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 $x_1, \dots, x_n \sim g(x)$ iid
- Compute weights

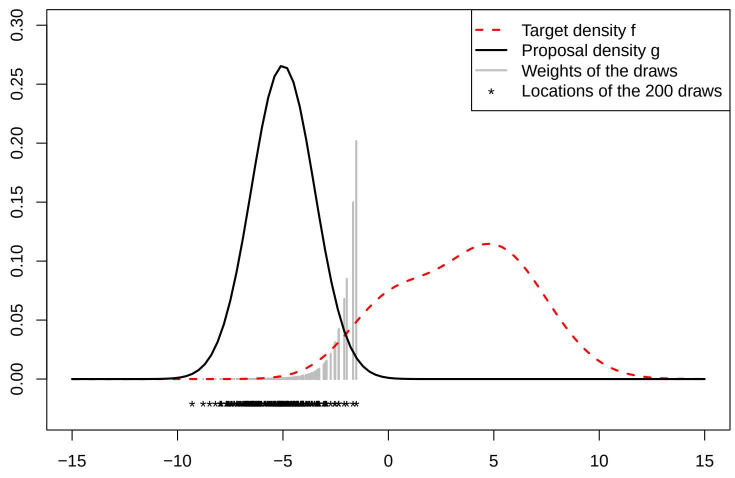
$$w_i = \frac{\frac{f(x_i)}{g(x_i)}}{\sum_{j=1}^{n} \frac{f(x_j)}{g(x_j)}}$$

- Generate a second sample of size m from the discrete distribution on $\{x_1, \dots, x_n\}$ with probabilities $w_1, \dots, w_n$.

Comments

- The resulting sample has approximate distribution f
- The resample can be drawn with or without replacement provided that $n >> m$, a suggestion is $n/m = 20$.
- The normalising constant is not needed.
- This approximate algorithm is sometimes called sampling importance resampling (SIR) algorithm.

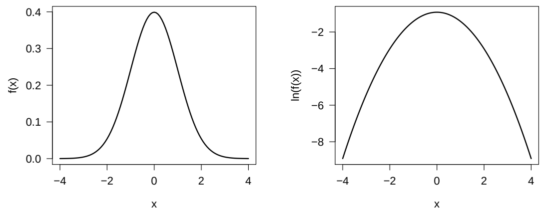
Illustration



Adaptive rejection sampling

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

$$(\ln f)''(x) \le 0, \quad \text{for all } x.$$

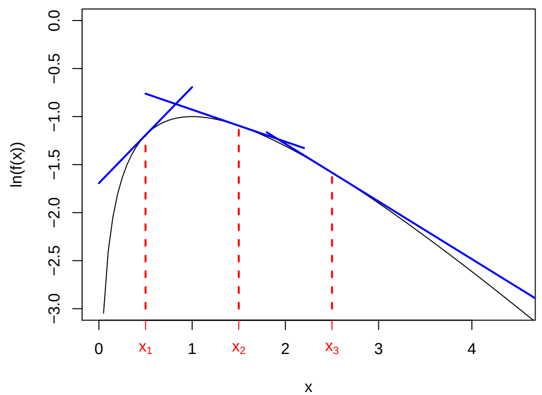


Many densities are log-concave, e.g. the normal, the gamma ($a > 1$), densities arising in GLMs with canonical link.

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 $x_1, x_2, \ldots, x_m$ (with at least one x_i on each side of the maximum of $\ln(f(x))$) 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.



Monte Carlo integration

Assumption

It is *easy* to generate independent samples $x^{(1)}, \ldots, x^{(M)}$ from a distribution $f(x)$ of interest.

A Monte Carlo estimate of the mean

$$\mathrm{E}(x) = \int x f(x) dx$$

is then given by

$$\hat{\mathrm{E}}(x) = \frac{1}{M} \sum_{m=1}^{M} x^{(m)}.$$

The strong law of large numbers ensures, that this estimate is consistent. This approach is called Monte Carlo integration

Monte Carlo integration (II)

Monte Carlo integration

Suppose $x^{(1)}, \ldots, x^{(M)}$ is an iid sample drawn from $f(x)$. Then the strong law of large numbers says:

$$\hat{\mathrm{E}}(g(x)) = \frac{1}{M} \sum_{m=1}^{M} g(x^{(m)}) \overset{a.s}{\to} \int g(x) f(x) dx = \mathrm{E}(g(x))$$

Examples

- Using $g(x) = x^2$ we obtain an estimate for $\mathrm{E}(x^2)$.
- An estimate for the variance follows as

$$\widehat{\mathrm{Var}}(x) = \hat{\mathrm{E}}(x^2) - \hat{\mathrm{E}}(x)^2$$

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 $p(z)$:

$$E(p) = \int p(z)f(z)dz$$

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.

$$\begin{aligned} E(p) &= \int p(z)f(z)dz \\ &= \int p(z)\frac{f(z)}{g(z)}g(z)dz \end{aligned}$$

whre $w(z) = f(z)/g(z)$ are known as importance weights.

Importance sampling estimators

The former expression suggests two different importance sampling estimators

$$\hat{E}(p) = \frac{1}{L}\sum_{l=1}^{L} p(z^{(l)}) \cdot w(z^{(l)}). \quad (1)$$

$$\hat{E}(p) = \frac{1}{\sum_{l=1}^{L} w(z^{(l)})}\sum_{l=1}^{L} p(z^{(l)}) \cdot w(z^{(l)}). \quad (2)$$

The difference between these two estimates is usually small. The main advantage of the second estimator is that it does not require the normalizing constants of f and g in order to be computed.

Importance sampling: Summary

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