Arbitrary p.d.f. generation

Marcin Chrząszcz
mchrzasz@cern.ch

University of Zurich[UZH]

Monte Carlo methods,
7 April, 2016

Reverting the c.d.f.

⇛ Let U be a random variable from $\mathcal{U}(0,1)$
⇛ Now let F be a non decreasing function such that:

$$F(-\infty) = 0 \quad F(\infty) = 1$$

then:

$$X = F^{-1}(U)$$

has a p.d.f. distribution with a c.d.f. function of F.

⇛ Prove:

$$F(x) = \mathcal{P}(U \leqslant F(x)) = \mathcal{P}(F^{-1}(U) \leqslant x) = \mathcal{P}(X \leqslant x) \quad \square$$

⇛ So it looks very simple if $x_1, X_2, ..., X_n$ are random variables from $\mathcal{U}(0,1)$ then: $\{X_i = F^{-1}(x_i)\}, i = 1, ..., n$ is the sequence that has a c.d.f. distirbution of F.

Reverting the c.d.f., examples

⇛ The exponential distribution: $E(0, 1)$.
⇛ The p.d.f.: $\rho(X) = e^{-X}, \ X \geqslant 0$.
⇛ The c.d.f.: $F(x) = \int_0^x e^{-X} dX = 1 - e^{-x}$.
⇛ Now let $R \in \mathcal{U}(0, 1): \ R = F(X) = 1 - e^{-X} \longrightarrow X = -\ln(1 - R)$
⇛ Now we can play a trick: if $R \in \mathcal{U}(0, 1)$ then $1 - R$ also in $\mathcal{U}(0, 1)$.
⇛ In the end we get: $X = -\ln(R)$

⇛ Use the reverting to generate the following distributions:

- E 6.1 $\rho(X) = \frac{c}{X}$ on the interval $[a, b]$, where $0 < a < b < \infty$.
- E6.2 The Breit-Wigner function:

$$\rho_{\theta,\lambda}(X) = \frac{\lambda}{\pi} \frac{1}{(X - \theta)^2 + \lambda^2}$$

Hit: First do $C(0, 1)$ then transform the variables.

Reverting the c.d.f., general case

⇛

Lets assume: F - non decreasing function such that: $F(-\infty) = 0$ and $F(\infty) = 1$. Then a random variable X:

$$X = \inf\{x : U \leqslant F(x)\}$$

has a distribution with a c.d.f. of F

```
int gen_discrete(double rn, double *p){
// generating a discrete distribution
// P{X = k } = p[k], k=0,1,...
// rn random number from U(0,1)
int k=0;
double sum=p[0];
while(suma < rn) suma +=p[++k]
return k;
}
```

⇛ E6.3 Using the above example please generate the p.d.f. accordingly to:

$$\mathcal{P}(X = k) = p_k = A\sin(\frac{\pi}{10}\left[k + \frac{1}{2}\right])$$

where A is the normalization constant (calculate it!). From the generated numbers make a histogram and compare to the exact function.

Reverting the c.d.f., pros and cons

⇛ Pros:

- Very accurate method.
- Fast and easy.
- To generate one random number from c.d.f. you need one random number from $\mathcal{U}(0,1)$.

⇛ Cons:

- Usually we require that the c.d.f. is revertible analytically(small number of functions).
- If we use the numerical method of reverting the function, then it's much slower and less accurate.
- Super hard fro multidimensional distributions.

⇛ Mathematical digression: Numerical reverting the c.d.f.:

- We look for the zero of the function: $F_u(X) \equiv F(X) - U$, where $U \in \mathcal{U}(0,1)$.
- $X_0 = F^{-1}(U)$

J. von Neumann elimination method, 1951

$\Rrightarrow$ Let X be a random variable with a p.d.f. $f(x)$ on an interval $[a,b]$, such that $= \infty < a < b < \infty$ and $f(x) \leqslant c,\ \forall_{x\in[a,b]},\ c < \infty$

$\Rrightarrow$ The algorithm:

1. We generate a point: $(U_1, U_2):\ U_1 \in \mathcal{U}_1(a,b),\ \mathcal{U}_2(0,c)$
2. If $U_2 \leqslant f(U_1)$ then $X = U_1$.
3. If not we eliminate the (U_1, U_2) pair and we start over.

$\Rrightarrow$ Prove:

X has a p.d.f. of $f(x)$.

$$\mathcal{P}(X \leqslant t) = (b-a)c \int_a^t \frac{d_{u_1}}{b-a} \int_0^c \frac{d_{u_2}}{c} \Theta(f(u_1) - u_2) = \int_a^t du_1 f(u_1)$$

J. von Neumann elimination method, 1951

⇛ Let X be a random variable with a p.d.f. $f(x)$ on an interval $[a,b]$, such that $= \infty < a < b < \infty$ and $f(x) \leqslant c,\ \forall_{x\in[a,b]},\ c<\infty$

⇛ The algorithm:

1. We generate a point: $(U_1, U_2):\ U_1 \in \mathcal{U}_1(a,b),\ \mathcal{U}_2(0,c)$
2. If $U_2 \leqslant f(U_1)$ then $X = U_1$.
3. If not we eliminate the (U_1, U_2) pair and we start over.

⇛ Prove:

X has a p.d.f. of $f(x)$.

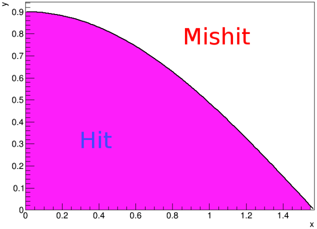
$$\mathcal{P}(X \leqslant t) = (b-a)c \int_a^t \frac{d_{u_1}}{b-a} \int_0^c \frac{d_{u_2}}{c} \Theta(f(u_1) - u_2) = \int_a^t du_1 f(u_1)$$

⇛ We know this already from the Buffon needle:

Generate 2 dim. distribution:

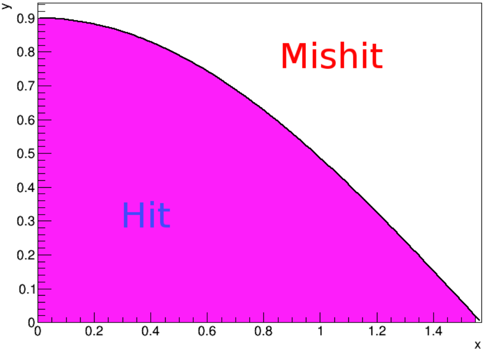
$$(x,y) : \mathcal{U}(0, \frac{\pi}{2}) \times \mathcal{U}(0,1) \text{ and}$$

$$y \begin{cases} \leqslant p(x): & \text{hit,} \\ > p(x): & \text{miss.} \end{cases}$$



⇛ In the end we managed to generate a p.d.f.: $f(x) = \cos(x)$.

Elimination method, example for the Buffon needle



- ⇛ We generate the points from (U_1, U_2) from $\mathcal{U}(0, \frac{\pi}{2}) \times \mathcal{U}(0, 1)$.
- ⇛ Points that are below the function we keep.
- ⇛ The ones above we reject.
- ⇛ The new variable will have the $f(x)$ p.d.f..

Elimination method, multi dimension case

⇛ Let $X = (X_1, X_2, ..., X_m)$ be a m dimensional random variable with a p.d.f.of $f(x_1, x_2, ..., x_m)$ of the domain of $\Omega \in \mathcal{R}^m$ and $f(x_1, x_2, ..x_m) < c < \infty$.
⇛ The algorithm:

- We generate the point $(U_1, U_2, ..., U_m) \in \mathcal{U}(\Omega)$ and $U_{m+1} \in \mathcal{U}(0, c)$
- If $U_{m+1} \leqslant f(U_1, U_2, U_3, ...U_m)$, then $X = (U_1, U_2, ..., U_m)$
- If not we start over.

⇛ The constructed random variable has the distribution of the p.d.f. $f(x_1, x_2, ..., x_m)$.
⇛ Now how do get an $\mathcal{U}(\Omega)$?

Elimination method, multi dimension case

⇛ Let $X = (X_1, X_2, ..., X_m)$ be a m dimensional random variable with a p.d.f.of $f(x_1, x_2, ..., x_m)$ of the domain of $\Omega \in \mathcal{R}^m$ and $f(x_1, x_2, ..x_m) < c < \infty$.

⇛ The algorithm:

- We generate the point $(U_1, U_2, ..., U_m) \in \mathcal{U}(\Omega)$ and $U_{m+1} \in \mathcal{U}(0, c)$
- If $U_{m+1} \leqslant f(U_1, U_2, U_3, ...U_m)$, then $X = (U_1, U_2, ..., U_m)$
- If not we start over.

⇛ The constructed random variable has the distribution of the p.d.f. $f(x_1, x_2, ..., x_m)$.

⇛ Now how do get an $\mathcal{U}(\Omega)$?

⇛ We generate a hypercube $K^m : \Omega \subset K^m$ and accept the points that are within the Ω domain $(\in \Omega)$.

⇛ E6.4 Using the Elimination method generate the following function:

$$\psi(x, y) = c\sqrt{1 - (x^2 + y^2)}, \quad x^2 + y^2 \leqslant 1$$

where c is a normalization constant(to be calculated). Using ROOT make a 2D histogram and compare with the $\psi(x, y)$ function.

Elimination method, general case

⇛ Let $f(x)$ be a p.d.f. accordingly to which we want to generate random variables. (x can be a multi dimension point).

⇛ The algorithm:

- We find a function $g(x)$ such that generating accordingly to it is fast and easy.
- We find a constant c such that:

$$f(x) \leqslant cg(x), \forall_x, \quad \text{optimal value of c} \quad c = \sup_x \frac{f(x)}{g(x)}$$

- We generate X accordingly to $g(x)$ and a random variable $U \in \mathcal{U}(0, c)$
- If $Ug(x) \leqslant f(X)$ then we accept X.
- If not the case we start over.

⇛ The alternative way:

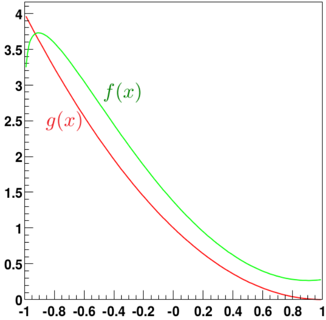
- We calculate the weight: $w(X) = \frac{f(X)}{g(X)}$. We find the maximum weight.
- If $w(X) \geqslant U w_{max}$ then we accept the X.
- If not go back :)

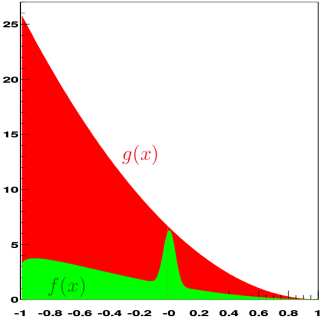
⇛ Remark:

If you are not scared of weighting you might not reject the events but make a weighted histogram.

Elimination method, limitations

⇛ The zeros of the $g(x)$ functions are very dangerous: $f(x) \neq 0$.
⇛ Peaks $f(x)$ are dangerous as they can screw up the efficiency of generating the distributions if they are not correctly approximated by $g(x)$.





Zeros of the function.

Inefficiency of the generation.

Elimination method, example

⇛ Example: $N(0,1)$:

$$f(x) = \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}},\ x \geqslant 0.$$

⇛ We choose the $g(x)$ to simulate the tails:

$$g(x) = e^{-x},\ x \geqslant 0$$

⇛ The weight:

$$w(X) = \frac{f(X()}{g(X)} = \sqrt{\frac{2}{\pi}} e^{\frac{x(2-x)}{2}}$$

⇛ And the maximum weight is:

$$w_{max} = w(1) = \sqrt{\frac{2e}{\pi}}$$

⇛ E6.5 Code up the above example. Use weighted and non weighted case. Compare the results.

Superposition of distribution

⇛ Continues superposition of distribution.
⇛ Let X be a random variable from the p.d.f. $f(x)$:

$$f(x) = \int g_y(x)h(y)dy$$

where $g_y(x)$ is a p.d.f. depending on y parameter;
$h(y)$ is a p.d.f..
⇛ The algorithm:

- Generate the Y accordingly to $h(y)$.
- For a given Y generate the events accordingly to $g_Y(x)$.

⇛ Example: $f(x) = n\int_1^{\infty} y^{-n}e^{-xy}dy,\ \ x, y > 0,\ n > 1.$
We can write down the functions: $g_Y(x) = ye^{-xy}$ and $h(y) = ny^{-(n+1)}$.

1. The numbers Y should be generated from the p.d.f.$h(y)$. One can use the Reverting c.d.f. for example: $Y = (1-U)^{1/n}$, where $U \in \mathcal{U}(0, 1)$.
2. The numbers X of a p.d.f.$g_Y(x)$ are generated from the exponential p.d.f. $E(0, \frac{1}{y}) : X = -\frac{1}{Y}\ln V$, where $V \in \mathcal{U}(0, 1)$.

Discrete superposition of distribution

⇛ Let's:

$$f(x) = \sum_{i=1}^{\infty} p_i g_i(x)$$

where p_i discrete p.d.f.. $p_i \geqslant 0$ and $\sum_{i=1}^{\infty} p_1 = 1$

⇛ The algorithm:

1. Generate the i number accordingly to the p_i p.d.f.. Using Reverting c.d.f. method.
2. For a given i generate the number X from the $g_i(x)$.

⇛ Example: $f(x) = \frac{5}{12}\left[1 + (x-1)^4\right],\ 0 \leqslant x \leqslant 2.$

$\rightarrow\ f(x) = \frac{5}{6} g_1(x) + 1/6 g_2(x),$

where $g_1(x) = \frac{1}{2}$ and $g_2(x) = \frac{5}{2}(x-1)^4$, so $p_1 = \frac{5}{6}$ and $p_2 = \frac{1}{6}$.

$$X = \begin{cases} 2U_2 & \text{if } U_1 < \frac{5}{6} \\ 1 + (2U_2 - 1)^{\frac{1}{5}} & \text{if } U_1 < \frac{5}{6} \end{cases}$$

$U_1, U_2 \in \mathcal{U}(0,1).$

Combination of the superposition and elimination method

⇛ Let's:

$$f(x) = \sum_{i=1}^{\infty} p_i f_i(x), \quad p_i \leqslant 0, \quad \sum_{i=1}^{n} p_i = 1$$

where $f_i(x)$ functions that depend on i.
For each f_i we find a weight function $g_i(x)$ and a constant $c_i > 0$ such as:

$$f_i(x) \leqslant c_i g_i(x) \forall_x .$$

⇛ The algorithm:

1. Generate i accordingly to the distribution p_i
2. For a given i generate X from the p.d.f. $g_i(x)$.
3. Generate a $U \in \mathcal{U}(0,1)$.
4. If $c_i U g_i(X) \leqslant f_i(X)$ then we accept the X if not then we start over.

⇛ Alternative:

1. One can also calculate the weights: $w(X) = \dfrac{f_i(X)}{g_i(X)}$.
2. Find the maximal weight: w_i^{max}.
3. If $w_i(X) \geqslant U w_i^{max}$ we accept X. If not then start over

Combination, warnings, example

⇛ We need a different maximal weight for each "branch". This can be done either numerically or analytically.

⇛ Example:

- Let our p.d.f. be in a form of:

$$f(x) = \sum_{i=1}^{n} c_i x^i, \quad 0 \leqslant x \leqslant 1$$

but $\exists_{j \in \{1,\dots,n\}} : c_j < 0$. We can do a transmigration:

$$cj = c_j^+ - c_j^- : c_j^+, c_j^- > 0$$

⇛ So:

$$f(x) = \sum_{i=1}^{n} c_i x^i \leqslant \sum_{i=1}^{n} c_i^+ x^i = g(x)$$

⇛ Now we use the weight function:

$$\bar{g}(x) = \frac{g(x)}{\sum_{i=1}^{n} c_i^+}$$

Combination, example

⇛ The algorithm:

- Generate X accordingly to $\bar{g}(x)$ p.d.f.. The same we did in the discrete superposition method.
- Generate $U \in \mathcal{U}(0,1)$
- If $Ug(X) \leqslant f(X)$ then we accept the X. If not start over.

⇛ Example $N(0,1)$:

$$f(x) = \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}},\ x \geqslant 0.$$

⇛ We write the function as:

$$f(x)\alpha_1 g_1(x)h_1(x) + \alpha_2 g_2(x)h_2(x)$$

where:

$$\alpha_1 = \sqrt{\frac{2}{\pi}} \qquad \alpha_2 = \sqrt{\frac{1}{2\pi}}$$

$$g_1 = \begin{cases} 1,\ 0 \leqslant x \leqslant 1 \\ 0,\ x \geqslant 1 \end{cases} \qquad g_2 = \begin{cases} 2e^{-2(x-1)},\ x \geqslant 1 \\ 0,\ 0 \leqslant x \leqslant 1 \end{cases}$$

Combination, example

$$h_1(x) = e^{\frac{-x^2}{2}} \qquad h_2(x) = e^{-\frac{(x-2)^2}{2}}$$

⇛ The variable X is chosen:

- With a probability: $\alpha_1/(\alpha_1+\alpha_2) = 2/3$ from the g_1. Then we do the elimination accordingly to h_1 function, aka. $h_1(X) \geqslant U \in \mathcal{U}(0,1)$
- With a probability: $\alpha_2/(\alpha_1+\alpha_2) = 2/3$ from the g_2. Then we do the elimination accordingly to h_2 function, aka. $h_2(X) \geqslant U \in \mathcal{U}(0,1)$

⇛ E6.6 Generate the above distribution using with the above method. Do a histogram and compare with desired p.d.f..

Acceptance - complement, Kornal & Peterson 1981

⇛ If $f = f_1 + f_2$ is a p.d.f. of a random variable X and $p = \int f_1(x)dx$.
⇛ f_1/p_1 and $f_2(1-p)$ are the p.d.f.of a discrete function.
⇛ Now if the functions f_1 and f_2 are easy to generate from then:

- We choose the numbers with the probability distribution f_1/p_1 from the p.d.f. f_1.
- We choose the numbers with the probability distribution $f_2/(1-p_1)$ from the p.d.f. f_2.

⇛ The method is super powerful if $f_2 = \mathrm{const}$.

Backup