Simulating Discrete Random Variables with MATLAB

Evgenii Samutichev

December 4, 2023

Sampling from a binomial distribution

Part 1 We have the (pseudo)random number generator for the uniform distribution on $[0,1]$, which is exactly the MATLAB `rand` function. Denote it as $R \sim \text{Unif}(0,1)$. Then, we can transform it to get the random number generator for $X \sim \text{Bern}(p)$ as follows

$$X = \mathbb{1}_{R \leq p} = \begin{cases} 1, & R \leq p \\ 0, & R > p \end{cases} \tag{1}$$

Then

$$\mathbb{P}(X = 1) = \mathbb{P}(R \in [0,p]) = p \tag{2}$$

and

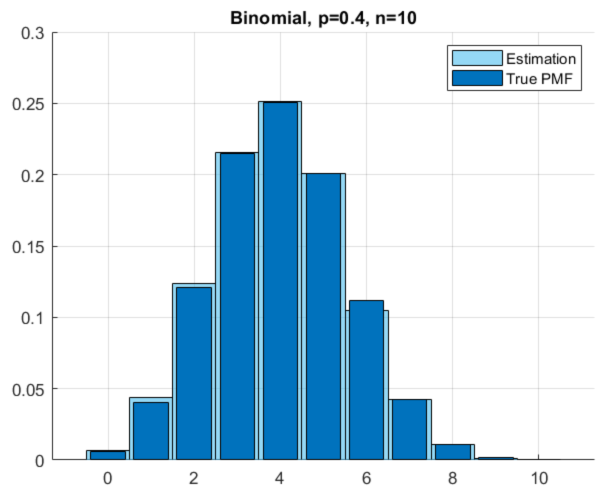
$$\mathbb{P}(X = 0) = \mathbb{P}(R \in (p,1]) = 1 - p \tag{3}$$

which proves that $X \sim \text{Bern}(p)$. **Code:** `bernoulli.m`

Part 2 By fixing n, we can also build the random number generator for $Y \sim \text{Binom}(n,p)$ as

$$Y = X_1 + ... + X_n \tag{4}$$

using the definition of Binomial distribution and where $X_i \sim \text{Bern}(p)$. So it's a sum of n samples from X. **Code:** `binomial.m`



Sampling from a Poisson distribution

Part 1 Let $X \sim \mathrm{Unif}(0,1)$, then for $Z = -\log X$ we have $Z \geq 0$ and so for $z \geq 0$

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) = \mathbb{P}(-\log X \leq z) = \mathbb{P}(e^{-\log X} \leq e^z) = \\ &= \mathbb{P}\left(\frac{1}{X} \leq e^z\right) = \mathbb{P}(X \geq e^{-z}) = 1 - e^{-z} \end{aligned} \tag{1}$$

where we can ignore case $X = 0$, since it has zero probability. Including $z < 0$ we get

$$F_Z(z) = \begin{cases} 1 - e^{-z}, & z \geq 0 \\ 0, & z < 0 \end{cases} \tag{2}$$

By taking the derivative of F_Z, we get

$$f_Z(z) = \frac{dF_Z}{dz}(z) = e^{-z} \tag{3}$$

for $z \geq 0$ and 0 otherwise, which is a density function of Exp(1), so $Z \sim \mathrm{Exp}(1)$.
Part 2 Now, let's look at the finite sums of i.i.d random variables $Z_i \sim \mathrm{Exp}(1)$. Using the convolution formula for the sum $S_2 = Z_1 + Z_2 \geq 0$, we get

$$f_{S_2}(z) = \int_{-\infty}^{\infty} f_Z(x) f_Z(z-x) dx = \int_0^z e^{-x} e^{-z+x} dx = \int_0^z e^{-z} dx = e^{-z} z \tag{4}$$

For $S_3 = Z_1 + Z_2 + Z_3 = S_2 + Z_3 \geq 0$ we get

$$\begin{aligned} f_{S_3}(z) &= \int_{-\infty}^{\infty} f_{S_2}(x) f_Z(z-x) dx = \int_0^z e^{-x} x e^{-z+x} dx = \\ &= e^{-z} \int_0^z x dx = \frac{e^{-z} z^2}{2} \end{aligned} \tag{5}$$

It's now tempting to assume that the general form is

$$f_{S_k}(z) = \frac{e^{-z} z^{k-1}}{(k-1)!} \tag{6}$$

let's prove this by induction. If it's true for k, then for $k+1$

$$\begin{aligned} f_{S_{k+1}}(z) &= \int_{-\infty}^{+\infty} f_{S_k}(x) f_Z(z-x) dx = \int_0^z e^{-x} \frac{x^{k-1}}{(k-1)!} e^{-z+x} dx = \\ &= \frac{e^{-z}}{(k-1)!} \int_0^z x^{k-1} dx = \frac{e^{-z}}{(k-1)!} \frac{z^k}{k} = \frac{e^{-z} z^k}{k!} \end{aligned} \tag{7}$$

so by induction this is true for all k and $S_k \sim \mathrm{Gamma}(k, 1)$
Part 3 Let $\lambda > 0$ and let $(X_n)_{n \geq 0}$ be a sequence of i.i.d as $X \sim \mathrm{Unif}(0,1)$ r.v.s. Then let Y be a random variable defined as

$$Y = \min\{n \geq 0 | X_0 \cdot \ldots \cdot X_n \leq e^{-\lambda}\} \tag{8}$$

Notice that the event

$$X_0 \cdot \ldots \cdot X_n \leq e^{-\lambda} \tag{9}$$

is equivalent to

$$\sum_{k=0}^{n} \log X_k = \log(X_0 \cdot ... \cdot X_n) \le \log\left(e^{-\lambda}\right) = -\lambda \tag{10}$$

and event $(X_0 = 0) \cup (X_1 = 0) \cup ... \cup (X_n = 0)$ has probability zero, so we can ignore it. Multiplying both sides by -1 we get

$$\sum_{k=0}^{n} -\log X_k = \sum_{k=0}^{n} Z_k = S_{k+1} \ge \lambda \tag{11}$$

where we used result of **Part 1** of this problem. Then

$$(Y = 0) = (S_1 \ge \lambda) \tag{12}$$

and for $n > 0$

$$(Y = n) = (S_{n+1} \ge \lambda) \setminus (S_n \ge \lambda) \tag{13}$$

From **Part 2** we get

$$\mathbb{P}(S_{n+1} \ge \lambda) = \int_{\lambda}^{\infty} \frac{e^{-x} x^n}{n!} dx \tag{14}$$

We can now compute the PMF of Y. First, notice that

$$\mathbb{P}(Y = 0) = \int_{\lambda}^{\infty} e^{-x} dx = -e^{-x}\Big|_{\lambda}^{\infty} = e^{-\lambda} = \frac{\lambda^0}{0!} e^{-\lambda} \tag{15}$$

so the formula holds for $n = 0$. For $n > 0$

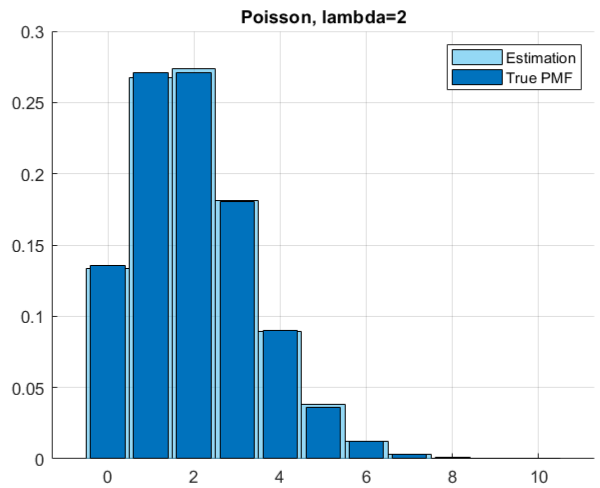
$$\begin{aligned} \mathbb{P}(S_{n+1} \ge \lambda) &= \int_{\lambda}^{\infty} \frac{e^{-x} x^n}{n!} dx = -e^{-x} \frac{x^n}{n!}\Big|_{\lambda}^{\infty} + \int_{\lambda}^{\infty} \frac{e^{-x} x^{n-1}}{(n-1)!} dx = \\ &= e^{-\lambda} \frac{\lambda^n}{n!} + \mathbb{P}(S_n \ge \lambda) \end{aligned} \tag{16}$$

then using (13)

$$\mathbb{P}(Y = n) = \mathbb{P}(S_{n+1} \ge \lambda) - \mathbb{P}(S_n \ge \lambda) = e^{-\lambda} \frac{\lambda^n}{n!} \tag{17}$$

So indeed $Y \sim \text{Pois}(\lambda)$.

Part 4 For the algorithm see `poisson.m`, it's vectorized to avoid for loops as much as possible.



The inversion method

Part 1 Let $P = (p_0, p_1, ...,)$ be a probability distribution on $\mathbb{N}$. Let $H : \mathbb{R} \to [0, 1]$ be a function defined as

$$\forall k \geq 0, \forall x \in [k, k+1) \text{ then } H(x) = p_0 + p_1 + ... + p_k \tag{1}$$

and let $U \sim \text{Unif}(0, 1)$. Define X as

$$X = \inf\{k \in \mathbb{N} \mid U \leq H(k)\} \tag{2}$$

then

$$\mathbb{P}(X = 0) = \mathbb{P}(U \leq p_0) = p_0 \tag{3}$$

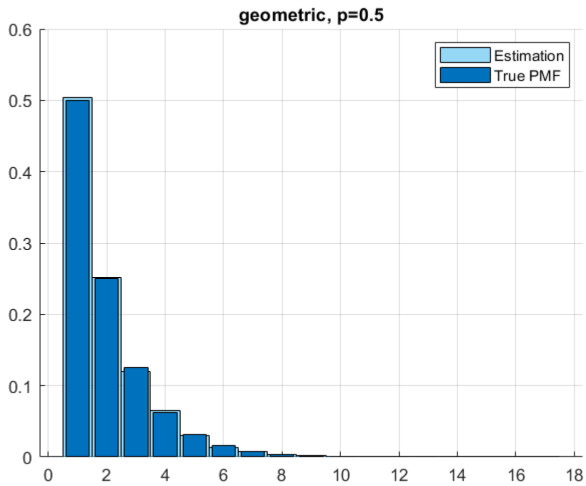
and notice that for $n > 0$

$$(X = n) = (U \leq H(n)) \setminus (U \leq H(n-1)) \tag{4}$$

since $H(x)$ is a nondecreasing function, so if n is an infimum, then for smaller value as $n - 1$ the inequality $U \leq H(n-1)$ can't hold. Then

$$\mathbb{P}(X = n) = \mathbb{P}(U \leq H(n)) - \mathbb{P}(U \leq H(n-1)) = \sum_{k=0}^{n} p_k - \sum_{k=0}^{n-1} p_k = p_n \tag{5}$$

Part 2 The result of **Part 1** allows us to sample from any discrete distribution. In particular, let's sample from geometric distribution with parameter p. For an algorithm see `geometric.m`.


geometric, p=0.5
Estimation
True PMF
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
18