1 Rare event simulation and importance sampling

Suppose we wish to use Monte Carlo simulation to estimate a probability $p = P(A)$ when the event A is "rare" (e.g., when p is very small). An example would be $p = P(D_k > b)$ with a very large b for delay D_k of the k^{th} customer in a queue. We could naively simulate n (large) iid copies of A, denoted by $A_1, A_2, \ldots, A_n$, then set $X_i = I\{A_i\}$ and use the crude estimate

$$p \approx \overline{p}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{1}$$

But this is not a good idea: $\mu \stackrel{\text{def}}{=} E(X_i) = P(A) = p$ and $\sigma^2 \stackrel{\text{def}}{=} Var(X_i) = p(1-p)$ and so, since p is assumed very small, the ratio $\sigma/\mu = \sqrt{p(1-p)}/p \sim 1/\sqrt{p} \longrightarrow \infty$ as $p \downarrow 0$; relative to μ, σ is of a much larger magnitude. This is very bad since when constructing confidence intervals,

$$\overline{p}(n) \pm \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}},$$

the length of the interval is in units of σ: If σ is much larger than what we are trying to estimate, μ, then the confidence interval will be way too large to be of any use. It would be like saying "I am 95% confident that he weighs 140 pounds plus or minus 500 pounds".

To make matters worse, increasing the number n of copies in the Monte Carlo so as to reduce the interval length, while sounding OK, could be impractical since n would end up having to be enormous.

Importance sampling is a technique that gets around this problem by changing the probability distributions of the model so as to make the rare event happen often instead of rarely. To understand the basic idea, suppose we wish to compute $E(h(X)) = \int h(x)f(x)dx$ for a continuous random variable X distributed with density $f(x)$. For example, if $h(x) = I\{x > b\}$ for a given large b, then $h(X) = I\{X > b\}$ and $E(h(X)) = P(X > b)$.

Now let $g(x)$ be any other density such that $f(x) = 0$ whenever $g(x) = 0$, and observe that we can re-write

$$\begin{aligned} E(h(X)) &= \int h(x)f(x)dx \\ &= \int \Big[h(x)\frac{f(x)}{g(x)}\Big]g(x)dx \\ &= \tilde{E}\Big[h(X)\frac{f(X)}{g(X)}\Big], \end{aligned}$$

where $\tilde{E}$ denotes expected value when g is used as the distribution of X (as opposed to the original distribution f). In other words: If X has distribution g, then the expected value of $h(X)\frac{f(X)}{g(X)}$ is the same as what we originally wanted. The ratio $L(X) = \frac{f(X)}{g(X)}$ is called the *likelihood ratio.* We can write

$$E(h(X)) = \tilde{E}(h(X)L(X)); \tag{2}$$

the left-hand side uses distribution f for X, while the right-hand side uses distribution g for X.

To make this work in our favor, we would want to choose g so that the variance of $h(X)L(X)$ (under g) is small relative to its mean.

We can easily generalize this idea to multi-dimensions: Suppose $h = h(X_1, \ldots, X_k)$ is real-valued where $(X_1, \ldots, X_k)$ has joint density $f(x_1, \ldots x_k)$. Then for an alternative joint density $g(x_1, \ldots x_k)$, we once again can write

$$E(h(X_1, \ldots, X_k)) = \tilde{E}(h(X_1, \ldots, X_k)L(X_1, \ldots, X_k)), \tag{3}$$

where $L(X_1, \ldots, X_k) = \frac{f(X_1, \ldots, X_k)}{g(X_1, \ldots, X_k)}$, and $\tilde{E}$ denotes expected value when g is used as the joint distribution of $(X_1, \ldots, X_k)$.

1.1 Application to the FIFO GI/GI/1 queue

As a concrete example, let's consider the FIFO GI/GI/1 queue with iid service times S_i distributed as $G(x) = P(S \le x)$ and iid interarrival times T_i distributed as $A(x) = P(T \le x)$. $\lambda = E(T)^{-1}$ is the arrival rate, $1/\mu = E(S)$ and $\rho \stackrel{\text{def}}{=} \lambda/\mu$. We assume the stability condition $\rho < 1$, which can be equivalently stated as $E(S - T) < 0$.

$\Delta_i \stackrel{\text{def}}{=} S_i - T_i,\ i \ge 0$, and the FIFO delay recursion is given by $D_{n+1} = (D_n + \Delta_n)^+,\ n \ge 0$, and we shall assume that $D_0 = 0$. If we form the random walk

$$R_k = \Delta_1 + \cdots + \Delta_k,\ R_0 = 0,$$

then since $E(\Delta) = E(S - T) < 0$, the random walk has *negative drift* and hence tends to $-\infty$ as time goes on: $R_k \to -\infty$ as $k \to \infty$ wp1. Before it drifts off to $-\infty$, however, it first reaches a finite maximum $M \stackrel{\text{def}}{=} \max_{k \ge 0} R_k$ which is a non-negative random variable. It is well known (*duality* between queue and risk) that the distribution of D_k (for any fixed k) is the same as the distribution of $M_k \stackrel{\text{def}}{=} \max_{0 \le j \le k} R_j$ (the maximum up to time k) and thus taking limits $(k \to \infty)$ yields that the distribution of stationary delay D is the same as the distribution of the all-time maximum M.

Thus for any $b \ge 0$,

$$P(D_k > b) = P(M_k > b) \tag{4}$$
$$P(D > b) = P(M > b). \tag{5}$$

For a given fixed $b \ge 0$ let

$$\tau(b) = \begin{cases} \min\{k \ge 0 : R_k > b\} & \text{if } R_k > b \text{ for some } k, \\ \infty & \text{if } R_k \le b \text{ for all } k. \end{cases}$$

$\tau(b)$ is the *first passage time* above b, it denotes the first time at which (if ever) the random walk goes above b. The case $\tau(b) = \infty$ must be included because the random walk has negative drift and thus might never reach a value above b (before eventually drifting to $-\infty$). Noting that $\{\tau(b) \le k\} = \{M_k > b\}$, and that $\{\tau(b) < \infty\} = \{M > b\}$, we have $P(\tau(b) \le k) = P(M_k > b)$, and $P(\tau(b) < \infty) = P(M > b)$. Thus (4)-(5) become

$$P(D_k > b) = P(M_k > b) = P(\tau(b) \le k) \tag{6}$$
$$P(D > b) = P(M > b) = P(\tau(b) < \infty). \tag{7}$$

1.1.1 Importance sampling in the light-tailed service case

Let $F(x) = P(\Delta \leq x),\ x \geq 0$, and *assume that it has a density function* $f(x)$. We shall also assume that service times are *light-tailed*: $E(e^{\epsilon S}) < \infty$ for some $\epsilon > 0$ (e.g., S has a finite moment generating function), which implies that the tail $P(S > x)$ tends to 0 fast like an exponential tail does. (A Pareto tail, however, such as x^{-3}, does not have this property; it is an example of a *heavy-tailed* distribution.)

We further shall assume the existence of a $\gamma > 0$ such that

$$E(e^{\gamma\Delta}) = \int_{-\infty}^{\infty} e^{\gamma x} f(x)dx = 1. \tag{8}$$

Defining the moment generating function $K(\epsilon) = E(e^{\epsilon\Delta}) = E(e^{\epsilon S})E(e^{-\epsilon T})$ and observing that $K(0) = 1$ and $K'(0) = E(\Delta) < 0$, and K is convex $K''(\epsilon) > 0$, we see that the condition (8) would hold under suitable conditions; conditions ensuring that K, while moving down below 1 for a while, shoots back upwards and hits 1 as ϵ increases. The value γ at which it hits 1 is called the *Lundberg constant*. Furthermore, since it is increasing when it hits 1 at γ, it must hold that $K'(\gamma) > 0$.

Let us change the distribution of Δ to have density

$$g(x) = e^{\gamma x} f(x). \tag{9}$$

We know that g defines a probability density, $\int g(x)dx = 1$, because of the definition of γ in (8). We say that we have *exponentially tilted* or *twisted* the distribution f to be that of g. In general we could take any value $\epsilon > 0$ for which $K(\epsilon) < \infty$, and change f to the new twisted density

$$g_\epsilon(x) = \frac{e^{\epsilon x} f(x)}{K(\epsilon)}. \tag{10}$$

Our g in (9) is the special case when ϵ is set to be the Lundberg constant in (8); $g = g_\gamma$.

It is easy to show that in fact, $\tilde{E}(\Delta) > 0$, that is, using distribution g for the random walk increments Δ_i makes the random walk now have positive drift! (To see this, note that $\tilde{E}(\Delta) = K'(\gamma)$ and recall that we argued above that $K'(\gamma) > 0$.)[1]

Thus for a given large $b > 0$ it is more likely that events such as $\{D_n > b\}$ will occur as compared to the original case when f is used. In fact $\tilde{P}(M > b) = 1 = \tilde{P}(\tau(b) < \infty)$, where $\tilde{P}$ denotes using g instead of f: the random walk will now with certainty tend to $+\infty$ and hence pass any value b along the way no matter how large; $R_n \to \infty$ wp1 under $\tilde{P}$.

Noting that the likelihood ratio function $L(x) = f(x)/g(x) = e^{-\gamma x}$ and using $h(x) = I\{x > b\}$, we can use (2) and conclude, for example, that

$$E(h(\Delta_1)) = P(\Delta_1 > b) = P(M_1 > b) = P(D_1 > b) = \tilde{E}[e^{-\gamma\Delta_1} I\{\Delta_1 > b\}]. \tag{11}$$

In two dimensions, utilizing (3), we can take $h(x_1, x_2) = I\{x_1 > b \text{ or } x_1 + x_2 > b\}$ yielding $h(\Delta_1, \Delta_2) = I\{M_2 > b\}$. We make the two increments iid each distributed as g in in (9), so that their joint distribution is the product $g_2(x_1, x_2) = g(x_1)g(x_2) = e^{\gamma(x_1+x_2)} f(x_1)f(x_2)$. The original joint distribution is the product $f_2(x_1, x_2) = f(x_1)f(x_2)$ and so $L(x_1, x_2) = f_2/g_2 = e^{-\gamma(x_1+x_2)}$, and therefore $L(\Delta_1, \Delta_2) = e^{-\gamma R_2}$. This then yields

$$E(h(\Delta_1, \Delta_2)) = P(M_2 > b) = P(D_2 > b) = \tilde{E}[e^{-\gamma R_2} I\{M_2 > b\}]. \tag{12}$$

[1] In the case when $S \sim exp(\mu)$ and $T \sim exp(\lambda)$, the M/M/1 case, with $\lambda < \mu$, it is easily seen that $\gamma = \mu - \lambda$ and that the twisted density has the effect of swapping the rates: under g, the service times S_i become iid with an exponential distribution at rate λ, while the interarrival times T_i become iid with an exponential distribution at rate μ; this yields an unstable queue.

Continuing analogously to higher dimensions then yields for any $k \geq 1$:

$$P(D_k > b) = P(M_k > b) = \tilde{E}[e^{-\gamma R_k} I\{M_k > b\}]. \tag{13}$$

At this point, let us discuss how to use (13) in a simulation. Our objective is to estimate $P(D_k > b)$ for a very large value of b. We thus simulate the first k steps of a random walk, $R_1, \ldots, R_k$, having iid positive drift increments distributed as g in (9). We compute $M_k = \max_{0 \leq j \leq k} R_j$ and obtain a first copy $X_1 = e^{-\gamma R_k} I\{M_k > b\}$. Then, independently, we simulate a second copy and so on yielding n (large) iid copies to be used in our Monte Carlo estimate

$$P(D_k > b) \approx \tilde{p}(n) = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{14}$$

(We of course need to be able to simulate from g, we assume this is so.)

If we had used the naive approach, we would have simulated iid copies of $X_1 = I\{M_k > b\}$ where the increments of the random walk would be distributed as f and thus have negative drift. The event in question, $I\{M_k > b\}$, would thus rarely happen; we would be in the bad situation outlined in the beginning of these notes. With our new approach, the random walk is changed to have positive drift and thus this same event, $I\{M_k > b\}$, now is very likely to occur. The likelihood ratio factor $e^{-\gamma R_k}$ must be multiplied along before taking expected values so as to bring the answer down to its true value $P(M_k > b)$ as opposed to the larger (incorrect) value $\tilde{P}(M_k > b)$.

It turns out (using martingale theory, see Section 1.2 below) that we can re-express the right-hand side of (13) as

$$\tilde{E}[e^{-\gamma R_{\tau(b)}} I\{M_k > b\}],$$

so that after taking the limit as $k \to \infty$ we obtain

$$P(D > b) = P(M > b) = \tilde{E}[e^{-\gamma R_{\tau(b)}} I\{M > b\}]. \tag{15}$$

But as we know, $\tilde{P}(M > b) = \tilde{P}(\tau(b) < \infty) = 1$ since the random walk has positive drift under $\tilde{P}$ (said differently, $M = \infty$ wp1 under $\tilde{P}$, so for any $b > 0$, $I\{M > b\} = 1$ wp1 under $\tilde{P}$). Thus (15) becomes

$$P(D > b) = P(M > b) = \tilde{E}(e^{-\gamma R_{\tau(b)}}). \tag{16}$$

But by definition, at time $\tau(b)$ the random walk has shot passed level b; $R_{\tau(b)} = b + B$, where $B = R_{\tau(b)} - b$ denotes the *overshoot*. We finally arrive at

$$P(D > b) = P(M > b) = e^{-\gamma b} \tilde{E}(e^{-\gamma B}). \tag{17}$$

In essence, we have reduced the problem of computing $P(D > b)$ to computing the Laplace transform evaluated at γ, $\tilde{E}(e^{-\gamma B})$, of the overshoot B of a positive drift random walk.

To put this to good use, we then use Monte Carlo simulation to estimate $\tilde{E}(e^{-\gamma B})$: Simulate the positive drift random walk with increments iid distributed as g until it first passes level b, and let B_1 denote the overshoot. Set $X_1 = e^{-\gamma B_1}$. Independently repeat the simulation to obtain another copy of the overshoot B_2 and so on for a total of n such iid copies, $X_i = e^{-\gamma B_i}$, $i = 1, 2, \ldots n$. Then use as the estimate

$$P(D > b) \approx e^{-\gamma b} \Big[\frac{1}{n} \sum_{i=1}^{n} X_i\Big]. \tag{18}$$

Note in passing that since $\tilde{E}(e^{-\gamma B}) < 1$ we conclude from (17) that $P(D > b) \le e^{-\gamma b}$, an exponential upper bound on the tail of delay. It turns out that under suitable further conditions, it can be proved that there exists a constant $C > 0$ such that

$$P(D > b) \sim Ce^{-\gamma b}, \text{ as } b \to \infty,$$

by which we mean that

$$\lim_{b\to\infty} \frac{P(D > b)}{Ce^{-\gamma b}} = 1.$$

This is known as the *Lundberg approximation*[2].

1.2 Martingales and the likelihood ratio identity

$L_k = e^{-\gamma R_k},\ k \ge 0$, is a mean 1 martingale under $\tilde{P}$ because $\tilde{E}[e^{-\gamma\Delta}] = \int e^{-\gamma x} e^{+\gamma x} f(x)dx = \int f(x)dx = 1$. Thus, by optional sampling, for each fixed k, $1 = \tilde{E}(L_k) = \tilde{E}(L_{\tau(b)\wedge k})$ since $\tau(b) \wedge k$ is a bounded stopping time. But $\tilde{E}(L_{\tau(b)\wedge k}) = \tilde{E}(L_{\tau(b)} I\{M_k > b\}) + \tilde{E}(L_k I\{M_k \le b\})$, while $\tilde{E}(L_k) = \tilde{E}(L_k I\{M_k > b\}) + \tilde{E}(L_k I\{M_k \le b\})$. Equating these two then yields $\tilde{E}(L_k I\{M_k > b\}) = \tilde{E}(L_{\tau(b)} I\{M_k > b\})$. Lurking in here is the famous *likelihood ratio identity*: Given any stopping time τ ($\tau = \tau(b)$ for example), it holds for any event $A \subseteq \{\tau < \infty\}$ that $P(A) = \tilde{E}(L_\tau I\{A\})$. (The filtration in our case is $\mathcal{F}_n = \sigma\{\Delta_1, \ldots, \Delta_n\} = \sigma\{R_1, \ldots, R_n\}$.)

Finally note that, meanwhile, $L_k^{-1} = e^{\gamma R_k}$ is a mean 1 martingale under P due to (8).

There is a back and forth between the two probabilities P and $\tilde{P}$:

$$\tilde{P}(A) = E(L_k^{-1} I\{A\}),\ A \in \mathcal{F}_k. \tag{19}$$

$$P(A) = \tilde{E}(L_k I\{A\}),\ A \in \mathcal{F}_k. \tag{20}$$

The general framework here: We use the canonical space $\Omega = \mathbf{R}^{\mathbf{N}} = \{(x_0, x_1, x_2, \ldots) : x_i \in \mathbf{R}\}$; the space of sequences of real numbers (the sample space for discrete-time stochastic processes) endowed with the standard Borel $\sigma-$ field and filtration $\{\mathcal{F}_k : k \ge 0\}$. Each random element on this space corresponds to a stochastic process, denoted by $\mathbf{R} = \{R_n : n \ge 0\}$. We start with the probability measure P corresponding to $\mathbf{R}$ being a random walk $R_k = \Delta_1 + \cdots + \Delta_k,\ R_0 = 0$, with iid increments Δ_i having density $f(x)$ with assumed Lundberg constant $\gamma > 0$ as defined in (8). Define the non-negative mean 1 martingale $L_k = e^{\gamma R_k},\ k \ge 1,\ L_0 = 1$, and then define a new probability on Ω via

$$\tilde{P}(A) = E(L_k I\{A\}),\ A \in \mathcal{F}_k. \tag{21}$$

(That (21) really defines a unique probability on Ω follows by *Kolomogorov's extension theorem* in probability theory: For each k (21) does define a probability on $\mathcal{F}_k$, denote this by $\tilde{P}_k$. Consistency of these probabilities follows by the martingale property, for $m < k$, $\tilde{P}_k(A) = \tilde{P}_m(A)$ $A \in \mathcal{F}_m$; that is, $\tilde{P}_k(A)$ restricted to $\mathcal{F}_m$ is the same as $\tilde{P}_m$.)

[2] $B = B(b)$ depends on b. If (under $\tilde{P}$) $B(b)$ converges weakly (in distribution) as $b \to \infty$ to (say) a rv B^*, then $C = \tilde{E}(e^{-\gamma B^*})$. The needed conditions for such weak convergence are that $\tilde{E}(\Delta) < \infty$ and that the first strictly ascending ladder height $H = R_{\tau(0)}$ have a non-lattice distribution. But it is known that H is non-lattice if and only if the distribution of Δ is so, and in our case it has a density g, hence is non-lattice. (We already know that $K'(\gamma) = \tilde{E}(\Delta) > 0$ but it could be infinite.)

$\tilde{P}$ turns out to be the distribution of a random walk as well, but with new iid increments distributed with the tilted density g defined in (10): $\tilde{P}(R_1 \leq x) = E(L_1 I\{R_1 \leq x\}) = G(x) = \int_{-\infty}^{x} g(x)dx$, and more generally, with $\Delta_i = R_{i-1} - R_i,\ i \geq 1$, $\tilde{P}(\Delta_1 \leq y_1, \ldots, Y_k \leq y_k) = E(L_k I\{\Delta_1 \leq y_1, \ldots, \Delta_k \leq y_k\}) = G(y_1) \times \cdots \times G(y_k)$. Moreover, we can go the other way by using the non-negative mean 1 martingale $L_k^{-1} = e^{-\gamma R_k}$:

$$P(A) = E(L_k L_k^{-1} I\{A\}) = \tilde{E}(L_k^{-1} I\{A\})),\ A \in \mathcal{F}_k. \tag{22}$$

The general change-of-measure approach allows as a starting point a given probability P, and a non-negative mean 1 martingale $\{L_k\}$. Then $\tilde{P}(A) = E(L_k I\{A\}),\ A \in \mathcal{F}_k$ is defined. Then the likelihood ratio identity is: Given any stopping time τ, it holds for any event $A \subseteq \{\tau < \infty\}$ that $P(A) = \tilde{E}(L_\tau^{-1} I\{A\})$. This works fine in continuous time $t \in [0, \infty)$, but then the canonical space used is $\mathcal{D}[0, \infty)$, the space of functions that are continuous from the right and have left-hand limits, equipped with the Skorohod topology.

Remark 1.1 The Lundberg constant and the change of measure using it generalizes nicely to continuous-time Levy processes, Brownian motion for example (and for Brownian motion such change of measure results are usually known as *Girsanov's Theorem*). For example, if $X(t) = \sigma B(t) + \mu t$ is a Brownian motion with negative drift, $\mu < 0$, then solving $1 = E(e^{\gamma X(1)}) = e^{\gamma\mu + \frac{\gamma^2\sigma^2}{2}}$ yields $\gamma = 2|\mu|/\sigma^2$. This yields the martingale $L(t) = e^{\gamma X(t)}$ and the new measure $\tilde{P}(A) = E(e^{\gamma X(t)} I\{A\}),\ A \in \mathcal{F}_t$. It is easily seen that under $\tilde{P}$, the process $X(t)$ remains a Brownian motion but with *positive* drift $\tilde{\mu} = |\mu|$, and the variance remains the same as it was. Using the likelihood ratio identity with $\tau(b) = \inf\{t \geq 0 : X(t) > b\} = \inf\{t \geq 0 : X(t) = b\}$ (via continuity of sample paths) then yields as in (17)

$P(M > b) = e^{-\gamma b}\tilde{E}(e^{-\gamma B})$. But now, $B = 0$ since $X(t)$ has continuous sample paths; there is no overshoot, b is hit exactly. Thus we get an exact exponential distribution, $P(M > b) = e^{-\gamma b},\ b > 0$ for the maximum of a negative drift Brownian motion, a well-known result that can be derived using more basic principles.

As a second example, we consider the Levy process $Z(t) = \sigma B(t) + \mu t + Y(t)$, where independently we have added on a compound Poisson process

$$Y(t) = \sum_{i=1}^{N(t)} J_i,$$

where $\{N(t)\}$ is a Poisson process at rate λ and the J_i are iid with a given distribution $H(x) = P(J \leq x),\ x \in \mathbf{R}$. Also, let $\hat{H}(s) = E(e^{sJ})$ denote the moment generating function of H, assumed to be finite for sufficiently small $s > 0$ (e.g., H is light-tailed). Choosing any ϵ for which $K_Z(\epsilon) \stackrel{\text{def}}{=} E(e^{\epsilon Z(1)}) = e^{\epsilon\mu + \frac{\epsilon^2\sigma^2}{2} + \lambda(\hat{H}(\epsilon) - 1)} < \infty$, we always obtain a martingale $L(t) = [K_Z(\epsilon)]^{-1} e^{\epsilon Z(t)}$, and a new measure $\tilde{P}(A) = E(L(t) I\{A\}),\ A \in \mathcal{F}_t$.

We now show that under $\tilde{P}$, Z remains the same kind of Levy process but with drift $\tilde{\mu} = \mu + \epsilon\sigma^2$, variance unchanged $\tilde{\sigma}^2 = \sigma^2$, $\tilde{\lambda} = \lambda\hat{H}(\epsilon)$ and the distribution H exponentially tilted to be $\tilde{H}(x)$ given by $d\tilde{H}(x) = \frac{e^{\epsilon x} dH(x)}{\hat{H}(\epsilon)}$.

To this end, we need to confirm that for $s \geq 0$,

$$\tilde{E}(e^{sZ(1)}) = e^{s\tilde{\mu} + \frac{s^2\sigma^2}{2} + \tilde{\lambda}(\hat{\tilde{H}}(s) - 1)},$$

where $\hat{\tilde{H}}(s) = E(e^{s\tilde{J}}) = \int e^{sx} d\tilde{H}(x) = \frac{\hat{H}(\epsilon + s)}{\hat{H}(\epsilon)}$.

Direct calculations yield

$$
\begin{aligned}
\tilde{E}(e^{sZ(1)}) &= E(L(1)e^{sZ(1)}) \\
&= [K_Z(\epsilon)]^{-1}E(e^{\epsilon Z(1)}e^{sZ(1)}) \\
&= [K_Z(\epsilon)]^{-1}K_Z(\epsilon+s) \\
&= e^{-\epsilon\mu-\frac{\epsilon^2\sigma^2}{2}}e^{-\lambda(\hat{H}(\epsilon)-1)}e^{(\epsilon+s)\mu+\frac{(\epsilon+s)^2\sigma^2}{2}+\lambda(\hat{H}(s+\epsilon)-1)} \\
&= e^{s\tilde{\mu}+\frac{s^2\sigma^2}{2}+\tilde{\lambda}(\hat{\tilde{H}}(s)-1)},
\end{aligned}
$$

as was to be shown.

As for the Lundberg constant: We assume apriori that Z has negative drift, $E(Z(1)) = \mu + \lambda E(J) < 0$, so that $M = \max_{t\geq 0} Z(t)$ defines a finite random variable. Solving for a $\gamma > 0$ such that $1 = K_Z(\gamma) = e^{\gamma\mu+\frac{\gamma^2\sigma^2}{2}+\lambda(\hat{H}(\gamma)-1)}$, leads to the equation

$$\gamma\mu + \frac{\gamma^2\sigma^2}{2} + \lambda(\hat{H}(\gamma) - 1) = 0.$$

Assuming a solution exists (this depends on H), then as the martingale we use $L(t) = e^{\gamma Z(t)}$. Under $\tilde{P}$, Z now has positive drift, $K'_Z(\gamma) = \tilde{E}(Z(1)) = \tilde{\mu} + \tilde{\lambda}E(\tilde{J}) > 0$.

Just as for the FIFO/GI/GI/1 queue, we obtain exactly the same kind of exponential bound for the tail of M: Using the likelihood ratio identity with $\tau(b) = \inf\{t \geq 0 : Z(t) > b\}$ then yields as in (17), $P(M > b) = e^{-\gamma b}\tilde{E}(e^{-\gamma B})$. Now, because of the "jumps" J_i, there is an overshoot B to deal with (unless the jumps are ≤ 0 wp1.). All of this goes thru with general negative drift Levy processes, the idea being that under $\tilde{P}$ the process remains Levy, but with new parameters making it have positive drift.