
1: Inverse method for Poisson Distribution (25%)

For discrete Poisson Distribution ($\lambda = 5$),

the p.m.f is $P(x|\lambda) = e^{-\lambda} \frac{\lambda^x}{x!}$ and the c.d.f is $F(x|\lambda) = \sum_{t \leq x} e^{-\lambda} \frac{\lambda^t}{t!}$.

Algorithm: Inverse method for the Poisson Distribution:

To generate $X \sim F(x)$:

STEP 1: Generate $U \sim \text{unif}[0, 1]$;

STEP 2: Transform $X = F^{-1}(U)$: if $F(x|\lambda) < U \leq F(x+1|\lambda)$, let $X = x+1$.

Plot :

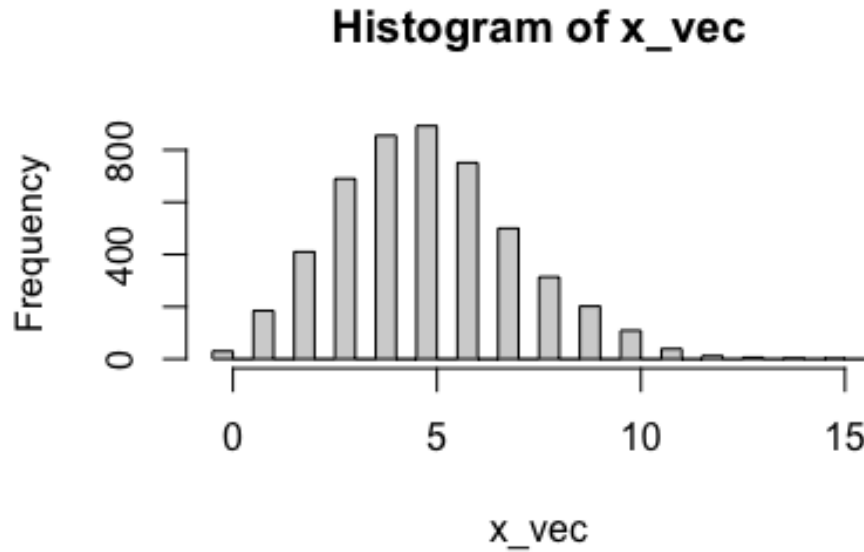


Figure 1: Histogram of 5000 samples

2: Accept-Reject method for truncated Gamma Distribution (25%)

For $X \sim \text{Gamma}(\frac{1}{2}, 1)I(x \geq 5)$, $f(x) = \frac{x^{-\frac{1}{2}} e^{-x} I(x \geq 5)}{\int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy}$.

We can define a shifted exponential distribution $g(x) = e^{-(x-5)} I(x \geq 5)$ and want to find a constant M such that $f(x) < Mg(x)$ for any x .

Then $M = \sup \frac{f(x)}{g(x)} = \sup \frac{\frac{x^{-\frac{1}{2}} e^{-x} I(x \geq 5)}{\int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy}}{e^{-(x-5)} I(x \geq 5)} = \frac{5^{-\frac{1}{2}} e^{-5}}{\int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy}$.

Algorithm: Accept-Reject method for truncated Gamma Distribution:

To generate $X \sim F(x) = \text{c.d.f of } f(x)$:

STEP 1: Generate $Y \sim g(y)$;

STEP 2: Generate $U \sim \text{unif}[0, 1]$;

STEP 3: Accept $X = Y$ if $U \leq \frac{f(Y)}{Mg(Y)}$.

Proof :

From the choice of constant M , we can know that $Mg(x) \geq f(x)$. The goal of this method is to generate $X \sim F(x) = \text{c.d.f of } f(x)$.

For the generating algorithm:

$$\begin{aligned}
 P(X \leq x) &= P(Y \leq x | Y \text{ is accepted}) \\
 &= P(Y \leq x | U \leq \frac{f(Y)}{Mg(Y)}) \\
 &= \frac{P(Y \leq x, U \leq \frac{f(Y)}{Mg(Y)})}{P(U \leq \frac{f(Y)}{Mg(Y)})} \\
 &= \frac{\int_{-\infty}^x g(y) \int_0^{\frac{f(y)}{Mg(y)}} 1 dudy}{\int_{-\infty}^{+\infty} g(y) \int_0^{\frac{f(y)}{Mg(y)}} 1 dudy} \\
 &= \frac{\int_{-\infty}^x g(y) \frac{f(y)}{Mg(y)} dy}{\int_{-\infty}^{+\infty} g(y) \frac{f(y)}{Mg(y)} dy} \\
 &= \frac{\int_{-\infty}^x f(y) dy}{\int_{-\infty}^{+\infty} f(y) dy} \\
 &= \frac{\int_{-\infty}^x f(y) dy}{1} \\
 &= F(x)
 \end{aligned}$$

Therefore, this AR method works.

Comparison :

Theoretical acceptance probability:

$$\begin{aligned}
 P(U \leq \frac{f(Y)}{Mg(Y)}) &= \int_{-\infty}^{+\infty} g(y) \int_0^{\frac{f(y)}{Mg(y)}} 1 dudy \\
 &= \frac{1}{M} \int_{-\infty}^{+\infty} f(y) dy \\
 &= \frac{1}{M}
 \end{aligned}$$

After computation, this acceptance probability is 0.9207851.

The actual acceptance rate is 0.918, which is a little bit lower than the theoretical value.

3: Importance Sampling for Estimation (25%)

(1)

Using 5000 samples from Q2 (l = length of samples obtained in Q2), the Monte Carlo estimate is

$$\begin{aligned}
 \int_5^{+\infty} \cos(x) x^{-\frac{1}{2}} e^{-x} dx &= \int_{-\infty}^{+\infty} \cos(x) \frac{x^{-\frac{1}{2}} e^{-x} I(x \geq 5)}{\int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy} dx \times \int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy \\
 &= \frac{\int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy}{l} \sum_{i=1}^l \cos(x_i) \\
 &= 0.001708
 \end{aligned}$$

(2)

Using the same notations in Q2, we define $h(x) = \cos(x)$.

$$\begin{aligned}
 \int_5^{+\infty} \cos(x) x^{-\frac{1}{2}} e^{-x} dx &= \int_{-\infty}^{+\infty} \cos(x) \frac{x^{-\frac{1}{2}} e^{-x} I(x \geq 5)}{\int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy} dx \times \int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy \\
 &= \int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy \int_{-\infty}^{+\infty} \frac{h(x)f(x)}{g(x)} dx
 \end{aligned}$$

Note that $\frac{f(x)}{g(x)} \leq M < \infty$, and $E_g h^2(x) = \int_{-\infty}^{+\infty} g(x) h^2(x) dx \leq \int_{-\infty}^{+\infty} g(x) dx = 1 < \infty$, we can use the importance sampling as follows:

Algorithm:

STEP 1: Generate $n = 5000$ samples from $g(x)$;

STEP 2: Compute the Monte Carlo estimate:

$$\int_{-\infty}^{+\infty} \frac{h(x)f(x)}{g(x)} dx = \frac{\sum_{i=1}^n \frac{h(x_i)f(x_i)}{g(x_i)}}{n} = \frac{e^{-5}}{n \int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy} \sum_{i=1}^n \cos(x_i) x_i^{-\frac{1}{2}}$$

Therefore,

$$\begin{aligned}
 \int_5^{+\infty} \cos(x) x^{-\frac{1}{2}} e^{-x} dx &= \int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy \int_{-\infty}^{+\infty} \frac{h(x)f(x)}{g(x)} dx \\
 &= \int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy \frac{e^{-5}}{n \int_5^{+\infty} y^{-\frac{1}{2}} e^{-y} dy} \sum_{i=1}^n \cos(x_i) x_i^{-\frac{1}{2}} \\
 &= \frac{e^{-5}}{n} \sum_{i=1}^n \cos(x_i) x_i^{-\frac{1}{2}} \\
 &= 0.00174
 \end{aligned}$$

4: Stratified Sampling (25%)

(1)

Randomly draw 100 samples from the date set, for each subpopulation:

Standard deviation for age interval 1 is 73.9823615320924;

Standard deviation for age interval 2 is 111.975202857371;

Standard deviation for age interval 3 is 204.753943423479.

(2)

The target of this sampling is to estimate the mean salary of this country.

Using Stratified Sampling, there're 3 strata in total, which are indicated by 1, 2, and 3 respectively. Define n_i = the sample number of the i strata, x_{ij} = the sampled j -th individual in strata i , μ_i = the i strata's proportion in population, and S_i = the strata i .

The target is

$$E(\vec{x}) = \sum_{i=1}^3 p(x_i \in S_i) E(x|x \in S_i) = \sum_{i=1}^3 \mu_i \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}.$$

It is supposed that the n_i 's can minimize $Var(E(\vec{x}))$.

$$Var(E(\vec{x})) = \sum_{i=1}^3 \mu_i^2 Var\left(\frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}\right) = \sum_{i=1}^3 \mu_i^2 \frac{1}{n_i} Var(x|x \in S_i)$$

Denote $Var(x|x \in S_i)$ by V_i , which can be estimated by variance we obtained in the last step. Then the minimization problem can be modeled as:

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^3 \frac{\mu_i^2 V_i}{n_i} \\ & \text{subject to} && \sum_{i=1}^3 n_i = 1000 \end{aligned}$$

By Lagrange Multiplier, $n_i = \frac{1000\mu_i\sqrt{V_i}}{\sum_{k=1}^3 \mu_k\sqrt{V_k}}$. We use standard deviations estimated in part (1) to compute variance here.

After computation,

$$n_1 = 68$$

$$n_2 = 307$$

$$n_3 = 625$$

(3)

By randomly drawing n_i samples in strata i for each i , the estimated population mean salary based on these 1000 samples is 4607.275, which is a little bit higher than the true mean salary 4307.189 but is very close. The Stratified Sampling can somehow reflect the real mean salary.