Sampling

Rejection sampling

We have some distribution that we wish to sample from, f(x). The problem is that we cannot sample from it. We have a distribution q(x) which we can easily sample from. If the distribution q(x) envelopes f(x) such that all values of f(x) are included in the distribution q(x) then we can sample from q(x) and apply some acceptance/rejection criteria to the samples.

To ensure that the distribution q(x) envelops f(x) we choose a constant scaling factor, m > 1, such that cq(x) > f(x) for all x.

m $= max \frac{f(x)}{q(x)}$

The above formula is basically taking the maximum distance between the distributions and scaling by that so q(x) is at or above f(x) at all points

A common criterion for accepting samples from x ~ q(x) is based on the ratio of the target distribution to that of the envelope distribution.

Aneels method of rejection is to generate a sample from q and the uniform distribution from (0, mq(x)). If f(x) < u then reject the sample x, otherwise accept. What's happening here is ratios. say 33% of the pdf for q(x) is above f(x), then there's a 66% chance that if we generated a value from the uniform distribution between (0, mq(x)) it would be within (0, f(x)). So the sample x generated from distribution q has a 66% chance of being accepted in this method. Following this logic, as many samples are generated the correct ratio of samples for f would be accepted and the pdf would be all g.

The problem with rejection sampling is that it's not very efficient. We reject lots of samples and so have to do more work to generate samples. That's why you try to keep the distance between the scaled distribution and f to a minimum so more samples are generated i.e pick a good enveloping distribution.

Useful resource: https://theclevermachine.wordpress.com/2012/09/10/rejection-sampling/

Inverse Transform Sampling

Method for generating random numbers from a probability distribution using its inverse cumulative distribution function, cdf, $F^{-1}(x)$

Way of generating independent samples from a given probability density

The cumulative distribution function for a random variable X is $F_X(x) = P(X \leq x)$

This method of sampling is 100% efficient, meaning all generated points are accepted as a sample.

Explanation in English

Required to know the inverse cdf of the distribution. . .

So you want to generate samples for a distribution and you have the cdf of the distribution. In other words you want to input a probability and receive a valid value. Well if you solve for x, the cdf of your desired distribution you'll be left with a formula where if you input a probability (which will be generated via a uniform dist between zero and 1) it'll output a sample that corresponds with the cdf of that probability

Watch this video: https://www.youtube.com/watch?v=rnBbYsysPaU

Continuous   We assume we want to generate a random variable X with cumulative distribution function (CDF) F_X. The inverse transform sampling algorithm is:

1. Generate U ~ Unif(0,1)
2. Let X = F_X^{-1}(U)

X will follow the distribution governed by the CDF F_X

Discrete

1. Generate U ~ Unif(0, 1)
2. Determine the index k such that $\sum_{j=1}^{k-1} p_j \leq U < \sum_{j=1}^{k} p_j and return X = x_k$

Off-point

to get the cdf of a value you take your pdf for the distribution (the probability of getting that value) and get the integral of the pdf at that point, which is the area under the curve up to that point, which means finding the probability that the value is less than or equal to your value.

Importance Sampling

Sampling from one distribution given another using weighting. . .

$E_g[x] = (1/n)\sum_x xg(x) = (1/n)\sum_x x\frac{g(x)}{f(x)}f(x) = (1/n)E_f[x\frac{g(x)}{f(x)}]$

f(x) must be non-zero whenever g(x) is non-zero. Divide by zero error.

Soooo sample from the alternate distribution and the weighting will give a better approximation than the naive solution.

Way of estimating properties of a particular distribution, while only having samples generated from a different distribution than the distribution of interest.

The idea is that is the alternate function f(x) is well chosen than any approximations made will be better than performing a naive approximation using g(x)

So you have the pdf of two distributions but can only sample from one. To sample from the other just take the pdf value of the distribution you can sample from for each x and multiply this by the ratio of the two pdfs for that x form the distributions

doesn't have to be normalized

Gibbs Sampling

Should know your bayesian probabilities

A way of sampling from probability distributions with greater than 2 dimensions

Markov chains

Accept all proposals -> 100% efficient

Sampling would refer to generating sequences of paired outcomes.

Relies on us knowing the conditional distributions and being able to sample from the conditional distributions.

Start off sampling at some sample of the joint distribution

1. sample a value of A from the conditional distribution of A given B
2. sample a value of B from the conditional distribution of B given the A just sampled

Looking at the number of values sampled from the distributions we start to see an approximation of the joint probabilities of A and B.

Algorithm:

$Define : P(\theta_1, \theta_2, \theta_3)$

$(\theta_1^0, \theta_2^0, \theta_3^0) \sim \pi(...)$ where pi is some distribution that can sample θ...

for t in 1:T θ_1^t $P(\theta_1|\theta_2^{t-1}, \theta_3^{t-1})$ θ_2^t $P(\theta_2|\theta_1^t, \theta_3^{t-1})$ θ_3^t $P(\theta_3|\theta_1^t, \theta_2^t)$

- counting the number of occurrences of $(\theta_1, \theta_2, \theta_3)$ as your joint distribution

Problems:

- must know the conditional distributions
- Slow for correlated parameters

Kolmogorov-Smirnov Test

non parametric test. Does not need to follow a normal distribution.

To test in disferences in the shape of two sample distributions

Compares the overall shape

Requires Univariate, continous data

D = the maximum absolute distance between the samples

the smaller the D the more similar the samples to each other

P value is obtained by finding the area under the expected distribution for values as least as extreme as the observed d statistic.

Acronyms