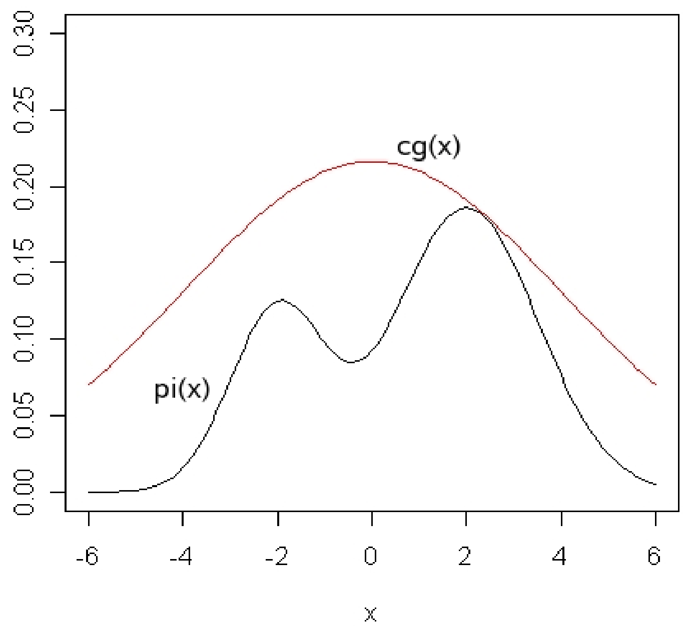
Rejection Sampling

STAT 525

9/4/18

The Rejection Method

- Suppose $\pi(\mathbf{x})$ is difficult to sample from. If we can find a constant c and an easy-to-sample distribution $g(\mathbf{x})$ such that $\pi(\mathbf{x}) \leq cg(\mathbf{x})$, then the rejection method can be used.



Rejection Sampling Algorithm [von Neumann (1951)]:

1. Draw $\mathbf{x} \sim g(\mathbf{x})$.

2. Draw u $\sim$Unif [0,1].

3. Accept $\mathbf{x}$ if
$$u \leq \frac{\pi(\mathbf{x})}{cg(\mathbf{x})};$$
otherwise, reject $\mathbf{x}$.

The accepted samples follow the target distribution $\pi(\mathbf{x})$.

The distribution $g(\mathbf{x})$ is called instrumental distribution.

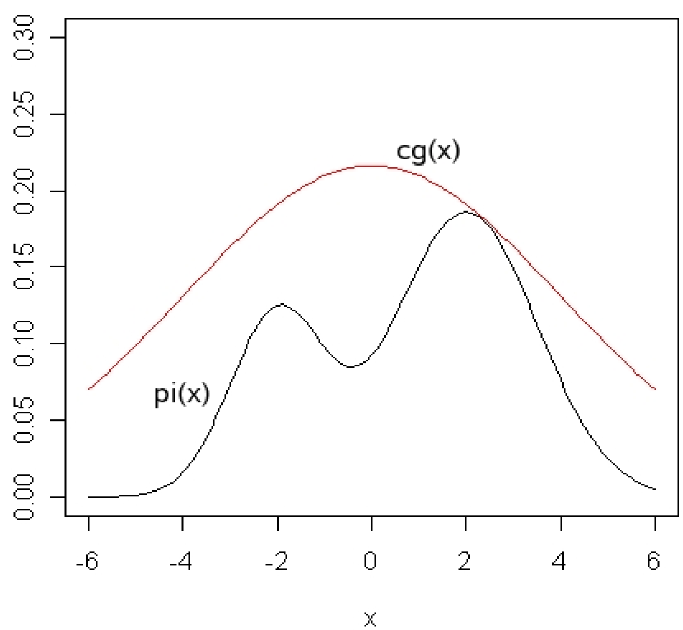
Proof

Call accepted X as Y. Then

$$
\begin{aligned}
& P(Y \leq y) \\
= \;& P\left(X \leq y | U \leq \frac{\pi(X)}{cg(X)}\right) = \frac{P\left(X \leq y, U \leq \frac{\pi(X)}{cg(X)}\right)}{P\left(U \leq \frac{\pi(X)}{cg(X)}\right)} \\
= \;& \frac{\int_{-\infty}^{y} g(\mathbf{x}) \int_0^{\frac{\pi(\mathbf{x})}{cg(\mathbf{x})}} du d\mathbf{x}}{\int_{-\infty}^{\infty} g(\mathbf{x}) \int_0^{\frac{\pi(\mathbf{x})}{cg(\mathbf{x})}} du d\mathbf{x}} = \frac{\int_{-\infty}^{y} \frac{\pi(\mathbf{x})}{cg(\mathbf{x})} g(\mathbf{x}) d\mathbf{x}}{\int_{-\infty}^{\infty} \frac{\pi(\mathbf{x})}{cg(\mathbf{x})} g(\mathbf{x}) d\mathbf{x}} \\
= \;& \frac{\frac{1}{c}\int_{-\infty}^{y} \pi(\mathbf{x}) d\mathbf{x}}{\frac{1}{c}\int_{-\infty}^{\infty} \pi(\mathbf{x}) d\mathbf{x}} = \int_{-\infty}^{y} \pi(\mathbf{x}) d\mathbf{x}. \qquad \text{So } Y \sim \pi(\mathbf{x}).
\end{aligned}
$$

Acceptance Rate

- The acceptance rate is $P\left(U \leq \frac{\pi(X)}{cg(X)}\right) = \frac{1}{c}$, i.e., the expected number of "operations" for obtaining one accepted sample is c. Choose c as small as possible.



Remarks

- The support of $g(\mathbf{x})$ should include the support of $\pi(\mathbf{x})$.
- Normalizing constant of $\pi(\mathbf{x})$ doesn't have to be known. If $\pi(\mathbf{x}) \propto l(\mathbf{x})$, just replace $\pi(\mathbf{x})$ by $l(\mathbf{x})$ in the algorithm.

  This is important especially for Bayesian inference because the posterior

  $$p(\theta|D) \propto p(D|\theta)p(\theta)$$

  is often only known up to a normalizing constant.

Another Way to View Rejection Sampling

Rejection sampling:

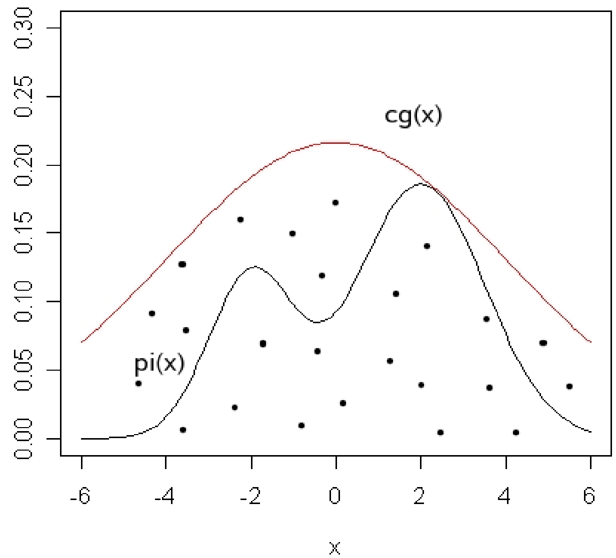
1. Draw $\mathbf{x} \sim g(\mathbf{x})$.
2. Draw u ∼Unif[0, 1].
3. Accept $\mathbf{x}$ if $u \leq \frac{\pi(\mathbf{x})}{cg(\mathbf{x})}$.

An equivalent algorithm:

1. Draw $\mathbf{x} \sim g(\mathbf{x})$.
2. Draw u^* ∼Unif[0, $cg(\mathbf{x})$].
3. Accept $\mathbf{x}$ if $u^* \leq \pi(\mathbf{x})$.

The equivalent algorithm:

1. Draw $\mathbf{x} \sim g(\mathbf{x})$.
2. Draw u^* $\sim$Unif[0, $cg(\mathbf{x})$].
3. Accept $\mathbf{x}$ if $u^* \leq \pi(\mathbf{x})$.



- In the equivalent algorithm, (x, u^*) is uniform in the area under $cg(\mathbf{x})$. We only accept those points that are under $\pi(\mathbf{x})$. So the marginal distribution of $\mathbf{x}$ of accepted points follow $\pi(\mathbf{x})$.
- Proportion of the area under $\pi(\mathbf{x})$ is the acceptance rate.

- Example 1. Want to draw sample from

$$\pi(x) = \frac{e}{e-1} e^{-x} e^{-e^{-x}}, \qquad 0 < x < \infty.$$

- Example 2. Want to draw sample from

$$\pi(x) \propto e^{\cos x}, \qquad 0 \leq x \leq 2\pi.$$

(This is a special Von Mises distribution for a random angle.)

- Example 3. Want to draw sample from truncated normal

$$\pi(x) \propto \phi(x) I_{\{x>d\}}.$$

where $\phi(x)$ is the standard normal density and I is the indicator function.

- A simple method is to generate from $N(0,1)$ and keep the samples that are larger than d. This works when $d < 0$ or small.

- When $d > 0$, especially when d is large, we can use the rejection method with a shifted exponential distribution as the instrumental distribution.

- Consider $g(x) = \lambda_0 e^{-\lambda_0 (x-d)}$. We want to find a small constant c such that

$$\frac{\phi(x)}{1-\Phi(d)} \leq c\lambda_0 e^{-\lambda_0(x-d)}, \;\; \forall\, x > d.$$

One choice of c is

$$c = \frac{\exp\{(\lambda_0^2 - 2\lambda_0 d)/2\}}{\sqrt{2\pi}\lambda_0(1-\Phi(d))}.$$

To achieve the maximum acceptance rate, the best λ_0 is

$$\lambda_0 = (d + \sqrt{d^2+4})/2.$$

- Comparison of the acceptance rates:

	d		
	0	1	2
Simple method	0.50	0.16	0.02
Rejection method	0.76	0.88	0.93

- The acceptance rate for the rejection method increases as d increases.

References

- Section 2.2 of Jun Liu's *Monte Carlo Strategies in Scientific Computing*.