Rare-event simulation: Assignment 3

Patrick Laub

March 24, 2020

Due: Wednesday, April 7, 1:30 pm

Submission by email, send completed Jupyter notebook.

Introduction: You now have a very good grasp of the rare-event concepts, algorithms, and the practical matters of using Python. For this last assignment, I'd like to focus on some real-world skills. Ideally, you should be able to:

1) find a research paper and implement the algorithm it proposes (without relying on one of my code demonstrations),
2) use a large (& somewhat cumbersome) software package to solve problems you care about,
3) solve rare-event problems using any algorithm (or combination of algorithms) that you deem appropriate.

The questions in this assignment address each of these learning goals in turn.

Sequential Monte Carlo (for rare-events)

1) Find the paper *Sequential Monte Carlo for rare event estimation* by Cérou et al from 2012. Use one (or both if you like) of the two main algorithms in here (i.e. the biased *adaptive* SMC or the unbiased SMC with predetermined levels) to solve

$$\ell = \mathbb{P}(X_1 + X_2 > \gamma)$$

where $X_i \overset{\text{i.i.d.}}{\sim} \mathsf{Pareto}(\alpha, 1)$. Take $\alpha = 7$ and $\gamma = 10$. We use the same definition as in previous classes, that the $\mathsf{Pareto}(\alpha, 1)$ has the p.d.f.

$$f(x) = \begin{cases} \alpha\,(1+x)^{-(\alpha+1)} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

2) Make a bivariate scatterplot of the $(X_1, X_2) \mid (X_1 + X_2 > \gamma)$ samples which were generated at the last stage of the SMC process (if there are many duplicates, perhaps resample and regenerate them). Shade the area of the plot red which satisfies $x_1 + x_2 \leq \gamma$.

Let's add a story to this problem. Imagine that $X_1 + X_2$ represents the aggregate claim size for an insurer with two risks, and the γ is some critical threshold (e.g. bankruptcy). Imagine the insurer has come to you and asked that you "describe the characteristics of the portfolio in the event of bankruptcy".

As such, take a look at the samples and try to characterise the behaviour of the samples. Write a short description (less than 3 sentences) that you can use to explain to the insurer which doesn't use any mathematical jargon.

Hint 1: Look up the "principle of the single big jump".

Hint 2: The two X's are identically distributed, so the scatterplots should look pretty symmetrical about the $x_1 = x_2$ line. They won't be exactly symmetrical, but if you have say $\geq 75\%$ of particles on one side of this line, then it indicates your SMC fell apart somewhere, and you should rerun with different hyperparameters.

PyMC3

3) Consider

$$\ell = \mathbb{P}(X_1 + X_2 > \gamma)$$

where $X_i \overset{\text{i.i.d.}}{\sim} \mathsf{Gamma}(r, m)$. Take $r = 10$, $m = 1$, and $\gamma = 50$. We again use the $\mathsf{Gamma}(r, m)$ with p.d.f.

$$f(x) = \begin{cases} \frac{x^{r-1}\mathrm{e}^{-\frac{x}{m}}}{\Gamma(r)m^r} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Solve this problem using the improved cross-entropy method but **use the PyMC3 library** to generate samples from the zero-variance IS distribution.

Hint: Look up the `Potential` function in PyMC3 to enforce the $X_1 + X_2 > \gamma$ constraint.

4) Repeat the analysis you did in question 2 for this new situation. That is, describe (without math jargon) the kind of outcomes which would lead to bankruptcy using the light-tailed gamma claims.

5) It seems criminal to have this library & not use it for it's intended purpose of Bayesian estimation. Find an example notebook which uses PyMC3 in a cool way & modify it (i.e. change it to work on a different dataset, or change the algorithm or the assumptions it uses) and put the result in here. Don't choose the same (or a very similar) starting notebook as any other team; once you have chosen an example to modify, write your names in this Google Doc along with a link to the example you've chosen to prevent anyone else taking your idea.

Open-ended

6) A man of age 55 signs a life annuity contract, according to which he has to pay a premium at (continuous) rate π until age 67 or the time T of death, and if still alive at 67, he will recieve a payment at rate 1 until death. Assuming a constant interest rate r, the equation determining π is therefore

$$\mathbb{E}\int_{45}^{\min(67,T)} \pi \mathrm{e}^{-rt}\mathrm{d}t = \mathbb{E}\int_{67}^{\max(67,T)} \mathrm{e}^{-rt}\mathrm{d}t$$

(the two sides are the present values of the contributions, respectively benefits). Assume that $r = 4\%$ and that T follows are $\mathsf{GompertzMakeham}(a, b, c)$ distribution conditioned on $T > 55$,

which is widely used lifetime distribution in life insurance. The $\mathsf{GompertzMakeham}(a, b, c)$ has p.d.f.

$$f(t) = \begin{cases} [a + b\,\mathrm{e}^{ct}] \exp\big[-at - \frac{b}{c}(\mathrm{e}^{ct} - 1)\big] & \text{if } t > 0 \\ 0 & \text{otherwise.} \end{cases}$$

and we'll use $a = 5 \cdot 10^{-4}$, $b = 7.5858 \cdot 10^{-5}$, $c = \log(1.09144)$ which is the "G82" base.

Estimate the fair value of π satisfying the equation above. Try to be accurate to 3-5 significant digits (with at least 75% confidence).

The question can originally be found in the book *Stochastic Simulation* by Asmussen & Glynn (2007).

7) Consider the risk posed by one contract with the fair value of π you calculated above. Calculate the 99.99% Value-at-Risk for the insurer and the associated (very old) age at death of the policyholder in this 99.99% worst-case scenario. Again, try to be as accurate as you can.