Rejection sampling

From Wikipedia, the free encyclopedia

In mathematics, **rejection sampling** is a basic technique used to generate observations from a distribution. It is also commonly called the **acceptance-rejection method** or "accept-reject algorithm" and is a type of Monte Carlo method. The method works for any distribution in $\mathbb{R}^m$ with a density.

Rejection sampling is based on the observation that to sample a random variable one can sample uniformly from the region under the graph of its density function.[1][2]

Contents

Description

To visualize the motivation behind rejection sampling, imagine graphing the distribution of a random variable onto a large rectangular board and throwing darts at it. Assume that the darts are uniformly distributed around the board. Now take off (reject) all of the darts that are outside the area under the curve. The remaining darts will be distributed uniformly within the area under the curve, and the x-positions of these darts will be distributed according to the random variable's density. This is because there is the most room for the darts to land where the curve is highest and thus the probability density is greatest.

The visualization as just described is equivalent to a particular form of rejection sampling where the proposal distribution is uniform (hence its graph is a rectangle). The general form of rejection sampling assumes that the board is not necessarily rectangular but is shaped according to some distribution that we know how to sample from (e.g. using inversion sampling), and which is at least as high at every point as the distribution we want to sample from, so that the former completely encloses the latter. (Otherwise, there will be parts of the curved area we want to sample from that can never be reached.) Rejection sampling works as follows:

1. Sample a point (an x-position) from the proposal distribution.
2. Draw a vertical line at this x-position, up to the curve of the proposal distribution.
3. Sample uniformly along this line (i.e. uniformly from 0 to the value of the proposal distribution (maximum of the probability density function)). If the sampled value is greater than the value of the desired distribution at this vertical line, return to step 1.

Note also that this algorithm can be used to sample from the area under *any* curve, regardless of whether the function integrates to 1. In fact, scaling a function by a constant has no effect on the sampled x-positions. This means that the algorithm can be used to sample from a distribution whose probability density function is only

known up to a constant (i.e. whose normalizing constant is unknown), which is common in computational statistics.

Examples

As a simple geometric example, suppose it is desired to generate a random point within the unit circle. Generate a candidate point (x, y) where x and y are independent uniformly distributed between -1 and 1. If it so happens that $x^2 + y^2 \leq 1$ then the point is within the unit circle and should be accepted. If not then this point should be rejected and another candidate should be generated.
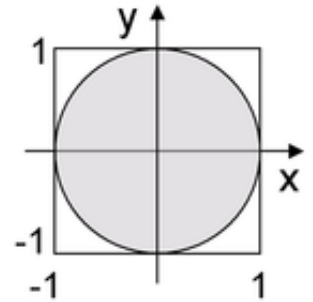
The ziggurat algorithm, a more advanced example, is used to efficiently generate normally-distributed pseudorandom numbers.

Theory

The rejection sampling method generates sampling values from an arbitrary probability distribution function $f(x)$ by using an instrumental distribution $g(x)$, under the only restriction that $f(x) < Mg(x)$ where $M > 1$ is an appropriate bound on $f(x)/g(x)$.

Rejection sampling is usually used in cases where the form of $f(x)$ makes sampling difficult. Instead of sampling directly from the distribution $f(x)$, we use an envelope distribution $Mg(x)$ where sampling is easier. These samples from $Mg(x)$ are probabilistically accepted or rejected.

This method relates to the general field of Monte Carlo techniques, including Markov chain Monte Carlo algorithms that also use a proxy distribution to achieve simulation from the target distribution $f(x)$. It forms the basis for algorithms such as the Metropolis algorithm.

The **unconditional acceptance probability** is the proportion of proposed samples which are accepted, which is $Pr(u < f(x)/Mg(x)) = E[f(x)/Mg(x)] = \int (f(x)/Mg(x))g(x)dx = (1/M)$ $\int f(x)dx = 1/M$. If M is low, fewer samples are rejected, and the required number of samples for the target distribution is obtained more quickly. Because M must be no less than the maximum of $f(x)/g(x)$, the unconditional acceptance probability is higher the less that ratio varies, however to obtain acceptance probability 1, $f(x) = g(x)$, which defeats the purpose of sampling.

Algorithm

The algorithm (used by John von Neumann and dating back to Buffon and his needle) is as follows:

- Sample x from $g(x)$ and u from $U(0, 1)$ (the uniform distribution over the unit interval)
- Check whether or not $u < f(x)/Mg(x)$.
  - If this holds, accept x as a realization of $f(x)$;
  - if not, reject the value of x and repeat the sampling step.

The validation of this method is the **envelope principle**: when simulating the pair $(x, v = u * Mg(x))$, one produces a uniform simulation over the subgraph of $Mg(x)$. Accepting only pairs such that $u < f(x)/Mg(x)$ then produces pairs (x, v) uniformly distributed over the subgraph of $f(x)$ and thus, marginally, a simulation from $f(x)$.

This means that, with enough replicates, the algorithm generates a sample from the desired distribution $f(x)$. There are a number of extensions to this algorithm, such as the Metropolis algorithm.

Drawbacks

Rejection sampling can lead to a lot of unwanted samples being taken if the function being sampled is highly concentrated in a certain region, for example a function that has a spike at some location. For many distributions, this problem can be solved using an adaptive extension (see adaptive rejection sampling). In addition, as the dimensions of the problem get larger, the ratio of the embedded volume to the "corners" of the embedding volume tends towards zero, thus a lot of rejections can take place before a useful sample is generated, thus making the algorithm inefficient and impractical. See curse of dimensionality. In high dimensions, it is necessary to use a different approach, typically a Markov chain Monte Carlo method such as Metropolis sampling or Gibbs sampling. (However, Gibbs sampling, which breaks down a multi-dimensional sampling problem into a series of low-dimensional samples, may use rejection sampling as one of its steps.)

Adaptive rejection sampling

For many distributions, finding a proposal distribution that includes the given distribution without a lot of wasted space is difficult. An extension of rejection sampling that can be used to overcome this difficulty and efficiently sample from a large number of distributions (provided that they are log-concave, which is in fact the case for most of the common distributions) is known as **adaptive rejection sampling**. The basic idea is to model the proposal distribution using a set of piecewise exponential distributions (i.e. segments of one or more exponential distributions, attached end to end). This can be easier visualized in log space (i.e. by looking at the logarithm of the distribution). The logarithm of an exponential distribution is a straight line, and hence this method essentially involves enclosing the logarithm of the density in a series of line segments. This is the source of the log-concave restriction: if a distribution is log-concave, then its logarithm is concave (shaped like an upside-down U), meaning that a line segment tangent to the curve will always pass over the curve. The method essentially involves successively determining an envelope of straight-line segments that approximates the logarithm better and better while still remaining above the curve, starting with a fixed number of segments (possibly just a single tangent line). Any time we choose a point that is rejected, we tighten the envelope with another line segment that is tangent to the curve at the point with the same x-coordinate as the chosen point.

See also

- Inverse transform sampling
- Pseudo-random number sampling

References

1. ^ Neal, Radford M. (2003). "Slice Sampling". *Annals of Statistics* **31** (3): 705–767. doi:10.1214/aos/1056562461 (http://dx.doi.org/10.1214%2Faos%2F1056562461). MR 1994729 (//www.ams.org/mathscinet-getitem?mr=1994729). Zbl 1051.65007 (http://www.zentralblatt-math.org/zmath/en/search/?format=complete&q=an:1051.65007).
2. ^ Bishop, Christopher (2006). "11.4: Slice sampling". *Pattern Recognition and Machine Learning*. Springer.

ISBN 0-387-31073-8.

- Robert, C.P. and Casella, G. "Monte Carlo Statistical Methods" (second edition). New York: Springer-Verlag, 2004.
- J. von Neumann, "Various techniques used in connection with random digits. Monte Carlo methods", Nat. Bureau Standards, 12 (1951), pp. 36–38.

Retrieved from "http://en.wikipedia.org/w/index.php?title=Rejection_sampling&oldid=600448482"
Categories: Monte Carlo methods │ Non-uniform random numbers