Rejection Sampling

Yutao Chen

⧖ Apr 12 2026
⧗ Apr 13 2026

Contents

Rejection Sampling 1
Adaptive Rejection Sampling 2

Rejection sampling is a basic method for sampling from distributions

$$p(\boldsymbol{x}) = \tilde{p}(\boldsymbol{x}) / Z_p,$$

with possibly unknown normalizing constants $Z_p = \int \tilde{p}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$.

Rejection Sampling

The core idea is to use a proposal distribution $q(\boldsymbol{x})$ such that $Cq(\boldsymbol{x}) \geq \tilde{p}(\boldsymbol{x})$ for some constant C. That is, $Cq(\boldsymbol{x})$ is an upper bound of $\tilde{p}(\boldsymbol{x})$.

> The proposal distribution $q(\boldsymbol{x})$ generally should be easy to sample from (e.g. Gaussians) compared to the target distribution $p(\boldsymbol{x})$.

To sample from the $p(\boldsymbol{x})$, we first sample $\boldsymbol{x} \sim q(\boldsymbol{x})$, and then either *accpet* or *reject* the sample $\boldsymbol{x}$ with probability

$$q(\text{accept} | \boldsymbol{x}) = \frac{\tilde{p}(\boldsymbol{x})}{Cq(\boldsymbol{x})}.$$

Consequently, using the Bayes theorem we can show that

$$\begin{aligned} q(\boldsymbol{x} | \text{accept}) &= q(\text{accept} | \boldsymbol{x}) \cdot q(\boldsymbol{x}) / q(\text{accept}) \\ &= \frac{\tilde{p}(\boldsymbol{x}) / C}{\int \tilde{p}(\boldsymbol{x}) / C \, \mathrm{d}\boldsymbol{x}} = \frac{\tilde{p}(\boldsymbol{x})}{Z_p} = p(\boldsymbol{x}). \end{aligned}$$

That is, the accepted samples will be distributed according to $p(\boldsymbol{x})$.

Adaptive Rejection Sampling

The efficiency of rejection sampling is mostly determined by how "tight" the upper bound $Cq(\boldsymbol{x})$ is.

> For example, one can in theory use an arbitrarily large C, but this results in a loose bound and hence high rejection rate.

Adaptive rejection sampling is a method for automatically constructing a tight upper bound $q(\boldsymbol{x})$ for any log-concave univariate density $p(\boldsymbol{x})$.

1. We discretize the support of $p(\boldsymbol{x})$ into finitely many grids;
2. We construct a linear upper bound of $\log \tilde{p}(\boldsymbol{x})$ for each grid.

The resulting upper bound $q(\boldsymbol{x})$ will be piecewise exponential as the log of the bound is piecewise linear.

Rejection sampling does not scale well to high dimensional space, as *(a)* it is challenging to come up with a tight upper bound, and *(b)* rejection rates are almost always ≈ 1 due to the curse of dimensionality.