Assignment 2 Solution

1. (a) To verify that M has the same distribution as X, we compute its CDF and compare it to that of X. For $0 \leq x \leq 1$:

$$\begin{aligned} P(M \leq x) &= P(\max(U_1, U_2, U_3) \leq x) = P(U_1 \leq x, U_2 \leq x, U_3 \leq x) \\ &= P(U_1 \leq x)P(U_2 \leq x)P(U_3 \leq x) = [P(U_1 \leq x)]^3 = x^3, \end{aligned}$$

which is the same as the CDF of X. We conclude that M and X have the same distribution.

(b) Inverse Transformation Method: The CDF of X is $F(x) = x^3$ for $0 \leq x \leq 1$, which has inverse, $F^{-1}(x) = x^{1/3}$. If U is distributed Uniform $[0, 1]$, then $F^{-1}(U) = U^{1/3}$ has the same distribution as X.

(c) Acceptance-Rejection Method: The PDF of X is $f(x) = 3x^2$ for $0 \leq x \leq 1$ and zero otherwise, and it achieves its maximum on $x = 1$, at which point the functions value is $L = 3$. The following steps will give us a random variable with the same distribution as X:

i. Generate two independent Uniform $[0, 1]$ random variables U_1, U_2, and set $X = U_1$ (since X takes values on $[0, 1]$).

ii. Accept X if $LU_2 \leq f(X)$ and reject otherwise. This is equivalent to accepting X if $3U_2 \leq 3X^2$, and rejecting otherwise. If X is rejected, go back to step i.

(d) Let k be the number of time units required to generate a uniform random variable.

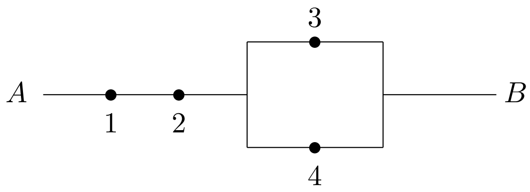
- Method (a): It requires 3 Uniform random variables (and computing the maximum, which well assume to require no computing time), so its computing time is $3k$ units of time.
- Method (b): It requires 1 Uniform random variable and one power operation, so its computing time is $k + 3k = 4k$ units of time.
- Method (c): Each implementation of Steps 1 and 2 require only 2 Uniform random variables, but in order to obtain a valid observation of X we first need to accept. The probability of accepting is

$$\begin{aligned} P(\text{Accept}X) = P(U_2 \leq X^2) = P(U_2 \leq U_1^2) = \int_0^1 P(U_2 \leq u^2 | U_1 = u)\, du \\ \int_0^1 P(U_2 \leq u^2)\, du = \int_0^1 u^2\, du = \frac{1}{3} \end{aligned}$$

So the number of times well have to repeat the two steps is distributed Geometric $(1/3)$, and the expected number of times well have to repeat the two steps is $1/p = 3$. Therefore, the expected number of Uniforms that well need to generate is 6, in which case the expected computing time is $6k$.

We conclude that the most efficient way of generating X is through the method from part (a).

2. Let X be the time at which the system fails (the moment when A gets disconnected from B). Our first goal is to simulate one observation of X.



Let Y_i be the time at which component i fails, $i = 1, 2, 3, 4$. The Y_i's have distribution Weibull with shape parameter $\alpha = 1/4$ and scale parameter $K = 1$, that is,

$$F(x) = P(X \leq x) = 1 - e^{-Kx^{\alpha}}$$

Note that sometimes it can be found written as $F(x) = 1 - e^{-(x/\beta)^{\alpha}}$, which means that in our notation, $K = \beta^{-\alpha}$. The inverse of F is given by

$$F^{-1}(x) = \left(-\frac{1}{K} \log(1 - y) \right)^{\frac{1}{\alpha}}$$

where $\log t$ is the natural logarithm of t. Let U_1, U_2, U_3 and U_4 be independent Uniform(0, 1) random variables; we can simulate Y_1, Y_2, Y_3 and Y_4 by setting

$$Y_1 = F^{-1}(U_1), \quad Y_2 = F^{-1}(U_2), \quad Y_3 = F^{-1}(U_3), \quad Y_4 = F^{-1}(U_4)$$

To compute X note that if $Y_1 = \min\{Y_1, Y_2, Y_3, Y_4\}$, then $X = Y_1$. Similarly, if $Y_2 = \min\{Y_1, Y_2, Y_3, Y_4\}$, then $X = Y_2$. Nonetheless, if Y_3 is the minimum, then at time Y_3 the system has not failed yet, so we need to check who the minimum of $\{Y_1, Y_2, Y_4\}$ is, and set $X = \min\{Y_1, Y_2, Y_4\}$. The same thing happens if $Y_4 = \min\{Y_1, Y_2, Y_3, Y_4\}$, in which case we set $X = \min\{Y_1, Y_2, Y_3\}$. The above information can be summarized as

$$X = \min\{Y_1, Y_2, \max\{Y_3, Y_4\}\}$$

The second part of the problem is to find an estimator for $\mu = E[X]$. Our natural choice is $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$, where the X_i's are independent copies having the same distribution as X (that is, we need to repeat the procedure described above n times). To find a 99% confidence interval for μ, we use the CLT approximation for $S_n = X_1 + \cdots + X_n$:

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \overset{D}{\approx} N(0, 1)$$

so for any $z > 0$,

$$\begin{aligned}
P(|N(0,1)| \leq z) &\approx P\left(\left| \frac{S_n - n\mu}{\sqrt{n}\sigma} \right| \leq z \right) \\
&= P\left(\left| \frac{S_n - n\mu}{n} \right| \leq \frac{\sigma z}{\sqrt{n}} \right) \\
&= P\left(\left| \frac{S_n}{n} - \mu \right| \leq \frac{\sigma z}{\sqrt{n}} \right) \\
&= P\left(|\hat{\mu} - \mu| \leq \frac{\sigma z}{\sqrt{n}} \right) \\
&= P\left(\hat{\mu} - \frac{\sigma z}{\sqrt{n}} \leq \mu \leq \hat{\mu} + \frac{\sigma z}{\sqrt{n}} \right)
\end{aligned}$$

Since we want,

$$0.99 = P\left(\hat{\mu} - \frac{\sigma z}{\sqrt{n}} \leq \mu \leq \hat{\mu} + \frac{\sigma z}{\sqrt{n}}\right)$$

we need to find z such that

$$0.99 = P(|N(0,1)| \leq z) = P(-z \leq N(0,1) \leq z) = \Phi(z) - \Phi(-z) = 2\Phi(z) - 1$$

This gives us

$$z = \Phi^{-1}\left(\frac{1.99}{2}\right) = \Phi^{-1}(0.995) = 2.5758$$

Since we don't know σ, we estimate it from our data with the formula:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

The solution below was obtained with $n = 10,000$:

$$\hat{\mu} = 0.4766, \quad \hat{\sigma}^2 = 8.3629, \quad 99\% \text{ C.I.} = (0.4021,\ 0.5511)$$

3. Suppose the lifetime X of a new product has distribution Exp(λ), i.e. X has PDF $f(x;\lambda) = \lambda e^{-\lambda x}$ and CDF $F(x;\lambda) = 1 - e^{-\lambda x}$. We'll apply maximum likelihood estimation (MLE) method to estimate the "best" parameter λ.

   Let $x_1 = 0.9$, $x_2 = 1.2$, $x_3 = 2.9$ and $T = 3$. The likelihood function is then given by

   $$L(\lambda) = f(x_1;\lambda)f(x_2;\lambda)f(x_3;\lambda)(1 - F(T;\lambda))^{47} = \lambda^3 e^{-\lambda(x_1+x_2+x_3+47T)}.$$

   The log-likelihood function is then

   $$g(\lambda) \triangleq \log(L(\lambda)) = 3\log(\lambda) - \lambda(x_1 + x_2 + x_3 + 47T).$$

   Setting $g'(\lambda) = 0$ yields the maximizer of $g(\lambda)$, i.e.

   $$\frac{3}{\lambda^*} - (x_1 + x_2 + x_3 + 47T) = 0$$

   so $\lambda^* = 3/(x_1 + x_2 + x_3 + 47T) \approx 0.0205$. The probability that a new product fails within one year is then

   $$p \triangleq P(X \leq 12) = 1 - e^{-12\lambda^*} \approx 0.2181.$$

   (a) The average cost per item of honoring the warranty is \$20$p \approx$ \$4.3620.

   (b) That the total warranty cost will be more than \$100000 is equivalent to that there will be more than 5000 products failing within one year. Let I_i denote whether ith product fails within one year or not, i.e. $I_i = 1$ if it fails and 0 otherwise. Then I_i has Bernoulli distribution and $P(I_i = 1) = p = 1 - P(I_i = 0)$. Moreover,, I_i's are independent and has common mean p and variance $\sigma^2 \triangleq p(1-p)$. Put $n = 300000$.

   $$\begin{aligned}
   &P(\text{More than 5000 products out of 300000 fail within one year}) \\
   &= P\left(\sum_{i=1}^{n} I_i \geq 5000\right) \approx P\left(np + \sqrt{n}\sigma\mathcal{N}(0,1) \geq 5000\right) \\
   &\approx P(\mathcal{N}(0,1) \geq -267.17) \approx 1
   \end{aligned}$$

(c) Suppose the expected cost per item of honoring the warranty is C. Then, we have the following recursion:

$$C = 20 + pC$$

so $C = 20/(1-p) \approx 25.58$.

4. Let Y_k be China's national income in year $1952+k$, $k = 0, 1, \ldots, 36$. The model we are trying to fit is

$$Y_k = ak + b$$

We use least squares to estimate a and b. By Handout " Forecasting and Regression Models", the estimators $\hat{a}$ and $\hat{b}$ solve:

$$\begin{pmatrix} \sum_{k=0}^{n} k^2 & \sum_{k=0}^{n} k \\ \sum_{k=0}^{n} k & n+1 \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \sum_{k=0}^{n} kY_k \\ \sum_{k=0}^{n} Y_k \end{pmatrix}$$

where $n = 36$. For reference we give below the values of each of the entries based on the data:

$$\sum_{k=0}^{n} k^2 = 16206, \quad \sum_{k=0}^{n} k = 666, \quad \sum_{k=0}^{n} kY_k = 1244684.2, \quad \sum_{k=0}^{n} Y_k = 46036.5$$

The estimators we obtain are

$$\hat{a} = 98.631389284021, \qquad \hat{b} = -531.135277382646$$

Our prediction for China's industry income in 1992 is then:

$$Y_{n+4} = \hat{a}(n+4) + \hat{b} = 40\hat{a} + \hat{b} = 3414.120293978189$$