Review: Rejection sampling

- We want $x \sim f(x)$
- We know how to $x \sim g(x)$
- Assume $\frac{f(x)}{g(x)} \leq c$ for all x where $f(x) > 0$.

Algorithm:

repeat

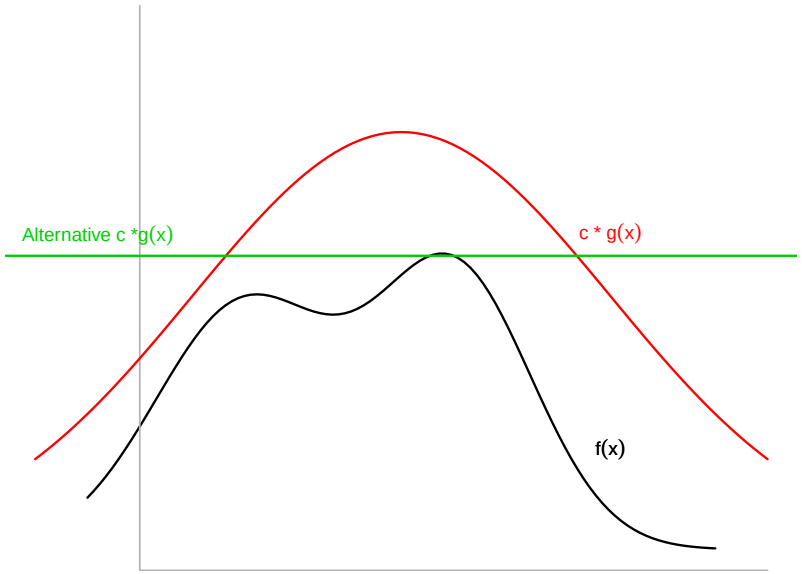
generate $x \sim g(x)$

compute $\alpha = \frac{1}{c} \cdot \frac{f(x)}{g(x)}$

generate $u \sim U[0, 1]$

until $u \leq \alpha$

Review: Rejection sampling



Rejection sampling - Acceptance probability

The acceptance probability can be written as

$$P(c \cdot U \cdot g(x) \leq f(x)) = \int_{-\infty}^{\infty} \frac{f(x)}{c \cdot g(x)} g(x)\, dx = \int_{-\infty}^{\infty} \frac{f(x)}{c}\, dx = c^{-1}.$$

The single trials are independent, so the number of trials up to the first success is geometrically distributed with parameter $1/c$. The expected number of trials up to the first success is therefore c.

Problem:

In high-dimensional spaces c is generally large so that many samples will get rejected.

Continuation: Standard Cauchy

How can we sample from the semi-unit circle?

Rejection sampling - Acceptance probability

Note: For c to be small, $g(x)$ must be similar to $f(x)$.
The art of rejection sampling is to find a g(x) that is similar to f(x) and which we know how to sample from.

Issues: c is generally large in high-dimensional spaces, and since $\alpha = 1/c$, many samples will get rejected.

Weighted resampling

A problem when using rejection sampling is to find a legal value for c. An approximation to rejection sampling is the following:

Let, as before:

- $f(x)$: target distribution
- $g(x)$: proposal distribution

Algorithm

- Generate $x_1, \ldots, x_n \sim g(x)$ iid
- Compute weights
$$w_i = \frac{\frac{f(x_i)}{g(x_i)}}{\sum_{i=1}^{n} \frac{f(x_i)}{g(x_i)}}$$
- For $i = 1, \ldots, m$ generate a sample from the discrete distribution on $\{x_1, \ldots, x_n\}$ with probabilities $w_1, \ldots, w_n$.

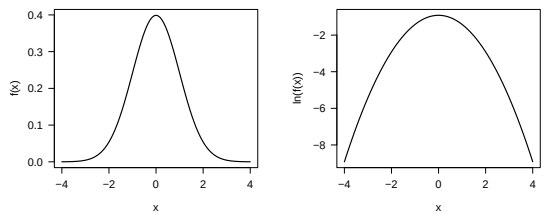
end.

Note: This is only an approximate algorithm.

Adaptive rejection sampling

This method works only for log concave densities, i.e.

$$\frac{\partial}{\partial x} \ln f(x) \leq 0, \quad \text{for all } x.$$

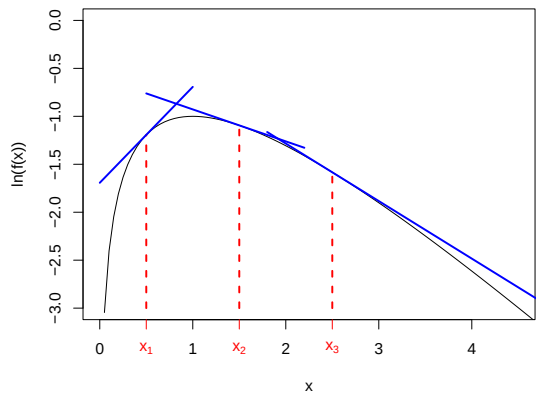


Many densities are log-concave, e.g. the normal, the gamma ($a > 1$), densities arising in GLMs with canonical link.

Basic idea: Form an upper envelope (the upper bound on $f(x)$) adaptively and use this in place of $c \cdot g(x)$ in rejection sampling.

Adaptive rejection sampling (2)

- Start with an initial grid of points $x_1, x_2, \ldots, x_m$ and construct the envelope using the tangents at $\ln(f(x_i))$, $i = 1, \ldots, m$.
- Draw a sample from the envelop function and if accepted the process is terminated. Otherwise, use it to refine the grid.



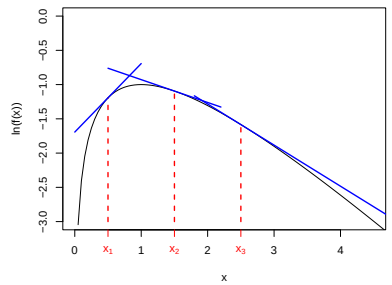
Adaptive rejection sampling (3)

The piecewise exponential distribution is defined as

$$g(x) = k_m \lambda_m \exp(-\lambda_m (x - t_{m-1})),$$

with $x \in (t_{m-1}, t_m], i = 1, \ldots, m$.

Here, $t_1, \ldots, t_m$ denotes the partition (intersection points of the tangent), λ_m is the slope of the tangent at x_m and k_m accounts for the corresponding offset. ($t_0 = 0, k_1 = 1$).



Importance sampling

One of the principal reasons for wishing to sample from complicated probability distributions $f(z)$ is to be able to evaluate expectations with respect to some function $p(z)$:

$$\mathrm{E}(p) = \int p(z) f(z) dz \approx \frac{1}{L} \sum_{l=1}^{L} p(z^{(l)}).$$

The technique of importance sampling provides a framework for approximating expectations directly but does not itself provide a mechanism for drawing samples from a distribution.

Importance sampling (2)

Importance sampling is based on the use of a proposal distribution $g(x)$ from which it is easy to draw samples.

We can then express the expectation in the form of a finite sum over samples $x^{(l)}$ drawn from $g(x)$:

$$\begin{aligned}
\mathrm{E}(p) &= \int p(z) f(z) dz \\
&= \int p(z) \frac{f(z)}{g(z)} g(z) dz \\
&\simeq \frac{1}{L} \sum_{l=1}^{L} \frac{f(z^{(l)})}{g(z^{(l)})} p(z^{(l)}).
\end{aligned}$$

The quantities $w_l = f(z^{(l)})/g(z^{(l)})$ are known as importance weights. The importance weights correct the bias from a wrong distribution.

Importance sampling (3)

As with rejection sampling, the success of importance sampling depends crucially on how well the proposal distribution $g(x)$ matches the target distribution $f(x)$.