Simulation - Lecture 3 - Rejection Sampling

Lecture version: Tuesday 4th February, 2020, 12:12

Robert Davies

Part A Simulation and Statistical Programming

Hilary Term 2020

Recap from previous lecture

- Monte Carlo is useful to calculate e.g. integrals when distributions are analytically difficult to work with. But we need iid rvs
- Assume we can always generate $U_i \sim \mathcal{U}[0,1]$
- Last time, we looked at two easy ways to get iid $X \sim P$
- **Inversion method**. Invert CDF, return $X_i = F_X^{-1}(U_i)$
- **Transformation method**. Find new distribution Q and function φ such that $X = \varphi(Y) \sim P$. Then draw $Y_i \sim Q$, and return $X = \varphi(Y_i)$
- Today: **Rejection sampling method**. Draw random variables from larger space defined by proposal pdf, and "reject" those not in region defined by target pdf

Idea

- Let X be a continuous r.v. on Ω with pdf f_X
- Consider a continuous rv variable $U > 0$ such that the conditional pdf of U given $X = x$ is

$$f_{U|X}(u|x) = \begin{cases} \frac{1}{f_X(x)} & \text{if } u < f_X(x) \\ 0 & \text{otherwise} \end{cases}$$

- The joint pdf of (X, U) is

$$\begin{aligned} f_{X,U}(x,u) &= f_X(x) \times f_{U|X}(u|x) \\ &= f_X(x) \times \frac{1}{f_X(x)} \mathbb{I}(0 < u < f_X(x)) \\ &= \mathbb{I}(0 < u < f_X(x)) \end{aligned}$$

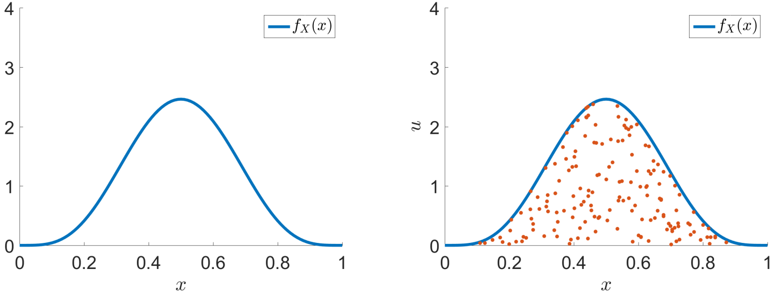
- Uniform distribution on the set $\mathcal{A} = \{(x,u) | 0 < u < f_X(x), x \in \Omega\}$

Fundamental Theorem of simulation

Theorem (Fundamental Theorem of simulation)

Let X be a rv on Ω with pdf or pmf f_X. Simulating X is equivalent to simulating

$$(X, U) \sim \mathrm{Unif}(\{(x, u) | x \in \Omega, 0 < u < f_X(x)\})$$



Rejection sampling idea

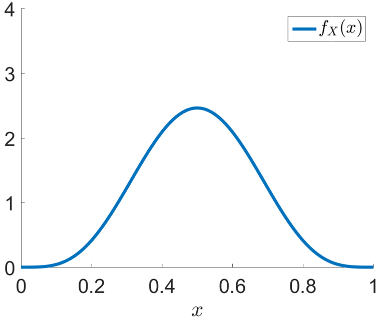
- Direct sampling of (X, U) uniformly over the set $\mathcal{A}$ is in general challenging
- Consider some superset $\mathcal{S}$ such that $\mathcal{A} \subseteq \mathcal{S}$, such that simulating uniform rv on $\mathcal{S}$ is easy
- Therefore, a uniform distribution on $\mathcal{A}$ can be obtained by drawing from a uniform distribution on $\mathcal{S}$, and *rejecting* samples in $\mathcal{S}$ not in $\mathcal{A}$
- Rejection sampling technique:
    1. Simulate $(Y, V) \sim \mathrm{Unif}(\mathcal{S})$, with simulated values y and v
    2. if $(y, v) \in \mathcal{A}$ then stop and return $X = y$,$U = v$,
    3. otherwise go back to 1.
- The resulting rv (X, U) is uniformly distributed on $\mathcal{A}$
- X is marginally distributed from f_X

Example: Beta density

- Let $X \sim \mathrm{Beta}(5,5)$ be a continuous rv with pdf

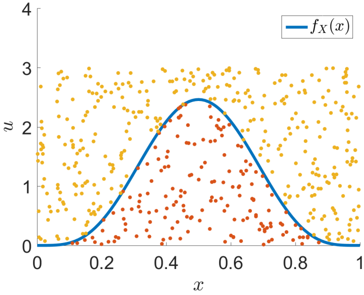
$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1$$

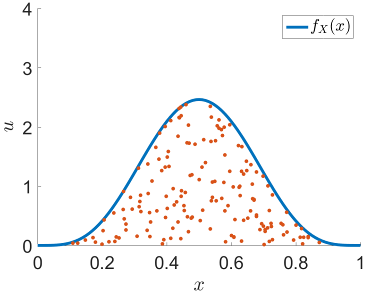
  where $\alpha = \beta = 5$.
- $f_X(x)$ is upper bounded by 3 on $[0, 1]$.



Example: Beta density

- Let $\mathcal{S} = \{(y, v) | y \in [0, 1], v \in [0, 3]\}$
  1. Simulate $Y \sim \mathcal{U}([0, 1])$ and $V \sim \mathcal{U}([0, 3])$, with simulated values y and v
  2. If $v < f_X(x)$, return $X = x$
  3. Otherwise go back to Step 1.
- Only requires simulating uniform random variables and evaluating the pdf pointwise





Rejection sampling more precisely

- Consider X a random variable on Ω with a pdf/pmf $f(x)$, a target distribution
- We want to sample from f using a proposal pdf/pmf q which we can sample.
- **Proposition**. Suppose we can find a constant M such that $f(x)/q(x) \leq M$ for all $x \in \Omega$.
- The following 'Rejection' algorithm returns $X \sim f$.

Rejection sampling method

Algorithm 1 Rejection sampling

- Identify proposal distribution Q that is easy to simulate from, with pdf q_Q, and find M such that $f_X(x)/q_Q(x) \leq M$ for all $x \in \Omega$
- Simulate $Y_i \sim Q$, and $U_i \sim \mathcal{U}[0,1]$
- For $U_i \leq f(Y_i)/q(Y_i)/M$, return an $X_i = Y_i$, otherwise do not return a value

Rejection Sampling: Proof for discrete rv

- We have

$$\begin{aligned}\Pr(X=x) &= \sum_{n=1}^{\infty} \Pr(\text{reject } n-1 \text{ times, draw } Y=x \text{ and accept it}) \\ &= \sum_{n=1}^{\infty} \Pr(\text{reject } Y)^{n-1} \Pr(\text{draw } Y=x \text{ and accept it})\end{aligned}$$

- We have

$$\begin{aligned}& \Pr(\text{draw } Y=x \text{ and accept it}) \\ =\ & \Pr(\text{draw } Y=x) \Pr(\text{accept } Y \mid Y=x) \\ =\ & q(x) \Pr\left(U \leq \frac{f(Y)}{q(Y)}/M \,\middle|\, Y=x\right) \\ =\ & \frac{f(x)}{M}\end{aligned}$$

- The probability of having a rejection is

$$\begin{aligned}\Pr(\text{reject } Y) &= \sum_{x\in\Omega} \Pr(\text{draw } Y = x \text{ and reject it}) \\ &= \sum_{x\in\Omega} q(x) \Pr\left(U \geq \frac{f(Y)}{q(Y)}/M \,\middle|\, Y = x\right) \\ &= \sum_{x\in\Omega} q(x)\left(1 - \frac{f(x)}{q(x)M}\right) = 1 - \frac{1}{M}\end{aligned}$$

- Hence we have

$$\begin{aligned}\Pr(X = x) &= \sum_{n=1}^{\infty} \Pr(\text{reject } Y)^{n-1} \Pr(\text{draw } Y = x \text{ and accept it}) \\ &= \sum_{n=1}^{\infty}\left(1 - \frac{1}{M}\right)^{n-1} \frac{f(x)}{M} = f(x).\end{aligned}$$

- Note the number of accept/reject trials has a geometric distribution of success probability $1/M$, so the mean number of trials is M.

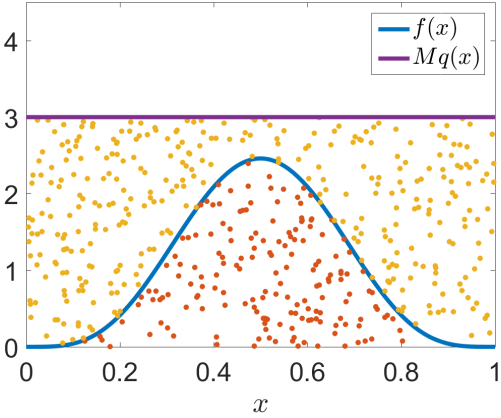
Rejection Sampling: Proof for continuous scalar rv

- Here is an alternative proof given for a continuous scalar variable X, the rejection algorithm still works but f, q are now pdfs.
- We accept the proposal Y whenever $(U, Y) \sim f_{U,Y}$ where $f_{U,Y}(u, y) = q(y)\mathbb{I}_{(0,1)}(u)$ satisfies $U \leq f(Y)/(Mq(Y))$.
- We have

$$
\begin{aligned}
\Pr(X \leq x) &= \Pr(Y \leq x | U \leq f(Y)/Mq(Y)) \\
&= \frac{\Pr(Y \leq x, U \leq f(Y)/Mq(Y))}{\Pr(U \leq f(Y)/Mq(Y))} \\
&= \frac{\int_{-\infty}^{x} \int_{0}^{f(y)/Mq(y)} f_{U,Y}(u, y) du dy}{\int_{-\infty}^{\infty} \int_{0}^{f(y)/Mq(y)} f_{U,Y}(u, y) du dy} \\
&= \frac{\int_{-\infty}^{x} \int_{0}^{f(y)/Mq(y)} q(y) du dy}{\int_{-\infty}^{\infty} \int_{0}^{f(y)/Mq(y)} q(y) du dy} = \int_{-\infty}^{x} f(y) dy.
\end{aligned}
$$

Example for target is beta, proposal is uniform

- $f(x)$ is the pdf of a $\mathrm{Beta}(5, 5)$ rv
- Proposal density q is the pdf of a uniform rv on $[0, 1]$



Calculating a minimal M

- Assume you have for $\alpha, \beta \geq 1$

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1$$

  which is upper bounded on $[0,1]$.
- We have the proposal $q(x) = \mathbb{I}_{(0,1)}(x)$ the uniform density on $[0,1]$.
- We need to find a bound M s.t. $f(x)/q(x) \leq M \iff f(x) \leq M$. We therefore want to set $M = \max_{0<x<1} f(x)$ and we obtain by solving for $f'(x) = 0$

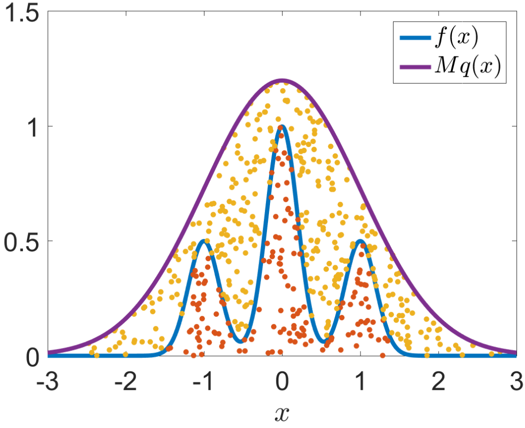
$$M = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \underbrace{\left(\frac{\alpha-1}{\alpha+\beta-2}\right)^{\alpha-1}\left(\frac{\beta-1}{\alpha+\beta-2}\right)^{\beta-1}}_{M'}$$

  which gives

$$\frac{f(y)}{Mq(y)} = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{M'}.$$

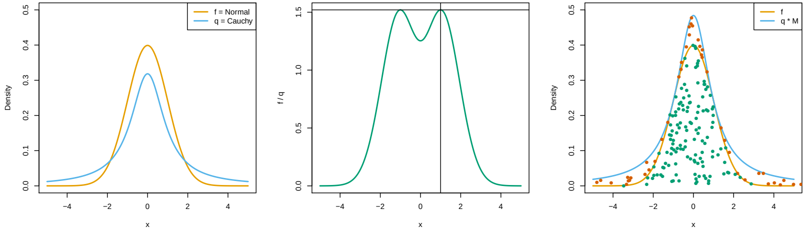
Illustrations for multimodal distribution

- $X \in \mathbb{R}$ with multimodal pdf
- Proposal density q is the pdf of a standardized normal



Normal example

- Let $X \sim N(0,1)$, with pdf $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$
- Consider as a proposal distribution $Y \sim q$ the Cauchy distribution, with pdf $q(x) = \frac{1}{\pi(1+x^2)}$
- We can work out that $\frac{f_X(x)}{q_Y(x)} \leq M$ for $M = \sqrt{2\pi}e^{-\frac{1}{2}}$
- We can generate Y from U using inversion $Y = \tan(\pi(U - \frac{1}{2}))$

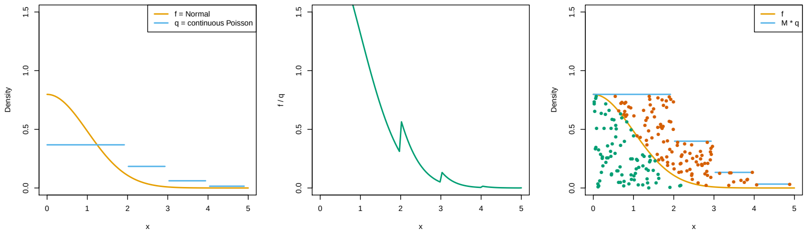


Normal from Cauchy code

```
f_X <- function(x) { 1 / sqrt(2 * pi) * exp(-0.5 * x ** 2)}
f_Y <- function(x) { 1 / pi / (1 + x ** 2)}
M <- sqrt(2 * pi) * exp(-1 / 2)
set.seed(914)
n <- 10000
x <- array(NA, n)
i <- 1 ## index
while(i <= n) {
    U1 <- runif(1)
    Xp <- tan(pi * (U1 - 0.5)) ## proposed
    U2 <- runif(1)
    if (U2 <= (f_X(Xp) / f_Y(Xp) / M)) {
        x[i] <- Xp
        i <- i + 1
    }
}
c(mean(x), var(x)) ## 0.007508397 0.985407347
```

Using a block uniform discrete proposal with a continuous target

- Let $X = |Z|$, with $Z \sim N(0,1)$, *i.e.* $f_X(x) = \frac{2}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ for $x > 0$
- Consider bounding it by Y from a continuous analog of a $Poisson(1)$ pmf
- Then using inversion for a discrete rv, we can sample from a $Poisson$ the normal way, sampling the continuous version using another random uniform



Code for inverse of a Poisson with rate 1

```
F_inverse_poisson <- function() {
    U <- runif(1)
    j <- 0
    done <- FALSE
    pL <- pU <- 0
    while(!done) {
        pU <- pU + exp(-1) / factorial(j)
        if ((pL <= U) & (U < pU)) {
            done <- TRUE
        } else {
            pL <- pU
            j <- j + 1
        }
    }
    return(j)
}
```

Code for truncated normal

```
set.seed(41)
n <- 10000
x <- array(NA, n)
i <- 1
M <- sqrt(2 / pi) * exp(1)
while(i <= n) {
    Xp <- F_inverse_poisson() + runif(1)
    f <- 2 / (sqrt(2 * pi)) * exp( - 0.5 * Xp**2)
    q <- exp(-1) / factorial(floor(Xp))
    if (runif(1) <= f / q / M) {
        x[i] <- Xp
        i <- i + 1
    }
}
c(mean(x), var(x)) ## 0.7981175 0.3605557
y <- abs(rnorm(n))
c(mean(y), var(y)) ## 0.7951576 0.3681160
```

Dealing with Unknown Normalising Constants

- In most practical scenarios, we only know $f(x)$ and $q(x)$ up to some normalising constants; i.e.

  $$f(x) = \tilde{f}(x)/Z_f \text{ and } q(x) = \tilde{q}(x)/Z_q$$

  where $\tilde{f}(x), \tilde{q}(x)$ are known but $Z_f = \int_\Omega \tilde{f}(x)dx$, $Z_q = \int_\Omega \tilde{q}(x)dx$ are unknown/expensive to compute.
- Rejection can still be used: Indeed $f(x)/q(x) \leq M$ for all $x \in \Omega$ iff $\tilde{f}(x)/\tilde{q}(x) \leq \tilde{M}$, with $\tilde{M} = Z_f M/Z_q$.
- Practically, this means we can ignore the normalising constants from the start: if we can find $\tilde{M}$ to bound $\tilde{f}(x)/\tilde{q}(x)$ then it is correct to accept with probability $\tilde{f}(x)/(\tilde{M}\tilde{q}(x))$ in the rejection algorithm. In this case the mean number N of accept/reject trials will equal $Z_q\tilde{M}/Z_f$ (that is, M again).

Example without normalization: gamma random variables

- We want to simulate a random variable $X \sim$Gamma(α, β) which works for any $\alpha \geq 1$ (not just integers);

$$f(x) = \frac{x^{\alpha-1}\exp(-\beta x)}{Z_f} \text{ for } x > 0, \quad Z_f = \Gamma(\alpha)/\beta^{\alpha}$$

  so $\tilde{f}(x) = x^{\alpha-1}\exp(-\beta x)$ will do as our unnormalised target.
- We saw that for $\alpha = a$ a positive integer we can simulate $X \sim$ Gamma(a, β) by adding a independent Exp(β) variables, $Y_i \sim$ Exp(β), $X = \sum_{i=1}^{a} Y_i$.
- So we can sample densities "close" in shape to Gamma(α, β) since we can sample Gamma$(\lfloor \alpha \rfloor, \beta)$. Perhaps we can use this as a proposal density?

Gamma rvs bound

- Let $a = \lfloor \alpha \rfloor$ and let's try to use Gamma(a, b) as the proposal, so $Y \sim$ Gamma(a, b) for integer $a \geq 1$ and some $b > 0$. The density of Y is

$$q(x) = \frac{x^{a-1} \exp(-bx)}{Z_q} \text{ for } x > 0, \quad Z_q = \Gamma(a)/b^a$$

  so we can use $\tilde{q}(x) = x^{a-1} \exp(-bx)$.
- We have to check whether the ratio $\tilde{f}(x)/\tilde{q}(x)$ is bounded over $\mathbb{R}_+$ where

$$\tilde{f}(x)/\tilde{q}(x) = x^{\alpha-a} \exp(-(\beta - b)x).$$

- Consider (a) $x \to 0$ and (b) $x \to \infty$. For (a) we need $a \leq \alpha$ so $a = \lfloor \alpha \rfloor$ is indeed fine. For (b) we need $b < \beta$ (not $b = \beta$ since we need the exponential to kill off the growth of $x^{\alpha-a}$).

Gamma rvs bound continued

- Given that we have chosen $a = \lfloor \alpha \rfloor$ and $b < \beta$ for the ratio to be bounded, we now compute the bound.
- $\frac{d}{dx}(\tilde{f}(x)/\tilde{q}(x)) = 0$ at $x = (\alpha - a)/(\beta - b)$ (and this must be a maximum at $x \geq 0$ under our conditions on a and b), so $\tilde{f}(x)/\tilde{q}(x) \leq \tilde{M}$ for all $x \geq 0$ if

$$\tilde{M} = \left(\frac{\alpha - a}{\beta - b}\right)^{\alpha - a} \exp(-(\alpha - a)).$$

- So accept Y if $U \leq \tilde{f}(Y)/\tilde{M}\tilde{q}(Y)$ where $\tilde{f}(Y)/\tilde{M}\tilde{q}(Y) = Y^{\alpha - a} \exp(-(\beta - b)Y)/\tilde{M}$.

Gamma rvs and the best choice of b

- Any $0 < b < \beta$ will do, but is there a best choice of b?
- Idea: choose b to minimize the expected number of simulations of Y per sample X output.
- Since the number N of trials is Geometric, with success probability $Z_f/(\tilde{M} Z_q)$, the expected number of trials is $\mathbb{E}(N) = Z_q \tilde{M}/Z_f$. Now $Z_f = \Gamma(\alpha)\beta^{-\alpha}$ where Γ is the Gamma function related to the factorial.
- Practice: Show that the optimal b solves $\frac{d}{db}(b^{-a}(\beta - b)^{-\alpha+a}) = 0$ so deduce that $b = \beta(a/\alpha)$ is the optimal choice.

Simulating normal random variables, revisited

- Recall $f(x) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}x^2)$ and $q(x) = 1/\pi/(1+x^2)$. We have

$$\frac{\tilde{f}(x)}{\tilde{q}(x)} = (1+x^2)\exp\left(-\frac{1}{2}x^2\right) \leq 2/\sqrt{e} = \tilde{M}$$

  which is attained at ± 1.
- Hence the probability of acceptance is

$$\mathbb{P}\left(U \leq \frac{\tilde{f}(Y)}{\tilde{M}\tilde{q}(Y)}\right) = \frac{Z_f}{\tilde{M}Z_q} = \frac{\sqrt{2\pi}}{\frac{2}{\sqrt{e}}\pi} = \sqrt{\frac{e}{2\pi}} \approx 0.66$$

  and the mean number of trials to success is approximately $1/0.66 \approx 1.52.$ (which matches our M from earlier)

Rejection Sampling in High Dimension

- Consider

$$\tilde{f}(x_1, ..., x_d) = \exp\left(-\frac{1}{2}\sum_{k=1}^{d} x_k^2\right)$$

and

$$\tilde{q}(x_1, ..., x_d) = \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{d} x_k^2\right)$$

- For $\sigma > 1$, we have

$$\frac{\tilde{f}(x_1, ..., x_d)}{\tilde{q}(x_1, ..., x_d)} = \exp\left(-\frac{1}{2}\left(1 - \sigma^{-2}\right)\sum_{k=1}^{d} x_k^2\right) \le 1 = \tilde{M}.$$

- The acceptance probability of a proposal for $\sigma > 1$ is

$$\mathbb{P}\left(U \le \frac{\tilde{f}(X_1, ..., X_d)}{\tilde{M}\tilde{q}(X_1, ..., X_d)}\right) = \frac{Z_f}{\tilde{M} Z_q} = \sigma^{-d}.$$

- The acceptance probability goes exponentially fast to zero with d.

Recap

- Rejection sampling relies on the idea that sampling uniformly from the area under your target density will return random variables distributed according to your target density
- To do this, you need a proposal density q that covers your target density f, where your proposal density is easily to sample from
- Given a bound M with $f(x)/q(x) \leq M \ \forall x$, then rejection sampling is
  - Draw $Y \sim q$ and $U \sim \mathcal{U}[0,1]$
  - Keep $X = Y$ if $U \leq \frac{f(Y)}{q(Y)M}$
- Then $X \sim f$, and M is the expected number of samples per returned X