Generating random variables I

Anatoli Iambartsev
IME-USP

Distribution of Random Variables.

Cumulative distribution function uniquely defines a random variable X:

$$F(x) = \mathbb{P}(X \leq x), \quad F : \mathbb{R} \to [0, 1];$$

1. it is non-negative and non-decreasing function with values from $[0, 1]$;

2. it is right-continuous and it has limit from the left (càdlàg function);

3. $\lim_{x \to +\infty} F(x) = 1; \lim_{x \to -\infty} F(x) = 0;$

4. $\mathbb{P}(a < X \leq b) = F(b) - F(a)$.

(Purely) Discrete random variable.

"1 kg" of probability is concentrated on finite or countable set of points $\mathcal{S}(X) = \{x_1, x_2, \dots\}$: for any $x_0 \in \mathbb{R}$

$$\mathbb{P}(X = x_0) = F(x_0) - \lim_{x \to x_0 -} F(x)$$

and for any $x_i \in \mathcal{S}(X)$ the probability $\mathbb{P}(X = x_i) > 0$.

$F(x)$ is discontinuous at the points $x_i \in \mathcal{S}(X)$ and constant in between:

$$F(x) = \sum_{x_i \le x} \mathbb{P}(X = x_i) = \sum_{x_i \le x} p(x_i)$$

$$\mathbb{P}(a < X \le b) = \sum_{a < x_i \le b} p(x_i)$$

(Purely) Discrete random variable.

- Bernoulli r.v. $X \sim B(p)$; Binomial r.v. $X \sim B(n,p)$;
- Poisson r.v. $X \sim Poi(\lambda)$;
- Geometric r.v. $X \sim G(p)$:

$$\mathbb{P}(X=k) = (1-p)^{k-1}p, k = 1,2,\ldots, \mathbb{E}(X) = \frac{1}{p},$$

$$\mathbb{P}(X=k) = (1-p)^{k}p, k = 0,1,2,\ldots, \mathbb{E}(X) = \frac{1-p}{p},$$

$$\mathbb{V}ar(X) = \frac{1-p}{p^2}$$

- ect.

Absolute continuous random variable.

If $F(x)$ is absolutely continuous, i.e. there exists a Lebesgue-integrable function $f(x)$ such that

$$F(b) - F(a) = \mathbb{P}(a < X \leq b) = \int_a^b f(x)dx,$$

for all real a and b. The function f is equal to the derivative of F almost everywhere, and it is called the probability density function of the distribution of random variable X.

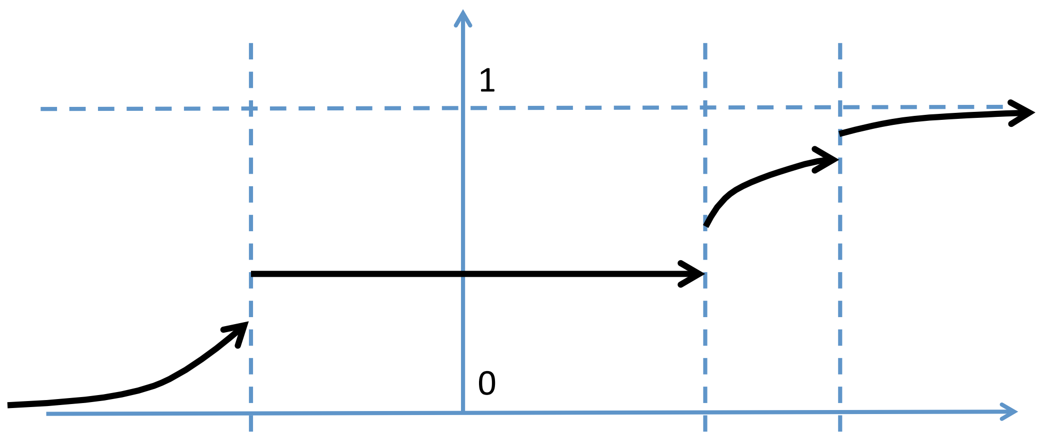
Absolute continuous random variable.

- Unifrom r.v. $X \sim U[a, b]$;
- Normal r.v. $X \sim N(\mu, \sigma^2)$;
- Exponential r.v. $X \sim Exp(\lambda)$;
- ect.

Mixed (absolute continuous and discrete) random variable.

Let X be a discrete r.v. with F_X distribution function, and let Y be absolute continuous r.v. with F_Y distribution function. Let $p \in (0,1)$ then in the course we also consider mixed random variable Z with following distribution function F_Z

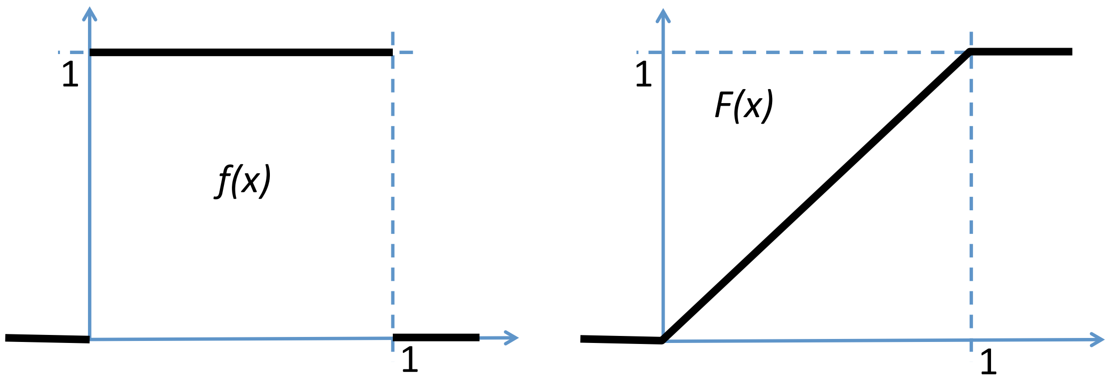
$$F_Z = pF_X + (1-p)F_Y.$$



Uniform random variable.

$X \sim U[0,1]$, with cumulative distribution function $F(x)$ and density function $f(x)$

$$F(x) = \mathbb{P}(X \le x) = \begin{cases} 0, & \text{if } x \le 0; \\ x, & \text{if } 0 < x \le 1; \\ 1, & \text{if } x > 1; \end{cases} \quad f(x) = \begin{cases} 0, & \text{if } x \notin [0,1]; \\ 1, & \text{if } x \in [0,1]. \end{cases}$$



Uniform random variable simulation.

> runif (100, min=3, max=4)

will generate 100 numbers from the interval $[3, 4]$ according uniform distribution.

[RC]: Strictly speaking, all the methods we will see (and this includes **runif**) produce *pseudo-random numbers* in that there is no randomness involved − based on an initial value u_0 and a transformation D, the uniform generator produce a sequence $(u_i, i = 0, 1, \ldots)$, where $u_i = D^i(u_0)$ of values on $(0, 1)$ − but the outcome has the same *statistical properties* as an iid sequence. Further details on the random generator of R are provided in the on-line help on RNG.

Uniform simulation.

```
> set.seed(2)
> runif(5)
[1] 0.1848823 0.7023740 0.5733263 0.1680519 0.9438393
> set.seed(1)
> runif(5)
[1] 0.2655087 0.3721239 0.5728534 0.9082078 0.2016819
> set.seed(2)
> runif(5)
[1] 0.1848823 0.7023740 0.5733263 0.1680519 0.9438393
```

Non-uniform random variable generation.

- The inverse transform method
- Accept-reject method
- others

The inverse transform method.

[RC] ... it is known also as *probability integral transform* – it allows us to transform any random variable into a uniform random variable and, more importantly, vice versa. For example, if X has density f and cdf F, then we have the relation

$$F(x) = \int_{-\infty}^{x} f(t)dt,$$

and if we set $U = F(X)$, then $U \sim U[0,1]$, indeed,

$$\begin{aligned}\mathbb{P}(U \leq u) &= \mathbb{P}(F(X) \leq F(x)) \\ &= \mathbb{P}(F^{-1}(F(X)) \leq F^{-1}(F(x))) = \mathbb{P}(X \leq x),\end{aligned}$$

here F has an inverse because it is monotone.

The inverse transform method.

Example. [RC] If $X \sim Exp(\lambda)$, then $F(x) = 1 - e^{-\lambda x}$. Solving for x in $u = 1 - e^{-\lambda x}$ gives $x = -\frac{1}{\lambda}\ln(1-u)$. Therefore, if $U \sim U[0,1]$, then

$$X = -\frac{1}{\lambda}\ln(U) \sim Exp(\lambda).$$

(as U and $1-U$ are both uniform)

The inverse transform method.

Example. [RC, Exercise 2.2] Some variables that have explicit forms of the cdf are logistic and Cauchy. Thus, they are well-suited to the inverse transform method.

- Logistic: $f(x) = \frac{1}{\beta}\frac{e^{-(x-\mu)/\beta}}{(1+e^{-(x-\mu)/\beta})^2}$, $F(x) = \frac{1}{1+e^{-(x-\mu)/\beta}}$;
- Cauchy: $f(x) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2}$, $F(x) = \frac{1}{2}+\frac{1}{\pi}\text{arctan}\left(\frac{x-\mu}{\sigma}\right)$;
- Pareto(γ): $f(x) = \frac{\gamma}{(1+x)^{\gamma+1}}$, $F(x) = 1\ldots$

The inverse transform method.

For an arbitrary random variable X with cdf F, define the generalized inverse of F by

$$F^{-1}(u) = \inf\{x : \ F(x) \geq u\}.$$

Lemma 1. Let $U \sim U[0,1]$, then $F^{-1}(U) \sim F$.

Proof. For any $u \in [0,1]$, $x \in F^{-1}([0,1])$ we have

$$F(F^{-1}(u)) \geq u, \ F^{-1}(F(x)) \geq x.$$

Therefore

$$\{(u,x) : \ F^{-1}(u) \leq x\} = \{(u,x) : \ F(x) \geq u\},$$

and it means

$$\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(F(x) \geq U) = F(x).$$

The inverse transform method.

Example. Let $X \sim B(p)$.

$$F(x) = \begin{cases} 0, & \text{if } x < 0; \\ 1-p, & \text{if } 0 \leq x < 1; \\ 1, & \text{if } 1 \leq x < \infty, \end{cases} \quad F^{-1}(x) = \begin{cases} 0, & \text{if } 0 \leq u \leq 1-p; \\ 1, & \text{if } p < u \leq 1, \end{cases}$$

Thus

$$X = F^{-1}(U) = \begin{cases} 0, & \text{if } U \leq 1-p; \\ 1, & \text{if } U > 1-p, \end{cases} \sim B(p).$$

The inverse transform method.

A main limitation of inverse method is that quite often F^{-1} is not available in explicit form. Sometimes approximations are used.

The inverse transform method.

Example. When $F = \Phi$, the standard normal cdf, the following rational polynomial approximation is standard, simple, and accurate for the normal distribution:

$$\Phi^{-1}(u) = y + \frac{p_0 + p_1 y + p_2 y^2 + p_3 y^3 p_4 y^4}{q_0 + q_1 y + q_2 y^2 + q_3 y^3 q_4 y^4}, 0.5 < u < 1,$$

where $y = \sqrt{-2 \ln(1-u)}$ and the p_k, q_k are given by table:

k	p_k	q_k
0	-0.322232431088	0.099348462606
1	-1	0.588581570495
2	-0.3422420885447	0.531103462366
3	-0.0204231210245	0.10353775285
4	0.0000453642210148	0.0038560700634

Accept-reject method.

[RC, Ch.2.3] When the inverse method will fail, we must turn to *indirect* methods; that is, methods in which we generate a candidate random variable and only accept it subject to passing a test... These so-called *accept-reject methods* only require us to know the functional form of the density f of interest (called the *target density*) up to a multiplicative constant. We use a simpler (to simulate) density g, called the *instrumental or candidate density*, to generate the random variable for which the simulation is actually done.

Accept-reject method.

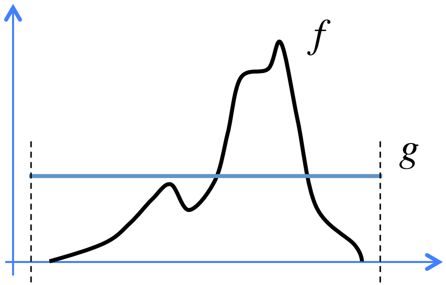
[RC, Ch.2.3] Constraints:

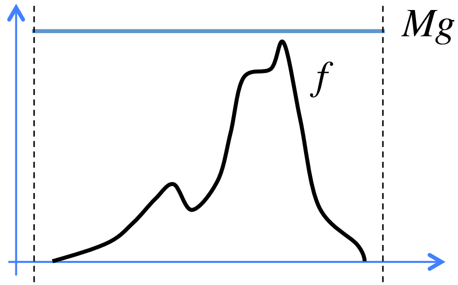
- f and g have compatible supports (i.e. $g(x) > 0$ when $f(x) > 0$);
- there is a constant M with $\frac{f(x)}{g(x)} \leq M$ for all x.

$X \sim f$ is simulated as follows.

- generate $Y \sim g$;
- generate $U \sim U[0,1]$;
- if $U \leq \frac{1}{M}\frac{f(Y)}{g(Y)}$, then we set $X = Y$, otherwise we discard Y and U and start again.

Accept-reject method.





Accept-reject method.

Proof.

$$
\begin{aligned}
\mathbb{P}(Y \le x) &= \mathbb{P}(Y \le x \mid Y \text{ accepted}) \\
&= \mathbb{P}(Y \le x \mid U \le f(Y)/Mg(Y)) \\
&= \frac{\int_{-\infty}^{\infty} \mathbb{P}(Y \le x \cap U \le f(Y)/Mg(Y) \mid Y = y) g(y) dy}{\mathbb{P}(U \le f(Y)/Mg(Y))} \\
&= \frac{\int_{-\infty}^{x} (f(y)/Mg(y))\, g(y) dy}{\mathbb{P}(U \le f(Y)/Mg(Y))} = \frac{\int_{-\infty}^{x} f(y) dy}{M\mathbb{P}(U \le f(Y)/Mg(Y))}
\end{aligned}
$$

in the same way

$$
\begin{aligned}
\mathbb{P}(U \le f(Y)/Mg(Y)) &= \int_{-\infty}^{\infty} \mathbb{P}(U \le f(Y)/Mg(Y) \mid Y = y) g(y) dy \\
&= \int_{-\infty}^{\infty} (f(y)/Mg(y))\, g(y) dy = \frac{1}{M} \int_{-\infty}^{\infty} f(y) dy = \frac{1}{M}.
\end{aligned}
$$

□

Accept-reject method.

The fact that

$$\mathbb{P}\left(U \leq \frac{f(Y)}{Mg(Y)}\right) = \frac{1}{M},$$

means that the number of iterations until an acceptance will be geometric random variable with mean M. Thus it is important to choose g so that M is small.

Accept-reject method.

Example. Let $X \sim G(\frac{3}{2}, 1)$, i.e.

$$f(x) = \frac{1}{\Gamma(3/2)} x^{1/2} e^{-x} = \frac{2}{\sqrt{\pi}} x^{1/2} e^{-x}, \;\; x > 0.$$

And let instrumental variable be $Y \sim Exp(\lambda)$

$$\frac{f(x)}{g(x)} = \frac{K x^{1/2} e^{-x}}{\lambda e^{-\lambda x}} = \frac{K}{\lambda} x^{1/2} e^{-(x-\lambda x)} \to \max_x .$$

The maximum is attained at $x = \frac{1}{2(1-\lambda)}$, and

$$M = \max_x \frac{f(x)}{g(x)} = \frac{K}{\lambda} \cdot \frac{e^{-1/2}}{(2(1-\lambda))^{1/2}} \to \min_\lambda$$

Thus, $\lambda(1-\lambda)^{1/2} \to \max$, which provide $\lambda = \frac{2}{3}$.

Accept-reject method.

Example. Let $X \sim f(x) = \frac{\sqrt{2}}{\sqrt{\pi}} e^{-x^2/2}, x > 0$. And let instrumental variable be $Y \sim Exp(1)$

$$M = \max_x \frac{f(x)}{g(x)} = \frac{f(1)}{g(1)} = \sqrt{\frac{2e}{\pi}} \approx 1.32$$

$$\frac{f(x)}{Mg(x)} = \exp\Big(x - \frac{x^2}{2} - \frac{1}{2}\Big) = \exp\Big(-\frac{(1-x)^2}{2}\Big).$$

Thus, the algorithm:

- generate $Y \sim Exp(1), U \sim U[0,1]$;
- accept Y, if $U \leq \exp\Big(-\frac{(Y-1)^2}{2}\Big)$.

Addition of a choice of sign with $1/2$ probability provides a method to generate normal distributed r.v.

References:

[RC] Cristian P. Robert and George Casella. *Introducing Monte Carlo Methods with R.* Series "Use R!". Springer

Some trick methods. Generation of Poisson(λ).

Let $\xi_i \sim Exp(\lambda)$. Let $S_n = \xi_1 + \cdots + \xi_n$.

$$\begin{aligned} N(1) &= \max\{n : \; S_n \le 1\}, \\ N(1) &\sim \text{Poi}(\lambda), \end{aligned}$$

and

$$\begin{aligned} N(1) &= \max\{n : \; \sum_{i=1}^{n} -\frac{1}{\lambda} \ln U_i \le 1\} \\ &= \max\{n : \; \sum_{i=1}^{n} \ln U_i \ge -\lambda\} \\ &= \max\{n : \; \ln(U_1 \ldots U_n) \ge -\lambda\} \\ &= \max\{n : \; U_1 \ldots U_n \ge e^{-\lambda}\} \\ N(1) &= \max\{n : \; U_1 \ldots U_n < e^{-\lambda}\} + 1. \end{aligned}$$

Some trick methods. Generation of Gamma $\Gamma(n,\lambda)$.

$X \sim \Gamma(n,\lambda)$ if $X = \xi_1 + \cdots + \xi_n$, where $\xi_i \sim Exp(\lambda)$.

By inverse transformation method generate exponential, and

$$X = -\frac{1}{\lambda}\ln(U_1 \cdot \dots \cdot U_n).$$

Some methods. Generation of $B(a,b), a,b \in \mathbb{N}$.

$X \sim B(a,b), a,b \in \mathbb{N} = \{1, 2, \dots\}$, if

$$X = \frac{\sum_{i=1}^{a} \xi_i}{\sum_{i=1}^{a+b} \xi_i},$$

where $\xi_i \sim Exp(1)$.

By inverse transformation method generate exponential r.v.s, and

$$X = \frac{\ln(U_1 \dots U_a)}{\ln(U_1 \dots U_a \cdot U_{a+1} \dots U_{a+b})}$$

Some trick methods. Mixture representation.

[RC] It is sometimes the case that a distribution can be naturally represented as *mixture distribution*; that is we can write it in the form

$$f(x) = \int_{\mathcal{Y}} g(x \mid y) p(y) dy \text{ or } f(x) = \sum_{i \in \mathcal{Y}} p_i f_i(x).$$

To generate a r.v. X using such representation, we can first generate a variable Y from mixing distribution and then generate X from selected conditional distribution. That is,

> if $Y \sim p(\cdot)$ and $X \sim g(\cdot \mid Y)$, then $X \sim f(\cdot)$ (continuous);
>
> if $Y \sim \mathbb{P}(Y = i) = p_i$ and $X \sim f_Y(\cdot)$, then $X \sim f(\cdot)$ (discrete).

Some trick methods. Mixture representation.

Example. Students density with ν degree of freedom can be represented as mixture, where

$$X \mid y \sim N(0, \nu/y), \quad \text{and} \quad Y \sim \chi^2_\nu.$$