Exercise 3 Stochastic Models of Manufacturing Systems 4T400, 6 May

1. Each week a very popular lottery in Andorra prints 10^4 tickets. Each tickets has two 4-digit numbers on it, one visible and the other covered. The numbers are randomly distributed over the tickets. If someone, after uncovering the hidden number, finds two identical numbers, he wins a large amount of money. What is the average number of winners per week?

   Answer: Let $X_i = 1$ if ticket i is a winning ticket, and $X_i = 0$ otherwise. Let n denote the number of tickets (in this case $n = 10^4$). Then $P(X_i = 1) = 1 - P(X_i) = \frac{1}{n}$, and thus $E(X_i) = \frac{1}{n}$. Let X be total number of winners, so $X = X_1 + \cdots + X_n$ and

   $$E(X) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = n \cdot \frac{1}{n} = 1.$$

   So the average number of winners does not depend on the size of the lottery.

2. Two people, strangers to one another, both living in Eindhoven, meet each other. Each has approximately 200 acquaintances in Eindhoven. What is the probability of the two people having an acquaintance in common?

   Answer: Let w be the number of acquaintances (so $w = 200$) and r denote the number of inhabitants of Eindhoven minus $w + 2$ (so assuming Eindhoven has 200.000 inhabitants, then $r = 199.798$). Let p be the probability of having at least one acquaintance in common, then

   $$1 - p = \frac{\binom{r}{w}}{\binom{r+w}{w}} = \frac{r(r-1)\cdots(r-w+1)}{(r+w)(r+w-1)\cdots(r+1)}.$$

   So for the specific values of r and w we get $p \approx 0.18$.

3. X is the distance to 0 of a random point in a disk of radius r.

   - What is the density of X?

     Answer: We have

     $$F(x) = P(X \le x) = \frac{\pi x^2}{\pi r^2} = \frac{x^2}{r^2}.$$

     Hence, for the density we get

     $$f(x) = \frac{d}{dx}F(x) = \frac{2x}{r^2}.$$

   - Calculate $E(X)$ and $\text{var}(X)$.

     Answer:

     $$E(X) = \int_0^r x f(x) dx = \int_0^r \frac{2x^2}{r^2} dx = \frac{2}{3}r.$$

     Similarly $E(X^2) = \frac{1}{2}r^2$, so

     $$\text{var}(X) = E(X^2) - (E(X))^2 = \frac{1}{2}r^2 - \frac{4}{9}r^2 = \frac{1}{18}r^2.$$

4. Let X be random on $(0, 1)$ and Y be random on $(0, X)$.

- What is the density of the area of the rectangle with sides X and Y?

  Answer: The joint density of X and Y is given by

  $$f(x,y) = \frac{1}{x}, \quad 0 < y < x < 1.$$

  Let $Z = XY$. Then

  $$\begin{aligned} F(z) &= P(Z \le z) = \int_0^{\sqrt{z}} \int_0^x \frac{1}{x} dy dx + \int_{\sqrt{z}}^1 \int_0^{\frac{z}{x}} \frac{1}{x} dy dx \\ &= 2\sqrt{z} - z. \end{aligned}$$

  Hence

  $$f(z) = \frac{d}{dz} F(z) = \frac{1}{\sqrt{z}} - 1.$$

- Calculate the expected value of the area of the rectangle with sides X and Y.

  Answer:

  $$E(Z) = \int_0^1 z f(z) dz = \int_0^1 z \left(\frac{1}{\sqrt{z}} - 1 \right) dz = \frac{1}{6}.$$

5. A batch consists of n items with probability $(1-p)p^{n-1}, n \ge 1$. The production time of a single item is uniform between 4 and 10 minutes. What is the mean production time of a batch?

   Answer: Let N denote the number in the batch, so

   $$P(N = n) = (1-p)p^{n-1}, \quad n \ge 1,$$

   and

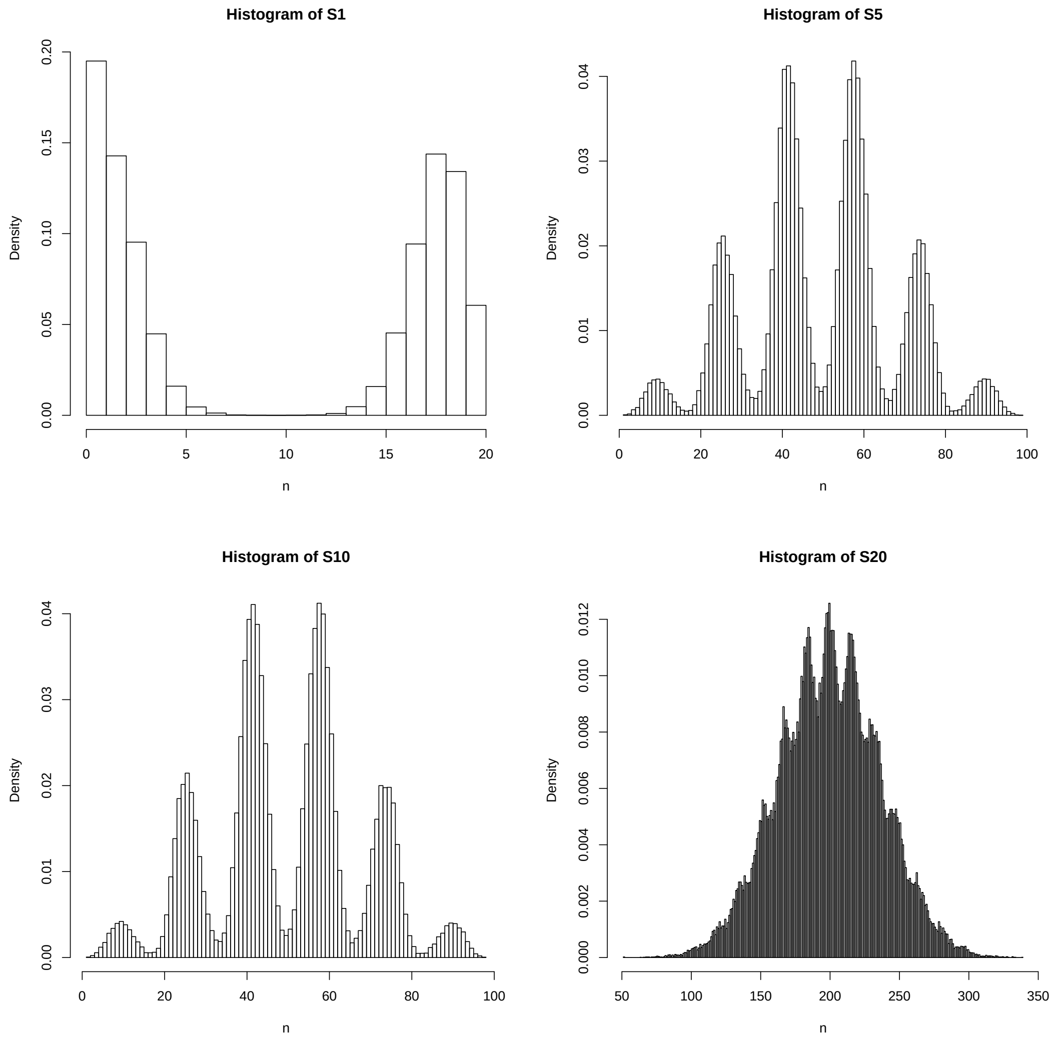
   $$E(N) = \sum_{n=1}^{\infty} n P(N = n) = \sum_{n=1}^{\infty} n(1-p)p^{n-1} = \frac{1}{1-p}.$$

   Let X_i denote the production time of item i in the batch, thus $E(X_i) = \frac{4+10}{2} = 7$ minutes. Further, let X denote the production time of the whole batch, so $X = X_1 + \cdots + X_N$, and

   $$\begin{aligned} E(X) &= E\left(\sum_{i=1}^{N} X_i \right) = \sum_{n=1}^{\infty} E\left(\sum_{i=1}^{N} X_i | N = n \right) P(N = n) \\ &= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^{n} X_i | N = n \right) P(N = n) = \sum_{n=1}^{\infty} E\left(\sum_{i=1}^{n} X_i \right) P(N = n) \\ &= \sum_{n=1}^{\infty} n E(X_1) P(N = n) = E(N)E(X_1) = \frac{7}{1-p} \text{ minutes.} \end{aligned}$$

6. Let X have a mixed binomial distribution: with probability 0.5, the random variable X is binomial with parameters $n = 20$ and $p = 0.1$, and otherwise, X is binomial with parameters $n = 20$ and $p = 0.9$. Let $X_i, i = 1, 2, \ldots$ be independent random variables, with the same distribution as X, and consider the sum $S_n = X_1 + \ldots + X_n$. For each $n = 1, 5, 10, 20$, generate many samples of S_n (for example, by using χ) and plot a histogram (for example in **R**). What is your conclusion?

   Answer: Below the histograms are plotted of S_1, S_5, S_{10} and S_{20}. Each histogram is based on 10^5 samples, generated by a χ program. The conclusion is that, although the mixed binomial distribution does not resemble the Guassian distribution at all, the distribution of the sum of a few mixed binomial random variables already closely matches that of a Gaussian one.



7. Consider a two-machine production line producing fluid. The production rate of machine 1 is 5, the rate of machine 2 is 4. Machine 2 never fails, but machine 1 is subject to breakdowns. The up times and down times have been collected during a long period. The size of the fluid buffer is K.

   - Fit some "simple" distributions to the sample mean and sample variance of the up and down times, and graphically compare the fitted and the empirical distributions.

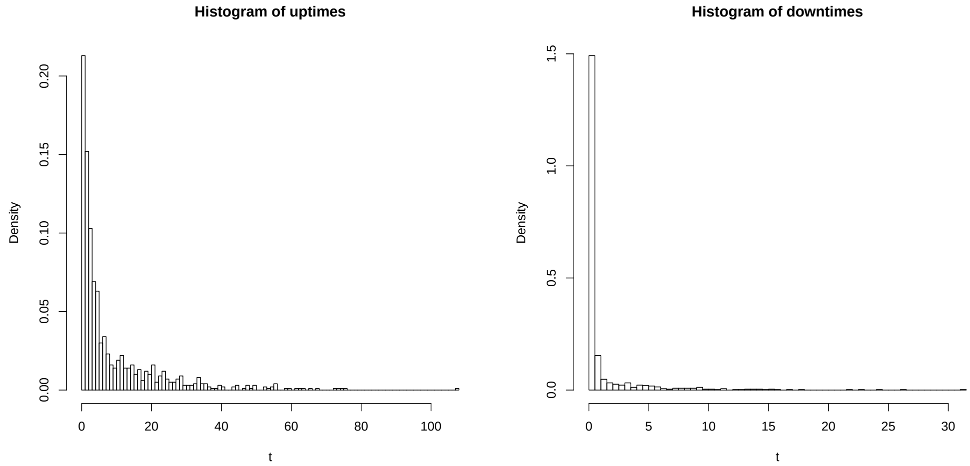
     Answer: Let the random variable U be the uptime, and D the downtime. The sample mean of the uptime $\overline{U}(n)$ with $n = 1000$ samples is equal to

     $$\overline{U}(1000) = \frac{1}{1000}\sum_{i=1}^{1} 000 U_i = 8.91$$

     and the sample variance $S^2(1000)$ is equal to

     $$S^2(1000) = \frac{1}{1000}\sum_{i=1}^{1} 000 (U_i - \overline{U}(1000))^2 = 163$$

Accordingly, the sample mean and sample variance of the 1000 samples of the downtime are $\overline{D}(1000) = 1.17$ and $S^2(1000) = 8.94$. The histograms of the samples of the uptimes (left) and the downtimes (right) are shown below.



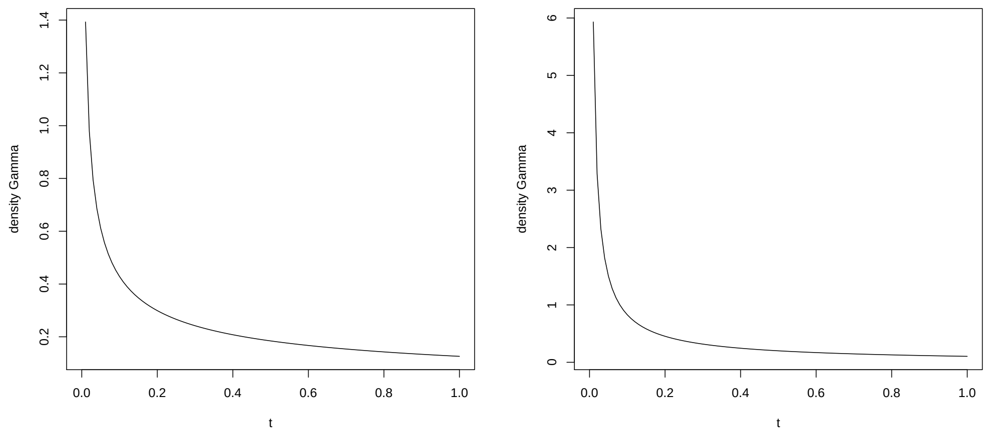
We can fit a Gamma distribution with density

$$f(t) = \frac{1}{\Gamma(a)} \left(\frac{t}{b}\right)^{a-1} \frac{1}{b} e^{-t/b}, \quad t \geq 0,$$

to the mean μ and variance σ^2 by setting

$$a = \frac{\sigma^2}{\mu^2}, \quad b = \frac{\mu}{a}.$$

This means that we get $a = 0.49$ and $b = 18.3$ for the uptimes, and $a = 0.15$ and $b = 7.64$ for the downtimes. The densities of the Gamma distribution fitted to the uptimes (left) and the downtimes (right) are shown below.



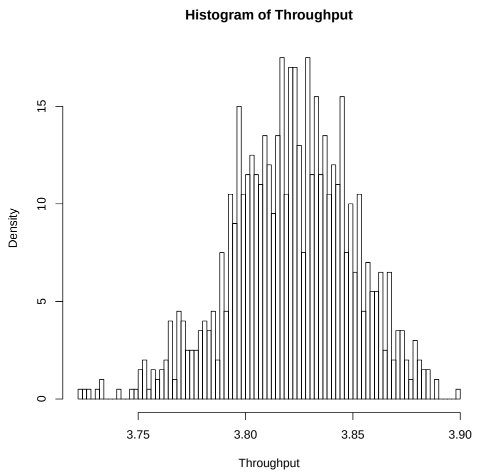
- Estimate, by simulation, the throughput for $K = 20$ over a long simulation horizon T, say $T = 10^4$ time units. Repeat this many (say 10^3) times, and plot a histogram of the estimates for the throughput. What is your conclusion about the distribution of these estimates and how can this be explained?

  Answer: Below we list a histogram of the estimates, showing that it is approximately normal distributed. The explanation is that the throughput is estimated by $4(1 - I/T)$

K	Empirical	Gamma	Exponential
0	3.53	3.54	3.53
1	3.59	3.58	3.62
2	3.62	3.61	3.69
5	3.67	3.67	3.81
10	3.73	3.74	3.90
20	3.81	3.82	3.97

Table 1: Throughput for up- and down times sampled from empirical, Gamma and exponential distributions.

where I is the sum of all idle times of machine 2 during the interval $[0, T]$. Hence, by the central limit theorem, one can expect that $4(1 - I/T)$ is approximately normal distributed as T (and thus the sum of idle times) is large.



- Estimate the long-run throughput of the production line for various values of the buffer size K, by using both the empirical and fitted distributions. What are your conclusions?

  Answer: Below we list the throughput for various values of the buffer size K for up- and down times sampled from their empirical distribution and fitted Gamma distribution; the throughput has been estimated by a simulation run of length 10^6, using a χ model. The results show that both distributions yield similar performance. If an exponential distribution is fitted to the mean only, then the performance is significantly different from the one for the empirical distribution; the throughput is larger, which might be expected, since the variability of the exponential (i.e. coefficient of variation) is less than that of the empirical.