and comparison results. One expects intuitively a heavier tail of $F$ to lead to a heavier $\overline{H}$. The precise statement of this is in terms of stochastic order (s.o.):

**Proposition 2.2** *Assume given two task time distributions $F_1$, $F_2$ such that $F_1$ is smaller than $F_2$ in s.o., that is, $\overline{F}_1(t)\le\overline{F}_2(t)$ for all $t$. Then also $H_{F_1,G}\le H_{F_2,G}$ in s.o. for any fixed $G$.*

This follows from (1) and the coupling characterization of s.o. ([17]) by noting that if $T_1\le T_2$, then (in obvious notation) $N(T_1)\le N(T_2)$ and hence $S(T_1)\le S(T_2)$, $X(T_1)\le X(T_2)$

Similarly, one expects a lighter tail of $G$ to lead to a larger $X$. However, stochastic ordering cannot be inferred since if $G_1$, $G_2$ are given ($F$ is fixed) such that $G_1$ is smaller than $G_2$ in s.o, then on one hand $N$ is smaller for $G_2$ than for $G_1$ for any $t$ but on the other the $U_i(t)$ are larger. However, we will establish an asymptotic order under a slightly stronger condition than $G_1$ being smaller than $G_2$ in s.o.:

**Proposition 2.3** *Assume that $G_1$ is smaller than $G_2$ in s.o. and that in addition $\limsup_{t\to\infty} \overline{G}_1(t)/\overline{G}_2(t) < 1$. Then for $F$ fixed,*

$$\limsup_{t\to\infty} \frac{\overline{H}_{F,G_2}(t)}{\overline{H}_{F,G_1}(t)} \;\le\; 1\,.$$

# 3 Geometric Sums. Bounded Job Time $T$

Given $T = t$, the number $N(t)$ of restarts is geometric with failure parameter $G(t) = \mathbb{P}(U_i \le t) = 1 - \overline{G}(t)$ so that

$$\mathbb{P}\bigl(N(t) > n\bigr) \;=\; G(t)^n, \;\; \mathbb{E}N(t) \;=\; \frac{G(t)}{\overline{G}(t)} \;\sim\; \frac{1}{\overline{G}(t)}\,.$$

It follows that given $T = t$, we can write

$$X \;\stackrel{\mathcal{D}}{=}\; t + S(t) \;\;\text{where}\;\; S(t) \;=\; \sum_{i=1}^{N(t)} U_i(t)$$

(here $\stackrel{\mathcal{D}}{=}$ means equality in distribution) where the $U_i(t)$ are independent of $N(t)$ and i.i.d. with the distribution $G_t$ being $G$ truncated to $[0,t)$, that is, with density $G(t)^{-1}g(s)I(s \le t)$ at $s$. Then $X \stackrel{\mathcal{D}}{=} T + S(T)$ so that

$$\overline{H}(x) \;=\; \int_0^\infty \mathbb{P}\bigl(S(t) > x - t\bigr)\, f(t)\,\mathrm{d}t\,. \tag{4}$$

This is the basic identity to be used in the following.

A first implication of (4) is that asymptotic properties of geometric sums must play a role for the asymtotics of $\overline{H}(x)$. We shall use Cramér-Lundberg theory, cf. [1], [2], [19], more precisely the following result:

**Proposition 3.1** *Let $V_1, V_2, \ldots$ be i.i.d. with common density $k(v)$, $N \in \mathbb{N}$ an independent r.v. with $\mathbb{P}(N = n) = (1-\rho)\rho^n$, and $S = V_1 + \cdots + V_N$. Then $\mathbb{P}(S > x) \sim C\mathrm{e}^{-\gamma x}$ where $\gamma$ is the solution of $\rho\int_0^\infty \mathrm{e}^{\gamma y}k(y)\,\mathrm{d}y \;=\; 1$ and $C = (1-\rho)/\gamma B$ where $B \;=\; \rho\int_0^\infty y\mathrm{e}^{\gamma y}k(y)\,\mathrm{d}y$. Furthermore, letting*

$$c_-(x) \;=\; \inf_{0\le z\le x, \overline{K}(z)>0} \frac{e^{\gamma x}\overline{K}(x)}{\int_x^\infty \mathrm{e}^{\gamma y}k(y)\,\mathrm{d}y}, \quad c_+(x) \;=\; \sup_{0\le z\le x, \overline{K}(z)>0} \frac{e^{\gamma x}\overline{K}(x)}{\int_x^\infty \mathrm{e}^{\gamma y}k(y)\,\mathrm{d}y},$$

*we have the Lundberg inequality*

$$c_-(x)\mathrm{e}^{-\gamma x} \;\le\; \mathbb{P}(S > x) \;\le\; c_+(x)\mathrm{e}^{-\gamma x}$$

*for all $x$.*

For a proof, see Willmot & Lin [19] pp. 108-109. Alternatively, Proposition 3.1 follows easily from

$$\begin{aligned}\mathbb{P}(S > x) &= \mathbb{P}(N \ge 1, V_1 > x) \;+\; \int_0^x \mathbb{P}(S > x - y)\mathbb{P}(N \ge 1, V_1 \in \mathrm{d}y)\\ &= \rho\overline{K}(x) \;+\; \int_0^x \mathbb{P}(S > x - y)\rho k(y)\,\mathrm{d}y\,,\end{aligned}$$

which is a defective renewal equation to which standard theory applies (see [2] V.7 and also [1] III.6c).

**Corollary 3.1** *In the* RESTART *setting,* $\mathbb{P}\big(S(t) > x\big) \sim C(t)\mathrm{e}^{-\gamma(t)x}$*,* $x \to \infty$*, where* $\gamma(t) > 0$ *is the solution of* $\int_0^t \mathrm{e}^{\gamma(t)y} G(\mathrm{d}y) = 1$ *and* $C(t) = \overline{G}(t)/\gamma(t)B(t)$ *where* $B(t) = \int_0^t y\mathrm{e}^{\gamma(t)y} g(y)\,\mathrm{d}y$*. This estimate is uniform in* $t_1 \le t \le t_2$ *for given* $0 < t_1 < t_2$*. Furthermore,*

$$\mathrm{e}^{-\gamma(t)t}\mathrm{e}^{-\gamma(t)x} \ \le\ \mathbb{P}\big(S(t) > x\big) \ \le\ \mathrm{e}^{-\gamma(t)t}\,.$$

*Proof.* The first statement is a trivial translation of the first statement of Proposition 3.1. For the two-sided Lundberg inequality, note that in the RESTART setting with $K(y) = K_t(y) = \mathbb{P}\big(U(t) \le y\big)$, the integral in the definition of $c_-(x)$ extends only up to $t$ which gives $c_-(x) \ge \mathrm{e}^{-\gamma t}$, and that $c_+(x) \le 1$. For the uniformity of the Cramér-Lundberg approximation, appeal to uniform estimates of the renewal functions corresponding to the $\mathrm{e}^{\gamma(t)y}K_t(\mathrm{d}y)$ as given, e.g., Kartashov [12], [13] (see also Wang & Woodroofe [20]). □

In particular, Corollary 3.1 settles the case of a fixed job size:

**Corollary 3.2** *Assume* $T \equiv t_0$ *and* $\overline{G}(t_0) > 0$*. Then*

$$\overline{H}(x) \ \sim\ C(t_0)\mathrm{e}^{\gamma(t_0)t_0}\mathrm{e}^{-\gamma(t_0)x}\,.$$

In the case of an infinite support of $f$, Corollary 3.2 shows that the tail of $H$ is heavier than $\mathrm{e}^{-\gamma(t)x}$ for all $t$ (note that $\gamma(t) \downarrow 0$ as $t \to \infty$; more precise estimates are given later). This observation proves Proposition 1.2.

If $T$ is random, we need to mix over $t$ with weights $f(t)$. If the support of $f$ has a finite upper endpoint $t_0$, Corollary 3.2 suggests that the asymptotics of $\overline{H}(x)$ is not too far from $\mathrm{e}^{-\gamma(t_0)x}$, and in fact, we shall show:

**Proposition 3.2** *Assume that the support of* $F$ *has upper endpoint* $0 < t_0 < \infty$*, that* $\overline{G}(t_0) > 0$ *and that*

$$f(t) \ \sim\ A(t_0 - t)^\alpha, \quad t \uparrow t_0, \tag{5}$$

*for some* $0 < A < \infty$ *and some* $\alpha \ge 0$*. Then*

$$\overline{H}(x) \ \sim\ \frac{AB(t_0)^\alpha \overline{G}(t_0)\Gamma(\alpha+1)}{\gamma(t_0)\mathrm{e}^{\alpha\gamma(t_0)}g(t_0)^{\alpha+1}}\,\frac{\mathrm{e}^{-\gamma(t_0)x}}{x^{\alpha+1}}\,.$$

*Proof.* For simplicity of notation, write $B = B(t_0)$, $\gamma = \gamma(t_0)$ etc.

It is easy to see that $\gamma(t)$ is continuous and differentiable in $t$. To obtain the asymptotics as $t \uparrow t_0$ we write

$$\begin{aligned}
1 = \int_0^t \mathrm{e}^{\gamma(t)y}G(\mathrm{d}y) \ &= \int_0^{t_0} \mathrm{e}^{\gamma(t)y}G(\mathrm{d}y) \ - \int_t^{t_0} \mathrm{e}^{\gamma(t)y}G(\mathrm{d}y) \\
&= \int_0^{t_0} \mathrm{e}^{\gamma y}\big[1 + (\gamma(t) - \gamma)y\big]\,G(\mathrm{d}y) \ - \ (t_0 - t)\mathrm{e}^{\gamma t_0}g(t_0) \ + \ \mathrm{o}\big(\gamma(t) - \gamma\big) \\
&= 1 \ + \ (\gamma(t) - \gamma)B \ - \ (t_0 - t)\mathrm{e}^{\gamma t_0}g(t_0) \ + \ \mathrm{o}\big(\gamma(t) - \gamma\big)
\end{aligned}$$

so that

$$\gamma(t) - \gamma \;\sim\; (t_0 - t)D \tag{6}$$

where $D = \mathrm{e}^{\gamma t_0} g(t_0)/B$. Appealing to the uniformity in Corollary 3.1, we therefore get

$$\begin{aligned}
\overline{H}(x) &= \int_0^{t_0} Z_t(x-t)f(t)\,\mathrm{d}t \;=\; \int_{t_0-\epsilon}^{t_0} Z_t(x-t)f(t)\,\mathrm{d}t \;+\; \mathrm{o}(\mathrm{e}^{-\gamma x}) \\
&= r_1(\epsilon)\int_{t_0-\epsilon}^{t_0} C\mathrm{e}^{-\gamma(t)(x-t)}A(t_0-t)^{\alpha}\,\mathrm{d}t \;+\; \mathrm{o}(\mathrm{e}^{-\gamma x}) \\
&= r_2(\epsilon)AC\mathrm{e}^{-\gamma(x-t_0)}\int_{t_0-\epsilon}^{t_0} \mathrm{e}^{-(\gamma(t)-\gamma)x}(t_0-t)^{\alpha}\,\mathrm{d}t \;+\; \mathrm{o}(\mathrm{e}^{-\gamma x})
\end{aligned}$$

where $Z_t(x) = P(S(t) > x)$ and $r_1(\epsilon), r_2(\epsilon), \ldots \to 1$ as $\epsilon \downarrow 0$. Thus substituting $y = (\gamma(t) - \gamma)x$ and noting that $\mathrm{d}y \sim -D\mathrm{d}t$ by (6), we get up to the $\mathrm{o}(\mathrm{e}^{-\gamma x})$ term that

$$\overline{H}(x) \;=\; r_2(\epsilon)ACD^{-\alpha-1}\frac{\mathrm{e}^{-\gamma(x-t_0)}}{x^{\alpha+1}}\int_0^{(\gamma(t_0-\epsilon)-\gamma)x} y^{\alpha}\mathrm{e}^{-y}\,\mathrm{d}y$$

Letting first $x \to \infty$, next $\epsilon \downarrow 0$, and rewriting the constants completes the proof. $\square$

# 4 Proof of Lemma 1.1

We will need the asymptotics of the Cramér root $\gamma(t)$:

**Lemma 4.1** *As $t \to \infty$, $\gamma(t) \;\sim\; \mu\overline{G}(t)$.*

*Proof.* Consider

$$\int_0^t \big(\mathrm{e}^{\gamma(t)y} - 1 - \gamma(t)y\big)\,G(\mathrm{d}y) \;=\; \overline{G}(t) - \gamma(t)\big(1/\mu - \mathrm{o}(1)\big)\,. \tag{7}$$

The non-negativity of the l.h.s. yields $\gamma(t) \;=\; \mathrm{O}\big(\overline{G}(t)\big)$. Since $t\overline{G}(t)$ because of $\mu > 0$, the integrand in (7) can therefore be writtes as $\gamma(t)y\epsilon(y,t)$ where $\epsilon(y,t) \to 0$ uniformly in $y \le t$ as $t \to \infty$. Therefore (7) equals $\gamma(t)\mathrm{o}(1)$ which shows the assertion. $\square$

*Proof of Proposition* 1.2. Given $\epsilon > 0$, choose $t_0$ such that $\gamma(t_0) < \epsilon$, cf. Lemma 4.1, and $a$ so large that $\gamma(t_0 + a) < \gamma(t_0)$. We then get

$$\begin{aligned}
\liminf_{x\to\infty} \mathrm{e}^{\epsilon x}\overline{H}(x) \;&\ge\; \liminf_{x\to\infty} \mathrm{e}^{\gamma(t_0)x}\overline{H}(x) \;\ge\; \liminf_{x\to\infty}\int_{t_0+a}^{\infty} \frac{\mathbb{P}\big(S(t) > x\big)}{\mathrm{e}^{-\gamma(t_0)x}} \\
&\ge \int_{t_0+a}^{\infty} \liminf_{x\to\infty} \frac{\mathbb{P}\big(S(t) > x\big)}{\mathrm{e}^{-\gamma(t_0)x}} f(t)\,\mathrm{d}t \;=\; \int_{t_0+a}^{\infty} \infty \cdot f(t)\,\mathrm{d}t \;=\; \infty
\end{aligned}$$

where we used Fatou's lemma in the third step and Corollary 3.1 in the next. $\square$

**Lemma 4.2** *For any $t_0 < \infty$, $\mathbb{P}\big(X > x,\, T \le t_0\big)$ goes to zero at least exponentially fast.*

*Proof.* By Lundberg's inequality,

$$\mathbb{P}\big(X > x,\, T \le t_0\big) \;\le\; F(t_0)\mathbb{P}\big(S(t_0) > x - t_0\big) \;\le\; F(t_0)\mathrm{e}^{-\gamma(t_0)(x-t_0)}\,. \qquad \square$$

**Lemma 4.3** *Define $S_n(t) = U_1(t) + \cdots + U_n(t)$, $m(t) = \mathbb{E}U_i(t) = \mathbb{E}[U_i \,|\, U_i \le t]$. Then*

$$\mathbb{P}\Big(\big|S_n(t)/n - m(t)\big| > \epsilon\Big) \;=\; \mathrm{o}(1), \quad n \to \infty,$$

*where the* $\mathrm{o}(1)$ *is uniform in $t > \delta$ for any $\delta > 0$.*

*Proof.* Define $U_i(t,n) = U_i(t)I\big(U_i(t) < n\big)$. Then, in obvious notation

$$\mathbb{P}\big(S_n(t,n) \ne S_n(t)\big) \;\le\; n\mathbb{P}\big(U_n(t,n) \ne U_n(t)\big) \;\le\; \frac{n}{G(t)}\mathbb{P}(U > n)$$

goes to zero uniformly in $t > \delta$ because of $\mathbb{E}U < \infty$. Further,

$$\frac{1}{n}\mathbb{E}U_i(t,n)^2 \;=\; \int_0^n \frac{2x}{n}\mathbb{P}\big(U_i(t,n) > x\big)\,\mathrm{d}x \le \frac{2}{G(t)}\int_0^n \frac{x}{n}\mathbb{P}\big(U > x\big)\,\mathrm{d}x \;=\; \mathrm{o}(1)$$

uniformly in $t > \delta$, as follows by dominated convergence with $\mathbb{P}\big(U > x\big)$ as majorant. Hence by Chebycheff's inequality,

$$\mathbb{P}\Big(\big|S_n(t,n)/n - m(t,n)\big| > \epsilon\Big) \;\le\; \frac{n\mathbb{E}U_i(t,n)^2}{n^2\epsilon^2} \;=\; \mathrm{o}(1)$$

uniformly in $t > \delta$. Also

$$m(t) - m(t,n) \;=\; \mathbb{E}U_i(t)I\big(U_i(t) \ge n\big) \;\le\; \frac{1}{G(t)}\mathbb{E}UI\big(U \ge n\big) \;=\; \mathrm{o}(1)$$

uniformly in $t > \delta$. Putting these estimates together completes the proof. $\square$

*Proof of Lemma* 1.1. Given $\epsilon > 0$, it follows by Lemma 4.1 that we can choose $t_0$ such that $\gamma(t) \ge \mu\overline{G}(t)(1-\epsilon)$ and (since $G$ has finite mean) $\gamma(t)t < \log(1+\epsilon)$ for $t \ge t_0$ . Thus by the upper Lundberg bound and Lemma 4.2,

$$\begin{aligned}
\overline{H}(x) &= \int_{t_0}^{\infty} \mathbb{P}\big(S(t) > x - t\big)f(t)\,\mathrm{d}t \;+\; \mathrm{o}(\mathrm{e}^{-rx}) \\
&\le \int_{t_0}^{\infty} \mathrm{e}^{-\gamma(t)(x-t)}f(t)\,\mathrm{d}t \;+\; \mathrm{o}(\mathrm{e}^{-rx}) \\
&\le (1+\epsilon)\int_{t_0}^{\infty} \mathrm{e}^{-\mu\overline{G}(t)(1-\epsilon)x}f(t)\,\mathrm{d}t \;+\; \mathrm{o}(\mathrm{e}^{-rx}) \\
&\le (1+\epsilon)I_{-}(x,\epsilon) \;+\; \mathrm{o}(\mathrm{e}^{-rx})
\end{aligned}$$

for some $r > 0$. Now note that $\overline{H}(x)$ decays slower than $\mathrm{e}^{-rx}$ by Proposition 1.2.

For the lower bound, let $\epsilon > 0$ be given and let $\epsilon_1, \epsilon_2 > 0$ satisfy $(1+\epsilon)(1-\epsilon_1) > 1+\epsilon_2 > 1$. Lemma 4.3 implies that there is an $n_0$ such that

$$\mathbb{P}\big(S_n(t) > nm(t)(1-2\epsilon_1)\big) \;\geq\; 1-\epsilon$$

for all $n \geq n_0$ and all $t \geq t_0$. Since $m(t) \to 1/\mu$, we have then also

$$\mathbb{P}\big(S_n(t) > n(1-\epsilon_1)/\mu\big) \;\geq\; 1-\epsilon$$

for all $n \geq n_0$ and all $t \geq t_0$. Choose next $\overline{g}_0$ such that $\mathrm{e}^{-(1+\epsilon_2)\overline{g}} \leq 1-\overline{g}$ for $0 < \overline{g} < \overline{g}_0$. Replacing $t_0$ by a larger $t_0$ if necessary, we may assume $\overline{G}(t) < \overline{g}_0$ for $t \geq t_0$ and get

$$\begin{aligned}
\overline{H}(x) &\geq \int_{t_0}^{\infty} \mathbb{P}\big(S(t) > x\big) f(t)\,\mathrm{d}t \\
&\geq (1-\epsilon)\int_{t_0}^{\infty} \mathbb{P}\big(N(t) > x\mu/(1-\epsilon_1)\big) f(t)\,\mathrm{d}t \\
&= (1-\epsilon)\int_{t_0}^{\infty} G(t)^{x\mu/(1-\epsilon_1)} f(t)\,\mathrm{d}t \\
&\geq (1-\epsilon)\int_{t_0}^{\infty} \exp\left\{-\overline{G}(t)x\mu(1+\epsilon_2)/(1-\epsilon_1)\right\} f(t)\,\mathrm{d}t \\
&\geq (1-\epsilon)\int_{t_0}^{\infty} \exp\big\{-\overline{G}(t)x\mu(1+\epsilon)\big\} f(t)\,\mathrm{d}t\,.
\end{aligned}$$

Since the last integral differs from $I_+(x,\epsilon)$ by a term which goes to zero exponentially fast and hence is $\mathrm{o}(\overline{H}(x))$, the proof is complete. □

*Proof of Proposition* 1.1. When $F = G$, we have $\mathrm{d}\overline{G}(t) = -f(t)$. Hence

$$\begin{aligned}
I_{\pm}(x,\epsilon) &= \left[\frac{1}{\mu x(1\pm\epsilon)}\mathrm{e}^{-\mu\overline{G}(t)x(1\pm\epsilon)}\right]_{t_0}^{\infty} \;=\; \frac{1}{\mu x(1\pm\epsilon)}\big(1-\mathrm{e}^{-\mu\overline{G}(t_0)x(1\pm\epsilon)}\big) \\
&\sim \frac{1}{\mu x(1\pm\epsilon)}\,.
\end{aligned}$$

The assertion now follows easily from Lemma 1.1 by letting first $x \to \infty$ and next $\epsilon \to 0$. □

*Proof of Proposition* 2.1. Under the assumptions of the last part of the Proposition, $\overline{G}(t) \leq c_1\overline{F}(t)^{1/\alpha}$ for $t \geq t_0$ and hence

$$\begin{aligned}
\int_0^{\infty} x^{\alpha} H(\mathrm{d}x) &= \alpha\int_0^{\infty} x^{\alpha-1}\overline{H}(x)\,\mathrm{d}x \\
&\geq c_2\int_0^{\infty} x^{\alpha-1}\mathrm{d}x\int_0^{\infty} \mathrm{e}^{-\mu\overline{G}(t)x} f(t)\,\mathrm{d}t \\
&\geq c_3\int_{t_0}^{\infty} \frac{1}{\overline{G}(t)^{\alpha}} f(t)\,\mathrm{d}t \;\leq\; c_4\int_{t_0}^{\infty} \frac{1}{\overline{F}(t)} f(t)\,\mathrm{d}t \\
&= c_4\int_0^1 \frac{1}{y}\mathrm{d}y \;=\; \infty\,,
\end{aligned}$$

proving the last part of the Proposition. For the first part, we get similarly

$$\int_0^\infty x^\alpha H(\mathrm{d}x) \;\le\; c_5 + c_6 \int_0^1 \frac{1}{y^{1-\alpha\epsilon}}\mathrm{d}y \;<\; \infty\,. \qquad \square$$

*Proof of Proposition* 2.3. Write $I^1_\pm(x,\epsilon)$, $\mu_1$ when $G = G_1$ and similarly for $G_2$. We may assume $G_1 \neq G_2$. Then the s.o. assumption implies $\mu_1 < \mu_2$. Hence if $\epsilon > 0$ is so small that $\mu_1(1+\epsilon) < \mu_2(1-\epsilon)$, we have $\mu_1\overline{G}_1(t)(1+\epsilon) \le \mu_1\overline{G}_2(t)(1-\epsilon)$ for all $t$. We then obtain

$$\overline{H}_{G_1}(x) \;\ge\; (1-\epsilon)I^1_+(x,\epsilon) \;\ge\; (1-\epsilon)I^2_-(x,\epsilon) \;\ge\; \frac{1-\epsilon}{1+\epsilon}\overline{H}_{G_2}(x)\,,$$

where the outer inequalities are asymptotic and the inner one exact. Let first $x \to \infty$ and next $\epsilon \to 0$. $\square$

# 5 Proofs: Integral Asymptotics

**Lemma 5.1** *For given constants $a, b, \gamma, \eta > 0$,*

$$I \;=\; \int_{t_0}^\infty \exp\left\{-\mathrm{e}^{-bt^\gamma}z - at^\eta\right\}\,\mathrm{d}t \;\approx_{\log}\; \mathrm{e}^{-ab^{-\eta/\gamma}\log^{\eta/\gamma} z}$$

*as $z \to \infty$.*

*Proof.* Let $c = ab^{-\eta/\gamma}$, $t_1 = (\log z/b)^{1/\gamma}$ and let $I_1, I_2$ be the contributions to $I$ from the intervals $(t_1,\infty)$, resp. $(t_0,t_1)$. In $I_1$, we bound the first term in the exponent below by 0 so

$$I_1 \;\le\; \int_{t_1}^\infty \mathrm{e}^{-at^\eta}\,\mathrm{d}t \;\approx_{\log}\; \mathrm{e}^{-at_1^\eta} \;=\; \mathrm{e}^{-c\log^{\eta/\gamma} z}\,.$$

In $I_2$, we substitute $y = \mathrm{e}^{-bt^\gamma}z$. Then

$$t \;=\; (\log z - \log y)^{1/\gamma}b^{-1/\gamma}, \quad \mathrm{d}t \;=\; -\frac{1}{\gamma y}(\log z - \log y)^{1/\gamma-1}b^{-1/\gamma}\mathrm{d}y$$

so that $I_2$ becomes

$$\frac{1}{\gamma b^{1/\gamma}}\int_1^{\mathrm{e}^{-t_0^\gamma}z} \frac{1}{y}(\log z - \log y)^{1/\gamma-1}\exp\left\{-y - c(\log z - \log y)^{\eta/\gamma}\right\}\,\mathrm{d}y$$

We split this integral into the contributions $I_3, I_4$ from the intervals $[1,2)$, $[2,\mathrm{e}^{-t_0^\gamma}z)$. Here

$$I_3 \sim \frac{1}{\gamma b^{1/\gamma}}\log^{1/\gamma-1} z\;\mathrm{e}^{-c\log^{\eta/\gamma} z}\int_1^2 \frac{1}{y}\mathrm{e}^{-y}\,\mathrm{d}y \;\approx_{\log}\; \mathrm{e}^{-c\log^{\eta/\gamma} z}\,.$$

For $I_4$, we write $q = \eta/\gamma$, $h(y) = -y/2 + c\log^q z - c(\log z - \log y)^q$. Then

$$h(y) \;=\; -y/2 + cq\int_1^y \frac{(\log z - \log y)^{q-1}}{y}\,\mathrm{d}y \;\le\; -y/2 + cq\log^{q-1} z\log y\,.$$

The r.h.s. is maximized for $y_z = 2cq\log^{q-1} z$, where

$$h(y_z) = \log^{q-1} z\mathrm{O}(\log\log z) = \mathrm{o}\big(\log^q z\big)\,.$$

Hence

$$\begin{aligned} I_4 &\le b^{-1/\gamma}\log^{1/\gamma-1} z\mathrm{e}^{-c\log^q z}\int_1^z \frac{1}{y}\mathrm{e}^{-y/2+h(y)}\,\mathrm{d}y \\ &\le \mathrm{e}^{-\big(c+\mathrm{o}(1)\big)\log^q z}\int_1^z \frac{1}{y}\mathrm{e}^{-y/2}\,\mathrm{d}y \ \approx_{\log}\ \mathrm{e}^{-c\log^q z}\,. \end{aligned}$$

Adding these estimates shows that $\mathrm{e}^{-ab^{-\eta/\gamma}\log^{\eta/\gamma} z}$ is an asymptotic upper bound in the logarithmic sense, and that it is also a lower one follows from the estimate for $I_3$. □

**Lemma 5.2** *For given constants $a, b$,*

$$\int_{t_0}^\infty \mathrm{e}^{-t^{-\beta}z}\frac{1}{t^{\alpha+1}}\,\mathrm{d}t \ \approx_{\log}\ \frac{1}{z^{\alpha/\beta}}\,.$$

*Proof.* Substitute $y = t^{-\beta}z$ to get

$$\int_{t_0}^\infty \mathrm{e}^{-t^{-\beta}z}\frac{1}{t^{\alpha+1}}\,\mathrm{d}t \ = \ \int_0^{t_0^{-\beta}z}\frac{y^{\alpha/\beta-1}}{\beta z^{\alpha/\beta}}\mathrm{e}^{-y}\,\mathrm{d}y \ \sim\ \frac{\Gamma\big(\alpha/\beta\big)}{\beta z^{\alpha/\beta}}) \ \approx_{\log}\ \frac{1}{z^{\alpha/\beta}}\,. \qquad □$$

**Lemma 5.3** *For given constants $a, b, \gamma > 0$,*

$$I = \int_{t_0}^\infty \exp\big\{-\mathrm{e}^{-bt^\gamma}z - (a+1)\log t\big\}\ \mathrm{d}t \approx_{\log} \frac{1}{\log^{a/\gamma} z}$$

*as $z \to \infty$.*

*Proof.* Let again $t_1 = t_1(z) = \big(\log z/b\big)^{1/\gamma}$ (then $\mathrm{e}^{-bt_1^\gamma}z = 1$) and let $I_1, I_2$ be the contributions to $I$ from the intervals $(t_1,\infty)$, resp. $(t_0, t_1)$. In $I_1$, $0 \le \mathrm{e}^{-bt^\gamma}z \le 1$ and so

$$\frac{\mathrm{e}^{-1}}{a\big(\log(z/b)\big)^{a/\gamma}} \le I_1 \le \frac{1}{a\big(\log(z/b)\big)^{a/\gamma}}\,.$$

For $I_2$, let $r_z(t) = \mathrm{e}^{-bt^\gamma}z + (a+1)\log t$, $s(t) = t^\gamma\mathrm{e}^{-bt^\gamma}$. Then

$$r_z'(t) = -b\gamma t^{\gamma-1}\mathrm{e}^{-bt^\gamma}z + \frac{a+1}{t} = \frac{1}{t}\big[-b\gamma s(t)z + (a+1)\big]\,.$$

Since $s$ is continuous with $s(t_0) > 0$ and $s(t)$ is monotonically decreasing for large $t$ with limit 0, we have $s(t) \ge s(t_1) = \log(z/b)/z$ for all $t_0 \le t \le t_1$ and all large $z$ because of $t_1(z) \to \infty$. Hence $r_z'(t) < 0$ for $t_0 \le t \le t_1$ and all large $z$ so that

$$\begin{aligned} I_2 = \int_{t_0}^{t_1}\mathrm{e}^{-r_z(t)}\,\mathrm{d}t \ &\le\ (t_1 - t_0)\mathrm{e}^{-r_z(t_1)} \\ &\le t_1\mathrm{e}^{-(a+1)\log t_1} \ = \ \frac{1}{t_1^a} \ = \ \frac{1}{\big(\log(z/b)\big)^{a/\gamma}}\,. \end{aligned}$$

Putting the upper bounds for $I_1, I_2$ together and noting that $\big(\log(z/b)\big)^{a/\gamma} \approx_{\log} \log^{a/\gamma} x$ shows that $\log^{-a/\gamma} z$ is an upper bound in the logarithmic sense, and that it is also a lower bound follows from the lower bound for $I_1$. □

**Lemma 5.4** *Let $\eta > 0$ be fixed. Then for any $a, b > 0$,*

$$I \;=\; \int_{t_0}^{\infty} \exp\left\{-t^{-b}z - at^{\eta}\right\} \mathrm{d}t \;\approx_{\log}\; \exp\left\{-c_{12}(a,b)z^{\theta_{12}(b)}\right\}$$

*as $z \to \infty$ where*

$$\theta_{12}(b) = \eta/(b+\eta), \quad c_{12}(a,b) \;=\; a^{1-\theta_{12}}\left[(\eta/b)^{1-\theta_{12}(b)} + (b/\eta)^{\theta_{12}}\right].$$

*Proof.* We choose $t_1 = t_1(z)$ to minimize $f(t) \;=\; t^{-b}z + at^{\eta}$ which gives

$$t_1 \;=\; \left(\frac{bz}{a\eta}\right)^{1/(b+\eta)}, \quad f(t_1) \;=\; c_{12}(a,b)z^{\theta_{12}(b)}.$$

Thus the claim of the lemma can be written as $I \;\approx_{\log}\; \mathrm{e}^{-f(t_1)}$. As lower bound, we use

$$\int_{t_1}^{t_1+1} \exp\left\{-t^{-b}z - at^{\eta}\right\} \mathrm{d}t \;\geq\; \exp\left\{-t_1^{-b}z - a(t_1+1)^{\eta}\right\} \;\approx_{\log}\; \mathrm{e}^{-f(t_1)}$$

where in the last step we used $(t+1)^{\eta} = t^{\eta}\bigl(1+\mathrm{o}(1)\bigr)$. For the upper bound, we write $I = I_1 + I_2 + I_3$ where $I_1, I_2, I_3$ are the contributions from the intervals $t_0 < t < t_1$, $t_1 < t < Kt_1$, resp. $Kt_1 < t < \infty$ where $K$ satisfies $aK^{\eta} > c_{12}(a,b)$. Since $f$ is decreasing in the interval $t_0 < t < t_1$ and increasing in $t_1 < t < \infty$, we have $I_1 \leq t_1\mathrm{e}^{-f(t_1)} \approx_{\log} f(t_1)$ and $I_2 \leq (K-1)t_1\mathrm{e}^{-f(t_1)} \approx_{\log} \mathrm{e}^{-f(t_1)}$. Finally,

$$I_3 \;\leq\; \int_{Kt_1}^{\infty} \mathrm{e}^{-at^{\eta}}\,\mathrm{d}t \;\approx_{\log}\; \mathrm{e}^{-aK^{\eta}t_1^{\eta}}$$

can be neglected because of the choice of $K$. □

*Proof of Theorem* 2.1. In (1:1), we can choose $t_0$ such that $\overline{G}(t) \leq \mathrm{e}^{-bt^{\gamma}}$ and $f(t) \geq \mathrm{e}^{-at^{\eta}}$, $t \geq t_0$, for any given $b < \beta$ and $a > \alpha$. With $I_{\pm}(x,\epsilon)$ as in Lemma 1.1, we then get

$$\begin{aligned}
&\liminf_{x\to\infty} \frac{\log I_+(x,\epsilon)}{-\log^{\eta/\gamma} x} \\
&\quad = \liminf_{x\to\infty} \frac{1}{-\log^{\eta/\gamma} x} \log \int_{t_0}^{\infty} \exp\left\{-\mu\overline{G}(t)x(1+\epsilon)\right\} f(t)\,\mathrm{d}t \\
&\quad \geq \liminf_{x\to\infty} \frac{1}{-\log^{\eta/\gamma} x} \log \int_{t_0}^{\infty} \exp\left\{-\mu\mathrm{e}^{-bt^{\gamma}}x(1+\epsilon) - at^{\eta}\right\} \mathrm{d}t \\
&\quad = \liminf_{x\to\infty} \frac{ab^{-\eta/\gamma}\log^{\eta/\gamma}\bigl(\mu x(1+\epsilon)\bigr)}{-\log^{\eta/\gamma} x} \;=\; ab^{-\eta/\gamma}
\end{aligned}$$

where we used Lemma 5.1 with $z = \mu x(1+\epsilon)$ in the third step. Letting $a \downarrow \alpha$, $b \uparrow \alpha$ shows that $\mathrm{e}^{-c_{11}\log^{\theta_{11}} x}$ is an asymptotic lower bound in the logarithmic sense. That it is also an asymptotic upper bound follows in the same way by noting that the contribution to $\overline{H}(x)$ from $(0,t_0)$ goes to zero exponentially fast by Proposition 1.2 for any $t_0$ and hence is negligible compared to $\mathrm{e}^{-c_{11}\log^{\theta_{11}} x}$.

Parts (2:2) and (2:1) follow in a similar way from Lemmas 5.2 and 5.3. For (1:2), we choose $b > \beta$, $a < \alpha$ and get

$$\begin{aligned}
\liminf_{x\to\infty} \frac{\log\bigl(-\log \overline{H}(x)\bigr)}{\log\bigl(-\log \mathrm{e}^{-x^{\theta_{12}}}\bigr)} &\geq \liminf_{x\to\infty} \frac{\log\bigl(-\log I_+(x,\epsilon)\bigr)}{\theta_{12}\log x} \\
&= \liminf_{x\to\infty} \frac{1}{\theta_{12}\log x}\Bigl(-\log\int_{t_0}^{\infty} \exp\bigl\{-\mu\overline{G}(t)x(1+\epsilon)\bigr\} f(t)\,\mathrm{d}t\Bigr) \\
&\geq \liminf_{x\to\infty} \frac{1}{\theta_{12}\log x}\Bigl(-\log\int_{t_0}^{\infty} \exp\bigl\{-\mu t^{-b}x(1+\epsilon) - at^{\eta}\bigr\}\,\mathrm{d}t\Bigr) \\
&= \liminf_{x\to\infty} \frac{\log\bigl(c_{12}(a,b)\bigl(\mu x(1+\epsilon)\bigr)^{\theta_{12}(b)}\bigr)}{\theta_{12}\log x} \;=\; \frac{\theta_{12}(b)}{\theta_{12}}\,.
\end{aligned}$$

Letting $a \uparrow \alpha$, $b \downarrow \alpha$ shows that $\mathrm{e}^{-x^{\theta_{12}}}$ is an asymptotic lower bound in the $\approx_{\log\log}$ sense. That it is also an asymptotic upper bound follows similarly. □

*Proof of Theorem* 2.2. In Lemma 1.1, we insert (2) and substitute $s = \overline{G}(t)$ to get

$$I_{\pm} \;=\; \int_0^{s_0} \exp\bigl\{-sx\mu(1\pm\epsilon)\bigr\}s^{\beta-1}L_0(s)\,\mathrm{d}s\,.$$

where $s_0 = \overline{G}^{-1}(t_0)$. Then by Karamata's Tauberian theorem ([3, Theorems 1.5.11 and 1.7.1]),

$$I_{\pm} \sim \Gamma(\beta)\frac{L\bigl(1/(x\mu(1\pm\epsilon)\bigr)}{x^{\beta}\mu^{\beta}(1\pm\epsilon)^{\beta}} \sim \Gamma(\beta)\frac{L(1/x)}{x^{\beta}\mu^{\beta}(1\pm\epsilon)^{\beta}}\,.$$

Let $\epsilon \downarrow 0$. □

*Proof of Corollary* 2.1. We have $\overline{G}(t) = L_G'(t)/t^{\alpha_G}$, where $L_G'(t) \sim L_G(t)/(\alpha_G+1)$ as $t \to \infty$. Then (2) holds with $\beta = \alpha_F/\alpha_G$ and $L_0\bigl(\overline{G}(t)\bigr) = L_F(t)/L_G(t){L_G'}^{\beta-1}(t)$. Note that $L_0$ is s.v. because the inverse of a s.v. function is again s.v. ([3, p. 28]) and because the composition of two s.v. functions is again s.v. Further, ([3, p. 29]) $L_F\bigl(\overline{G}^{-1}(s)\bigr) \sim L_F\bigl(s^{-1/\alpha_G}\bigr)$ as $s \downarrow 0$ and similarly for $L_G$. Thus, $L_0(s) \sim L_F\bigl(s^{-1/\alpha_G}\bigr)(\alpha_G+1)^{\beta-1}/L_G^{\beta}\bigl(s^{-1/\alpha_G}\bigr)$. □

*Proof of Corollary* 2.2. We have

$$\overline{G}(t) = \int_t^{\infty} \mathrm{e}^{-\lambda_G y^{\eta}} y^{\alpha_G} L_G(y)\,\mathrm{d}y \sim \frac{1}{\eta\lambda_G}\mathrm{e}^{-\lambda_G t^{\eta}} t^{\alpha_G+1-\eta_G} L_G(t)$$

(e.g., substitute $z = \mathrm{e}^{-\lambda_G y^{\eta}}$ and apply Karamata's theorem). From this it is easy to see that

$$\overline{G}^{-1}(s) \;\sim\; (-\log s)^{1/\eta}/\lambda_G^{1/\eta}\,, \quad s \downarrow 0.$$

In particular, $L_F\bigl(\overline{G}^{-1}(s)\bigr) \sim L_F\bigl((-\log s)^{1/\eta}\bigr)$ and similarly for $L_G$. Thus if $L_0$ is defined by (2) with $\beta = \lambda_F/\lambda_G$, we have

$$L_0(s) \;\sim\; \lambda_G^{\beta-1-\omega}\eta^{\beta-1}(-\log s)^{\omega}\frac{L_F\bigl((-\log s)^{1/\eta}\bigr)}{L_G^{\beta}\bigl((-\log s)^{1/\eta}\bigr)}$$

which in particular shows that $L_0$ is s.v. at $s = 0$. Now just replace $s$ by $1/x$ and $\beta$ by $\alpha_H$ to obtain the Corollary. □

Corollary 1.1 is a special case of Corollary 2.2.

# 6 Concluding Remarks

**A)** The representation (1) easily gives a proof of the asymptotics $\overline{H}(x) \sim 1/\mu x$ for the diagonal case $F = G$ (Proposition 1.1). Indeed, the event $N = n$ corresponds to the ordering $U_1 < T, \ldots, U_n < T$, $U_{n+1} > T$. Since the $n+2$ random variables $T, U_1, \ldots, U_n, U_{n+1}$ are i.i.d. when $F = G$, we therefore have $\mathbb{P}(N = n) = 1/(n+2) \cdot (n+1)$, $\mathbb{P}(N > n) = 1/(n+2)$ (we are grateful to Clive Anderson for a remark triggering this observation). One can now argue that in order for $X$ to be large, $N$ has to be large which in turn is only possible if $T$ is large. Then the distribution of the $U_i$ is close to $G$, so that the geometric sum is approximately $N/\mu$. Since $\mathbb{E}T = 1/\mu < \infty$ implies that the tail of $T$ is lighter than $1/x$, we therefore get

$$\mathbb{P}(X > x) \;\approx\; \mathbb{P}\Big(\sum_{i=1}^{N} U_i > x\Big) \;\approx\; \mathbb{P}(N > \mu x) \;\sim\; \frac{1}{\mu x}\,.$$

Th argument is not hard to make rigorous, but we omit the details since the further results of the paper are much more general than Proposition 1.1 and require different proofs.

**B)** An application of the above results occurs in parallel computing. Assume that a job of length $NT$ is split into $N$ subjobs of length $T$ which are placed on $N$ parallel processors ($N$ may run in the order of hundreds or thousands). If one processor fails, the corresponding subjob is restarted on a new processor. With $X_1, \ldots, X_N$ the total times of the subjobs, the total job time is then $M_N = \max(X_1, \ldots, X_N)$. The asymptotic behaviour of $M_N$ as $N \to \infty$ is available from extreme value theory once the tails of the $X_n$ is known, which is precisely what has been the objective of this paper.

Whether this asymptotic scheme is the most relevant one is, however, questionable. One could equally well assume the job length fixed at $T$ and the length of the $N$ subjobs to be $T/N$, and intermediate possibilities. This leads into specific questions on extreme value theory in a triangular array setting, which are currently under investigation.

**C)** An alternative to the Cramér-Lundberg theory for geometric sums that has been one of our main tools is what could suitably be called Renyi theory, cf. [11]. One considers there a weak convergence triangular setting where still $x \to \infty$ but the parameters of the geometric sum depend on $x$; this is also related to the heavy-traffic or diffusion limit setting of risk and queueing theory, cf. e.g. [1] and [2] X.7. Renyi theory (e.g. [11]) provides the following alternative to Proposition 3.1:

**Proposition 6.1** *For each $x$, let $V_1(z), V_2(z), \ldots$ be i.i.d. with common density $k(v; z)$ and $N(z) \in \mathbb{N}$ an independent r.v. with $\mathbb{P}(N(z) = n) = (1 - \rho(z))\rho(z)^n$, $S(z) = V_1(z) + \cdots + V_N(z)$. If $\rho(z) \to 1$ as $z \to \infty$ and $V_k(z) \xrightarrow{\mathcal{D}} V$, $\mathbb{E}V_k(z) \to 1/\mu$ for some r.v. $V$ with finite mean $\mu^{-1}$, then $\mu(1 - \rho(z)S(z)$ has a limiting standard exponential distribution.*

**Corollary 6.1** *In the RESTART setting, $\mu\overline{G}(t)S(t)$ has a limiting standard exponential distribution as $t \to \infty$.*

The implication is that

$$\mathbb{P}\big(\mu\overline{G}(t)S(t) > y\big) \;\to\; \mathrm{e}^{-y} \tag{8}$$

for any fixed $y$. Noting that $\mu\overline{G}(t) \sim \gamma(t)$ and replacing $y$ by $\gamma(t)x$, this suggests $\mathbb{P}(S(t) > x) \approx \mathrm{e}^{-\gamma(t)x}$, i.e. the Cramér-Lundberg approximation. Of course, the derivation is not rigorous since (8) requires that $y$ is fixed. Nevertheless, it is indeed possible to derive some of our results from (8). The main reason that we have chosen Cramér-Lundberg asymptotics as our basic vehicle is that simple bounds are available (Lundberg's inequality) which is not the case for Renyi theory.

**D)** Since $H$ is a mixture of the $H_t$ given by $H_t(x) = \mathbb{P}\big(S(t) + t \le x\big)$ and the tail of $H_t$ obeys the Cramér-Lundberg asymptotics, we are dealing with the problem of determining the tail of a mixture where the tails of the mixing components are known. Looking for literature on this problem, we found a set of papers emerging from reliability and survival analysis (Finkelstein & Esaulova [9] and references there) which suggest our logarithmic asymptotics results but do not prove them because the assumptions are too stringent to apply to our setting.

**E)** The $\approx_{\log\log}$ asymptotics in part (1:2) identifies $\theta_{12}$ as the correct exponent to $x$ in $\log\overline{H}(x)$, but does not allow sharpenings like $\overline{H}(x) \approx_{\log} \mathrm{e}^{-c_{12}x^{\theta_{12}}}$, $\overline{H}(x) \approx_{\log} \mathrm{e}^{-c_{12}\log^{q_{12}}x\, x^{\theta_{12}}}$ etc. Inspection of the proof shows that to obtain such strengthenings, one needs first of all to be able to replace the $b$ in Lemma 5.4 with a fixed value $\beta$ rather than considering $b$'s arbitrarily close to $\beta$. This would be the case if, e.g., one assumed $\overline{G}$ to be regularly varying with index $-\beta$ rather than just $\overline{G}(t) \approx_{\log} t^{-\beta}$. This does not appear to be all that restrictive, but does not suffice since one also needs to replace the $\pm\epsilon$ in Lemma 1.1 with sharper bounds. This amounts to second-order asymptotics of the $\overline{H}_t(x)$, i.e. to obtain second-order uniform Cramér-Lundberg expansions which does not appear easy at all.

# References


[1] S. Asmussen (2000) *Ruin Probabilities.* World Scientific.

[2] S. Asmussen (2003) *Applied Probability and Queues* (2nd ed.). Springer-Verlag.

[3] N.H. Bingham, C.M. Goldie, & J.L. Teugels (1987) *Regular Variation.* Cambridge University Press.

[4] A. Bobbio & K. Trivedi (1990) Computation of the distribution of the completion time when the work requirement is a PH random variable. *Stochastic Models* **6**, 133–150.

[5] X. Castillo & D.P. Siewiorek (1980) A performance-reliability model for computing systems. *Proc FTCS-10, Silver Spring, MD, IEEE Computer Soc.*, 187–192.

[6] P.F. Chimento, Jr. & K.S. Trivedi (1993) The completion time of programs on processors subject to failure and repair. *IEEE Trans. on Computers* **42**(1).

[7] B.S. Chlebus, R. De Prisco, & A.A. Shvartsman (2001) Performing tasks on synchronous restartable message-passing processors. *Distributed Computing* **14**, 49–64.

[8] R. DePrisco, A. Mayer, & M. Yung (1994) Time-optimal message-efficient work performance in the presence of faults. *Proc. 13th ACM PODC*, 161–172.

[9] M. Finkelstein & V. Esaulova (2006) Asymptotic behavior of a general class of mixture failure rates. *Adv. Appl. Probab.* **38**, 244–262.

[10] D. Jagerman (1982) An inversion technique for the Laplace transform. *Bell Syst.Tech. J.* **61**, 1995–2002.

[11] V. Kalashnikov (1997) *Geometric Sums: Bounds for Rare Events with Applications. Risk Analysis, Reliability, Queueing.* Kluwer.

[12] N.V. Kartashov (1980) A uniform asymptotic renewal theorem. *Th. Probab. Appl.* **25**, 589–592.

[13] N.V. Kartashov (1982) Equivalence of uniform renewal theorems and their criteria. *Teor. Veoryuatnost. i Mat. Statist.* **27**, 51–60 (in Russian).

[14] V. Kulkarni, V. Nicola, & K. Trivedi (1986) On modeling the performance and reliability of multimode systems. *The Journal of Systems and Software* **6**, 175–183.

[15] V. Kulkarni, V. Nicola, & K. Trivedi (1987) The completion time of a job on a multimode system. *Adv. Appl. Probab.* **19**, 932–954.

[16] L. Lipsky (2007) *Queueing Theory. A Linear Algebraic Approach* (2nd ed.). Springer-Verlag.

[17] A. Müller & D. Stoyan (2002) *Comparison Methods for Stochastic Models and Risks.* Wiley.

[18] R. Sheahan, L. Lipsky, P. Fiorini & S. Asmussen (2006) On the distribution of task completion times for tasks that must restart from the beginning if failure occurs. *SIGMETRICS Performance Evaluation Review* **34**, 24–26.

[19] G. Willmot & X. Liu (2001) *Lundberg Approximations for Compound Distributions with Insurance Applications.* Lectute Notes in Statistics **156**. Springer-Verlag.

[20] M. Wang & M. Woodroofe (1996) A uniform renewal theorem. *Sequential Anal.* **15**, 21–36.