MAST90045 Assignment 3

This assignment counts for 17% of the total assessment in MAST90045. It is due on Sunday, the 7th of June. Assignments up to one week late will receive a 50% penalty. Assignments more than one week late will receive no marks.

Your assignment consists of two parts: a pdf/word report (saved as e.g., YourNameAssignment1Report.pdf) and an R submission of all your programs (saved as e.g., YourNameAssignment1Code.R).
The report should consist of **up to 7 pages** giving your solution to the problem, including figures and tables. You are expected to present your solution in a structured and coherent report (clearly written style, using complete sentences, etc.). Your report should contain appendices (which do not count towards the page limit). One appendix should consist of all the program code you used to produce the results presented in the body of the proposal. Your programs should be properly formatted and commented. A technical appendix (for the maths) is advisable as well.
Ideally you submit one R file per assignment, which if i source, produces all reported results.

Good luck!

NOTE 1: Whenever you are asked to estimate quantities(rather than calculate exactly), please consider giving a measure of certainty in your estimates.
NOTE 2: Please discuss expectations (not in a statistical sense) and results, before and after each task, respectively.

Conserving water

In this assignment you will simulate the operation of a domestic rainwater and greywater system, then use the simulation to decide the most cost effective system you can use to achieve certain water savings. The system consists of a rainwater tank, which collects rain from the roof, and a greywater tank, which collects water from the shower and the washing machine. Rainwater and greywater can be used either for flushing the loo or watering the garden (assuming that garden friendly laundry detergents are used).

Our simulation will be on a daily basis: each day we will simulate the volumes of rainwater and greywater that are collected, then simulate the volumes of water used for the loo and on the garden. Rainwater is preferred for the garden, but greywater will do, and vice versa for flushing the loo. If there is not enough stored water for flushing the loo or watering the garden, then mains water is used.

The first thing we need is a simulation of daily rainfall. The following data is taken from the Bureau of Meteorology website `http://www.bom.gov.au/climate/averages/tables/cw_086071_All.shtml`. The data is for Melbourne, based on monthly records since 1908.

Month	mean_rainfall	decile_1_rainfall	median_rainfall	decile_9_rainfall	mean_days_rain
Jan	47.6	10.9	36.9	99.2	8.4
Feb	48.0	6.8	32.6	108.5	7.5
Mar	50.4	11.8	38.8	104.9	9.4
Apr	57.3	17.7	49.8	114.4	11.8
May	55.8	21.3	55.1	91.2	14.6
Jun	49.0	25.1	42.6	85.3	15.4
Jul	47.5	23.3	44.4	72.1	16.1
Aug	50.0	23.5	49.2	77.7	16.1
Sep	58.1	27.8	53.0	92.4	14.9
Oct	66.4	25.4	67.0	111.3	14.2
Nov	60.4	21.5	53.8	114.7	11.8
Dec	59.5	17.6	52.3	110.3	10.4

We will assume that the rainfall is independent from one day to the next[1]. Let $X_{i,j}$ be the rainfall on the j-th day of month i ($i \in \{1, 2, \ldots, 12\}$), then we will assume that

$$\begin{aligned} X_{i,j} &\sim A_{i,j}B_{i,j}, \text{ for } A_{i,j} \text{ and } B_{i,j} \text{ independent, with} \\ A_{i,j} &\sim Bernoulli(p_i) \\ B_{i,j} &\sim Gamma(m_i, \lambda_i) \end{aligned}$$

Task 1 Because our data concerns monthly rather than daily rainfall, we also need a model for monthly rainfall. Let n_i be the number of days in month i, then the rainfall in month i is

$$Y_i = \sum_{j=1}^{n_i} X_{i,j}.$$

Show that

$$E(Y_i) = n_i p_i m_i / \lambda_i.$$

Put $N_i = \sum_{j=1}^{n_i} A_{i,j} \sim binom(n_i, p_i)$, then using the Law of Total Probability, show that for $x \geq 0$,

$$Pr(Y_i \leq x) = Pr(N_i = 0) + \sum_{k=1}^{n_i} Pr(C_{i,k} \leq x) Pr(N_i = k)$$

where $C_{i,k} \sim gamma(k\, m_i, \lambda_i)$. You may use the fact that the sum of independent $gamma(m_1, \lambda)$ and $gamma(m_2, \lambda)$ random variables has a $gamma(m_1 + m_2, \lambda)$ distribution.

Complete the following functions. Do *not* use `F_Y` in `Ysim`, instead use `rbinom` and `rgamma`.

```
Ysim <- function(p, m, la, n) {
  # simulate monthly rainfall
  ...
}

F_Y <- function(x, p, m, la, n) {
  # cdf of monthly rainfall
  ...
}
```

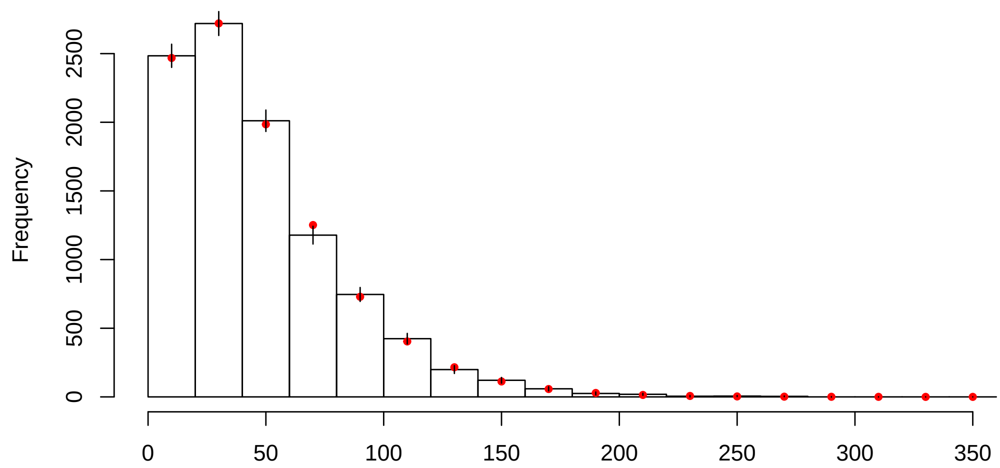
Next use `Ysim` to check that `F_Y` is correct, using parameters $p = 0.27$, $m = 0.24$, $\lambda = 0.043$ and $n = 31$. The idea is to plot a histogram using simulated values of Y, then compare the observed number in each bin with the expected number obtained from F_Y. A CI for the observed number

[1] This is not true, but we do not have data good enough to fit a more accurate model.

in each bin can be obtained using the following result. Let $Y(1), \ldots, Y(k)$ be an i.i.d. sample of Y's, then

$$\sum_{i=1}^{k} 1_{\{a<Y(i)\leq b\}} \approx N(k(F_Y(b) - F_Y(a)), k(F_Y(b) - F_Y(a))(1 - F_Y(b) + F_Y(a))).$$

Your histogram should look something like that below. Here $k = 10000$, the histogram bins have width 20, the vertical line attached to each bar of the histogram is a 95% CI, and the solid circle is the expected number of observations in a given bin.



Task 2 Next we need to estimate the parameters of $X_{i,j}$. We use an ad hoc approach. We will assume here, and after, that February always has 28 days.

mean_days_rain gives sample averages for the N_i. Use these to estimate the p_i.

Use $E(Y_i)$ and mean_rainfall to obtain an expression for m_i in terms of λ_i (and p_i).

To find λ_i we use the observed deciles. Let $\hat{F}_i$ be the empirical cdf of Y_i, obtained from observations of month i rainfall since 1908. The decile $d_i(k)$, $k = 1, \ldots, 9$, is the $10k$ percentage point of the sample. That is

$$\hat{F}_i(d_i(k)) = k/10.$$

Clearly $d_i(5)$ is just the median.

We fit λ_i by minimising the following loss function, where $F_i = F_{Y_i}$ is the theoretical cdf of Y_i (which depends on λ_i),

$$L(\lambda_i) = \frac{1}{0.09}(F_i(d_i(1)) - 0.1)^2 + \frac{1}{0.25}(F_i(d_i(5)) - 0.5)^2 + \frac{1}{0.09}(F_i(d_i(9)) - 0.9)^2.$$

There isn't a really good justification for the weights used in this loss function. They were chosen using the widely observed principle that weights should be inversely proportional to variances, and the observation that $Var\hat{F}(x) \propto F(x)(1 - F(x))$.

For month 1 your estimates of p_1, m_1 and λ_1 should be 0.27, 0.24 and 0.043.

Task 3 The next task is to complete the following function, to simulate the operation of a rainwater and a greywater tank in a household of four people.

```
tanksim <- function(num_years, maxraintank, maxgreytank, phat, mhat, lahat, plotflag = F) {
  # Daily simulation of domestic rainwater and greywater tanks
  #
  # num_years is length of simulation in years
  # maxraintank and maxgreytank are max capacity of each tank (in litres)
  # phat, mhat and lahat are parameters for the daily rainfall sim, one for each month
  #
  # Returns volume of water saved (in litres) for each year of the simulation
  # If plotflag is TRUE then the levels of each tank at the end of each day are plotted

  Xsim <- function(month) {
    # simulate rainfall for a day in given month
    rbinom(1, 1, phat[month])*rgamma(1, mhat[month], lahat[month])
  }

  # Constants
  days_in_month <- c(31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31)
  mean_max_temp <- c(25.9, 25.8, 23.9, 20.3, 16.7, 14.1, 13.5, 15.0, 17.2, 19.7, 22.0, 24.2)
  roofarea <- 100                      # m^2
  gardenarea <- 200                    # m^2
  flushsize <- 5                       # litres per flush
  numflush <- function() rbinom(1, 15, .8) # flushes per day for four people, at home half the day
  showersize <- 35                     # in litres
  numshower <- 4                       # showers per day for four people
  washsize <- 35                       # litres per load
  numwash <- function() rbinom(1, 8, .125) # washes per day for four people

  ...
}
```

For this simulation it is sufficient to increment time in units of one day. Each day the state of the simulation is updated as follows:

1. Daily rainfall (in mm) is simulated. The rainwater tank gets a volume of water (in litres) equal to the area of the roof (in square metres) times the depth of rain.

2. Daily greywater (in litres) is simulated. The greywater tank gets a volume of water equal to the number of showers times the shower size, plus the number of laundry washes time the wash size.

3. Water used for flushing toilets is calculated as the number of flushes times the flush size. This water is taken first from the greywater tank, then from the rainwater tank, then from the mains. Record the amount of water saved, that is, water for flushing that comes from the tanks and not the mains.

4. Water for the garden is calculated as follows. In month i we wish the average depth of water given to the garden over the last three days, to be at least the mean maximum temperature for that month divided by 15 (in mm). (So in January we want, over any consecutive three day period, an average of at least $25.9/15 = 1.727$ mm of rain to be given to the garden.) The depth of water given to the garden in the previous two days must include rainfall as well as watering. The volume of water used (in litres) is the required depth of water times the area of the garden (in metres squared). This water is taken first from the rainwater tank, then from the greywater tank, then from the mains. Record the amount of water saved (coming from tanks rather than the mains).

Here is some typical output, using `tanksim(1, 3000, 1000, phat, mhat, lahat, T)`