

Accept/Reject Algorithm

Suppose you want to generate a draw of a random variable, i.e. from a distribution of a random variable Y but the direct method (integral transform, i.e. $F_Y^{-1}(U) \sim Y$ for $U \sim \text{uniform}(0, 1)$, which is from $F_Y(Y) \sim \text{uniform}(0, 1)$) doesn't work because you don't have a closed form for F_Y . What do you do?

How about this for an idea: print out a picture of the “target” density f_Y (let's suppose compact support), draw a tight bounding box around it, and throw darts at it. Here's a picture to keep in mind (for now pretend that it's completely shaded)

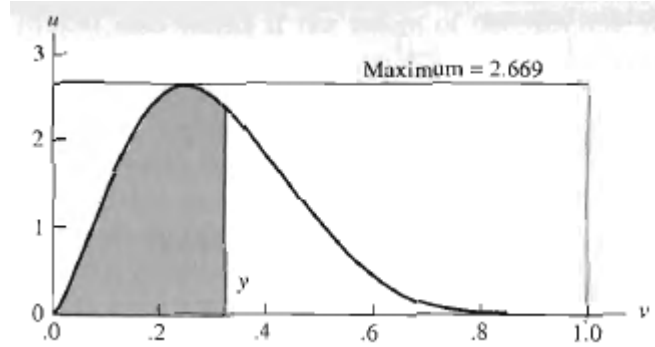


Figure 1:

Let's also pretend we can always nail the bounding box and the darts will be uniformly distributed within, i.e. let V be the random variable that corresponds to the x -coordinate of the dart hit, which is uniformly distributed on the support of the target density f_Y (remember that the support is the range of Y , not the domain) and let U be the random variable that corresponds to the y -coordinate, which is uniformly distributed on $(0, \text{height-of-the-bounding-box}) \equiv (0, M)$ where

$$M = \sup_{y \in \mathcal{Y}} f_Y(y)$$

Then what's the probability that we hit “inside the density”, i.e. under the curve of the density? Because we're throwing the darts uniformly, with uniform distribution on the entire bounding box, it's the area subtended between the curve of density and the x -axis, divided by the area of the entire box. Since U corresponds to the y -coordinate of a dart hit and V corresponds to the x -coordinate we want to know what the probability of the event $\{U \leq f_Y(V)\}$ is, i.e

$$P(U \leq f_Y(V)) = ?$$

Note this is no different than something much more mundane like $P(H \leq G)$ when H, V are for example uniformly distributed on $(0, 1)$ and then $P(H \leq G)$ is just the lower or upper triangle (depending on how you orient the h and g axes). Now V is only defined on \mathcal{Y} so let $\lambda(\mathcal{Y})$ be the Lebesgue measure of \mathcal{Y} and then

$$\begin{aligned}
P(U \leq f_Y(V)) &= \int_{\mathcal{Y}} \left(\int_0^{f_Y(v)} f_{U,V}(u, v) du \right) dv \\
&= \int_{\mathcal{Y}} \left(\int_0^{f_Y(v)} f_U(u) f_V(v) du \right) dv \\
&= \int_{\mathcal{Y}} \left(\int_0^{f_Y(v)} \frac{1}{M} \frac{1}{\lambda(\mathcal{Y})} du \right) dv \\
&= \frac{1}{M} \frac{1}{\lambda(\mathcal{Y})} \int_{\mathcal{Y}} f_Y(v) dv \\
&= \frac{1}{M} \frac{1}{\lambda(\mathcal{Y})}
\end{aligned}$$

Think about whether this makes sense. If $\mathcal{Y} = (0, 1)$ and then the area of the bounding box is $M \cdot 1$, and the density f_Y , which integrates to 1 on its support is fraction

$$\frac{\text{its area}}{M \cdot 1} = \frac{1}{M}$$

*(Skip) Alternatively

$$\begin{aligned}
P(U \leq f_Y(V)) &= E(1_{U \leq f_Y(V)}) \\
&= E(E(1_{U \leq f_Y(V)} | V)) \\
&= E(P(U \leq f_Y(V) | V)) \\
&= \int_{\mathcal{Y}} P(U \leq f_Y(v)) f_V(v) dv \\
&= \int_{\mathcal{Y}} \left(\int_0^{f_Y(v)} f_U(u) du \right) f_V(v) dv \\
&= \frac{1}{M} \frac{1}{\lambda(\mathcal{Y})} \int_{\mathcal{Y}} f_Y(v) dv \\
&= \frac{1}{M} \frac{1}{\lambda(\mathcal{Y})}
\end{aligned}$$

Now fix some x coordinate v . Given that a dart landed somewhere on the line $x = v$ what's the probability it landed underneath the curve? Since M corresponds to the height of our bounding box, the probability that a dart with x coordinate equal to v hit underneath the curve of the pdf is $f_Y(v)/M$. Why? Picture the bounding box and an x -slice of it (cf. Figure 1: picture the line denoting the slice at $v = y$ in Figure 1 extended all the way to the top of the

bounding box). The slice has height M and $f_Y(v)$ is the height of the density there. So the ratio of the length, on that slice, of the portion “underneath the density” to height of the bounding box is $f_Y(v)/M$ and since the darts were thrown with uniform distribution in each coordinate it’s true that the probability of landing underneath is the ratio of the length/total-height. So the claim is

$$P(U \leq f_Y(v)) = f_Y(v)/M$$

When stated this way it’s maybe obvious, since we said the y -coordinate of the dart is uniformly distributed on $(0, M)$ and we’re basically computing

$$\begin{aligned} P(U \leq f_Y(v)) &= \int_0^{f_Y(v)} \left(\frac{1}{M} \right) du \\ &= f_Y(v)/M \end{aligned}$$

So what’s the punchline? Almost there: let’s compute the conditional cumulative distribution of V , the random variable associated with the x -coordinate of our dart hit, conditioned on the dart hitting below the pdf of Y , i.e. conditioned on the event $\{U \leq f_Y(V)\}$ (by analogy with the $P(H \leq G)$ example imagine computing $P(G \leq g | H \leq G)$, i.e. $P(G \leq g)$ given that we’re constrained to the half triangle as mentioned before, except it’s $\{H \leq f(G)\}$ instead of simply $\{H \leq G\}$):

$$\begin{aligned} P(V \leq y | U \leq f_Y(V)) &= \frac{P(V \leq y, U \leq f_Y(V))}{P(U \leq f_Y(V))} \\ &= \frac{\int_A \int_0^{f_Y(v)} f_{U,V}(u, v) du dv}{1/M\lambda(\mathcal{Y})} \text{ where } A = \{v \in \mathcal{Y} : v \leq y\} \\ &= M\lambda(\mathcal{Y}) \int_A \int_0^{f_Y(v)} f_U(u) f_V(v) du dv \\ &= M\lambda(\mathcal{Y}) \int_A \left(\int_0^{f_Y(v)} \frac{1}{M} \frac{1}{\lambda(\mathcal{Y})} du \right) dv \\ &= \int_A \int_0^{f_Y(v)} du dv \\ &= \int_A f_Y(v) dv \\ &= P(Y \leq y) \end{aligned}$$

What does this mean? It means that the conditional distribution of V , conditioned on the event $\{U \leq f_Y(V)\}$, is the same as the distribution of Y , i.e. the x -coordinate of all of the darts that hit below the curve of the pdf is distributed the same as Y . The distilled algorithm then is: throw darts at your pdf-dartboard. If one hits below the curve of the pdf then “accept” it and take its x -coordinate as a true draw from the distribution of Y but otherwise “reject” it and throw again.

How many darts do you have to throw before this will happen? Well since $P(U \leq f_Y(V)) = \frac{1}{M} \frac{1}{\lambda(\mathcal{Y})}$ and $1 - P(U \leq f_Y(V)) = P(U > f_Y(V)) = 1 - \frac{1}{M} \frac{1}{\lambda(\mathcal{Y})}$ we see that the random variable

$N = \text{number of pairs darts with coordinates } (v_i, u_i) \text{ before } u_i \leq f_Y(v_i)$

is geometrically distributed with success probability $\frac{1}{M} \frac{1}{\lambda(\mathcal{Y})}$ and therefore expectation equal to $M\lambda(\mathcal{Y})$, so you'd have to throw on average $M\lambda(\mathcal{Y})$ darts.

Particular Example

Take for example $Y \sim \text{beta}(2.7, 6.3)$ (note that for integer values of the shape and scale parameters α, β the integral transform will work).

Let

$$M = \sup_{y \in \mathcal{Y}} f_Y(y)$$

Take two independent uniform random variables $U \sim \text{uniform}(0, M)$ and $V \sim \text{uniform}(0, 1)$. Then

$$\begin{aligned} P(V \leq y, U \leq f_Y(V)) &= \int_0^y \left(\int_0^{f_Y(v)} f_{U,V}(u, v) du \right) dv \\ &= \frac{1}{M} \frac{1}{1} \int_0^y \left(\int_0^{f_Y(v)} du \right) dv \\ &= \frac{1}{M} \int_0^y f_Y(v) dv \\ &= \frac{1}{M} P(Y \leq y) \end{aligned}$$

and

$$P(U \leq f_Y(V)) = \frac{1}{M}$$

and so reaffirming our calculations from the motivational discussion

$$P(V \leq y | U \leq f_Y(V)) = \frac{P(V \leq y, U \leq f_Y(V))}{P(U \leq f_Y(V))} = P(Y \leq y)$$

Recapitulating: if you want draw from a $Y \sim \text{beta}(2.7, 6.3)$ distribution then draw pairs (v, u) from (U, V) (where U and V are as defined above) and if

$$u \leq f_Y(v) = v^{2.7-1} (1-v)^{6.3} / \text{B}(2.7, 6.3)$$

where $\text{B}(2.7, 6.3)$ is the beta function defined by the

$$\text{B}(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$$

then “accept” v as a true draw from a $\text{beta}(2.7, 6.3)$ distribution and if $u > f_Y(v) = v^{2.7-1} (1-v)^{6.3} / B(2.7, 6.3)$ then “reject” and simply draw again.

Generalization

Now why did this work? Well many reasons (the entire argument in the first part) but one in particular is that the support \mathcal{Y} of $Y \sim \text{beta}(2.7, 6.3)$, which is $(0, 1)$, is the same as the support of V (otherwise there would be no hope of them being equal, regardless of the conditional). But of course you couldn’t use a uniform as the “candidate” (what V is called) to draw from a distribution that has non-compact support (e.g. something like a Gamma distribution which has support \mathbb{R}) since there’s no uniform distribution on unbounded support (nothing like $\text{uniform}(0, \infty)$). So you need another candidate but it has to “cover” the target density, i.e. we need a bounding shape not a bounding box, and we need to make sure that we throw darts only in this bounding shape. Then conditionally the x -coordinate of our dart throws should again be distributed the same as the target density.

First of all we need to scale our candidate density so that it definitely covers the one we want to sample from, the target density. If we compute

$$M = \sup_{y \in \mathcal{Y}} \frac{f_Y(y)}{f_V(y)}$$

we can scale our candidate density by M so that it’s always greater than or equal to our intractable density. The effect of the scaling is only seen insofar as when we throw darts underneath the “bounding shape”, i.e. the candidate density, the y -coordinate of our darts now has to be constrained differently for each x -coordinate:

$$U|V \sim \text{uniform}(0, M \cdot f_V(V))$$

Picture two densities, one covering the other like this (where Proposal=candidate)

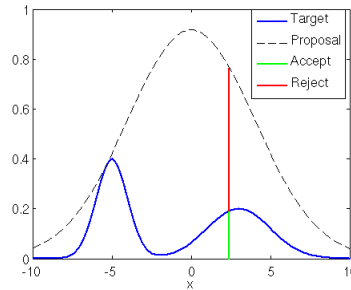


Figure 2:

and we have to again find the probability of the event $\{U \leq f_Y(V)\}$

$$\begin{aligned}
P(U \leq f_Y(V)) &= \int_{\mathcal{Y}} \left(\int_0^{f_Y(v)} f_{U,V}(u, v) du \right) dv \\
&= \int_{\mathcal{Y}} \left(\int_0^{f_Y(v)} f_{U|V}(u|v) f_V(v) du \right) dv \\
&= \int_{\mathcal{Y}} \left(\int_0^{f_Y(v)} \frac{1}{M \cdot f_V(v)} f_V(v) du \right) dv \\
&= \frac{1}{M} \int_{\mathcal{Y}} \left(\int_0^{f_Y(v)} du \right) dv \\
&= \frac{1}{M} \int_{\mathcal{Y}} f_Y(v) dv \\
&= \frac{1}{M}
\end{aligned}$$

and then

$$\begin{aligned}
P(V \leq y | U \leq f_Y(V)) &= \frac{P(V \leq y, U \leq f_Y(V))}{P(U \leq f_Y(V))} \\
&= \frac{\int_A \int_0^{f_Y(v)} f_{U,V}(u, v) dudv}{1/M} \text{ where } A = \{v \in \mathcal{Y} : v \leq y\} \\
&= M \int_A \int_0^{f_Y(v)} f_{U|V}(u|v) f_V(v) dudv \\
&= M \int_A \left(\int_0^{f_Y(v)} \frac{1}{M \cdot f_V(v)} f_V(v) du \right) dv \\
&= \int_A \left(\int_0^{f_Y(v)} du \right) dv \\
&= \int_A f_Y(v) dv \\
&= P(Y \leq y)
\end{aligned}$$

How many darts do you have to throw before this will happen? Well since $P(U \leq f_Y(V)) = \frac{1}{M}$ and $1 - P(U \leq f_Y(V)) = P(U > f_Y(V)) = 1 - \frac{1}{M}$ we see that the random variable

N = number of pairs darts with coordinates (v_i, u_i) before $u_i \leq f_Y(v_i)$

is geometrically distributed with success probability $\frac{1}{M}$ and therefore expectation equal to M , so you'd have to throw on average M darts.