MAST30027: Modern Applied Statistics

Assignment 3, 2021.

Due: 5pm Wednesday October 6th

- This assignment is worth 8% of your total mark.
- To get full marks, show your working including 1) R commands and outputs you use, 2) mathematics derivation, and 3) rigorous explanation why you reach conclusions or answers. If you just provide final answers, you will get zero mark.
- Your assignment must be submitted on Canvas LMS as a single PDF document only (no other formats allowed). Your answers must be clearly numbered and in the same order as the assignment questions.
- The LMS will not accept late submissions. It is your responsibility to ensure that your assignments are submitted correctly and on time, and problems with online submissions are not a valid excuse for submitting a late or incorrect version of an assignment.
- We will mark a selected set of problems. We will select problems worth $\geq$ 50% of the full marks listed.
- Answers including images of screen-captured R codes or figures won't be marked.
- Also, please read the "Assessments" section in "Subject Overview" page of the LMS.

---

1. The file `assignment3_prob1.txt` contains final exam scores of 100 students in Modern Applied Statistics. We can read the scores as follows.

```
> X = scan(file="assignment3_prob1.txt", what=double())
Read 100 items
> length(X)
[1] 100
> mean(X)
[1] 75.726
```

Suppose that the 100 scores are independent to each other and they follow Normal distribution with mean = 75 and unknown precision τ. Specifically, let $x_1, \ldots, x_{100}$ be the final exam scores, and

$$x_i \sim \mathrm{N}(75, \frac{1}{\tau}) \quad \text{for} \quad i = 1, \ldots, 100.$$

Suppose that the precision τ has a Gamma(2, 1) prior distribution, where Gamma(α, β) has the pdf

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x).$$

   (a) **(5 marks)** Derive the posterior distribution of the precision τ conditioned on the final exam scores of 100 students, $p(\tau|x_1, \ldots, x_{100})$. Evaluate parameters in the posterior distribution using the data from `assignment3_prob1.txt`.

(b) **(6 marks)** Derive the posterior predictive distribution for a new score $\tilde{x}$, $p(\tilde{x}|x_1, \ldots, x_{100})$. Evaluate parameters in the posterior predictive distribution using the data from `assignment3_prob1.txt`.

[Hint for (b)] A three-parameter version of a t distribution (Jackman, S. (2009)), denoted by $t(\nu, a, b)$, has the pdf

$$p(x|\nu, a, b) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu b}} \left(1 + \frac{1}{\nu}\frac{(x-a)^2}{b}\right)^{-\frac{\nu+1}{2}}.$$

2. **(9 marks)** Gamma random variables can be used to simulate chi-square, t, F, beta, and Dirichlet distributions, as well as being useful in their own right. Hence it is important to be able to generate gamma r.v.s as efficiently as possible. In this assignment we investigate a popular algorithm due to Marsaglia and Tsang, for $\alpha \geq 1$.[1]

(1) If $X \sim \text{Gamma}(\alpha, 1)$ then $X/\lambda \sim \text{Gamma}(\alpha, \lambda)$.
Note that Gamma(α, β) has the pdf

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x).$$

(2) Assume that $h(x)$ is strictly increasing and maps the range of x onto $[0, \infty)$. If X has density $h(x)^{\alpha-1}e^{-h(x)}h'(x)/\Gamma(\alpha)$ then $Y = h(X) \sim \text{Gamma}(\alpha, 1)$.

(3) Given $\alpha \geq 1$ put $d = \alpha - 1/3$ and $c = 1/\sqrt{9d}$ and then define $h : [-1/c, \infty) \to [0, \infty)$ by $h(x) = d(1 + cx)^3$. Then,

$$\begin{aligned} h(x)^{\alpha-1}e^{-h(x)}h'(x) &\propto \exp(g(x)) \quad \text{where} \\ g(x) &= d\log((1+cx)^3) - d(1+cx)^3 + d \end{aligned}$$

and

$$\exp(g(x)) \leq \exp(-x^2/2) \quad \text{on} \quad [-1/c, \infty).$$

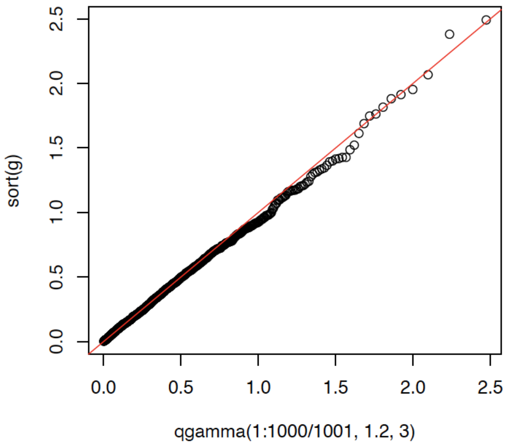
Using these facts (checking these facts is not required for the assignment 3), come up with an algorithm for simulating from Gamma(α, λ) for $\alpha \geq 1$. You may assume that you can already simulate from the standard normal distribution. 1) Provide a brief description of your algorithm (e.g., which algorithm you use. If you use a rejection method, explain which function has been used as envelop). 2) Code up your algorithm and use it to generate 1000 Gamma(1.2, 3) pseudo-random variables. Demonstrate that your algorithm is working using a q-q plot.

The following R commands show how to make q-q plot for 1000 samples generated by

```
> g <- rgamma(1000, 1.2, 3)
```

For your assignment, instead of rgamma(1000, 1.2, 3), you should write your own code which implements your algorithm.

```
> g <- rgamma(1000, 1.2, 3)
> plot(qgamma(1:1000/1001, 1.2, 3), sort(g))
> abline(0, 1, col="red")
```



3. **(2 marks)** Read the "Assessments" section of "Subject Overview" page of the LMS. Provide the requested information in the first page of your assignment.

[1] G. Marsaglia and W.W. Tsang, A simple method for generating gamma variables. ACM Trans. Math. Software, 26:363–371, 2000.