Monte Carlo Inference

Lesson No. 11

Leonardo Alves de Melo - 156188

1 Introduction

There are some methods to find the posterior using deterministic algorithms, using derivative to reach their purpose. The problem is that find the derivative sometimes is not the wisest way to resolve some kind of problems, due to intractable calculations. A way to solve this issue is use the Monte Carlo Approximation, which the main idea is to get some samples from the posterior and use them to computed what is needed. The approximation can be seen in Equation 1, where f is some suitable function.

$$\mathbb{E}[f|D] \approx \frac{1}{S}\sum_{s=1}^{S} f(x^s) \tag{1}$$

In this summary it will be presented some technics to generate samples from a probabilistic distribution even in high dimensions.

2 Curiosities

This technic was developed in physics for the creation of the first atomic bomb. The name *Monte Carlo* comes from a casino in Monaco where the uncle of Stanislaw Ulam would borrow money from relatives to gamble.

3 Sampling in Standard Distributions

The best way to start is talking about the methods for sampling for low dimensions, like one or two in standard form.

3.1 Sampling with Cumulative Distribution Function

For a one dimensional function, we can use the Theorem 1 to confirms that for a random number $u \sim U(0,1)$ representing the height up of y axis, when the axis reachs the curve denoted in Figure 1 it can be found the value of x, denoted by $x = F^{-1}(u)$.

Theorem 1 *if $U \sim$ U(0,1) is a uniform random variable, then* $F^{-1}(U) \sim F$

3.2 Sampling Using Box-Muller Method

This method uses a gaussian to sample. Given a sample z_1 and $z_2 \in (-1,1)$ and exclude the pairs that $z_1^2 + z_2^2 \leq 1$ is not valid. Now defining Equation 3, where i is 1 or 2 and $r^2 = z_1^2 + z_2^2$, and using the multivariate change of variables formula, we get the Box-Muller Method given by Equation 2.

$$p(x_1,x_2) = p(z_1,z_2)|\frac{\delta(z_1,z_2)}{delta(x_1,x_2)}| = [\frac{1}{\sqrt{2\pi}}exp(-\frac{1}{2}x_1^2)][\frac{1}{\sqrt{2\pi}}exp(-\frac{1}{2}x_2^2)] \tag{2}$$

To sample, we get the Cholesky decomposition of the covariance matrix and calculate the Equation 4.

$$x_i = z_i(\frac{-2lnr^2}{r^2})^{\frac{1}{2}} \tag{3}$$

$$y = Lx + \mu \tag{4}$$

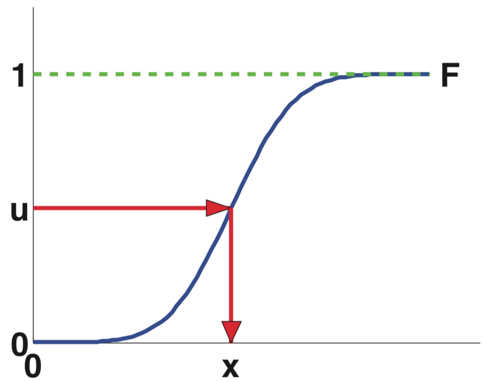

Figure 1. A Cumulative Distribution

4 Sampling in Non-Standard Distributions

4.1 Rejection Sampling

In this sampling is created a proposal distribution $q(x)$ that satisfies $Mq(x) \geq \tilde{p}(x)$, where M is a constant and $\tilde{p}(x)$ is an unnormalized version of $p(x)$.

Then we pick a sample $x \sim q(x)$ and $u \sim U(0,1)$, and if $u > \frac{\tilde{p}(x)}{Mq(x)}$, we accept the sample, and reject otherwise.

4.2 Importance Sampling

In this method we will sample approximating integrals of the form of Equation 5.

$$I = \mathbb{E}[f] = \int f(x)p(x)dx \tag{5}$$

We will sample on the important parts of the space, i.e., where both $p(x)$ and $|f(x)|$ is large, getting a result that is super efficient.

To handle unnormalized distributions but without its normalization constant, Z_p, using an unnormalized proposal $\tilde{q}(x)$ with possibly unknown normlization constant Z_q by the Equation 6, where w_s is given by Equation 7.

$$\hat{I} = \sum_{s=1}^{S} w_s f(x^s) \tag{6}$$

$$w_s \triangleq \frac{\tilde{w}_s}{\sum_{s'} \tilde{w}_{s'}} \tag{7}$$

In this case, when $S \to \infty$, it comes that $\hat{I} \to I$, but with under weak assumptions.

4.3 Particle Filtering

It is for recursive Bayesian inference, and is used in many areas, like time-series, online parameter learning and others. The main idea is to approximate using the Equation 8, where $\hat{w}_t^s$ is the normalized weight of sample s at time t.

$$p(z_{1:t}|y_{1:t}) \approx \sum_{s=1}^{S} \hat{w}_t^s \delta_{z_{1:t}^s}(z_{1:t}) \tag{8}$$

The posterior can be approximated by Equation 9, and for $S \to \infty$ it becames the true posterior.

$$p(z_t|y_{1:t}) \approx \sum_{s=1}^{S} \hat{w}_t^s \delta_{z_t^s}(z_t) \tag{9}$$

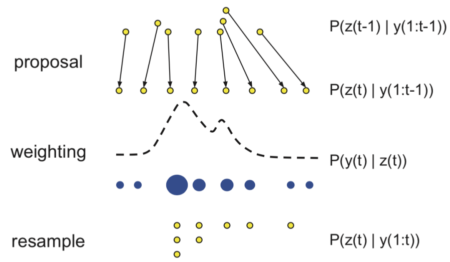

Figure 2. Ilustration of Algoritm 1.

Now, we have to propose for each old sample s extension using $z_t^s \sim q(z_t|z_{t-1}^s, y_t)$ to get w_t^s.

There is a problem called *degeneracy problem* using this propuse, it fails after a few steps because most of the particles will have negligible weight. Thats occurs because we are sampling in a high-dimensional space using a myopic proposal distribution. One way to solve that is to use the resampling step, where its algorithm can be seen in Algorithm 1, and an ilustrative image can be seen in Figure 2.

Algorithm 1 One step of a generic particle filter

for $s = 1 : S$ **do**
    Draw $z_t^s \sim q(z_t|z_{t-1}^s, y_t)$;
    Compute weight $w_t^s \propto w_{t-1}^s \frac{p(y_t|z_t^s)p(z_t|z_{t-1}^s)}{q(z_t|z_{t-1}^s, y_t)}$;
Normalize weights: $w_t^s = \frac{w_t^s}{\sum_{s'} w_t^{s'}}$;
Compute $\hat{S}_{eff} = \frac{1}{\sum_{s=1}^{S}(w_t^s)^2}$;
if $\hat{S}_{eff} < S_{min}$ **then**
    Resample S indices $\pi \sim w_t$;
    $z_t^{:} = z_t^{\pi}$;
    $w_t^s = 1/S$;

Some kind of possible application is in robot localization, where a robot movement itself in a place like an office and try to know where exactly it is given the map of the place. Another example of application is visual object tracking, as we can see in Figure 3, where we want to track the position of an object in sequence of video frames. This method can also be used in time series forecasting.

4.4 Rao-Blackwellised Particle Filtering

This method can be used in models that it's hidden variables can be splitted in a way that the first kind, denoted as z_t, can be analytically integrated out provided we know the values of $q_{1:t}$, which is the second kind, representing $p(z_t|q_{1:t})$ parametrically. A distributional particle s is the name of a particle that is both a value of $q_{1:t}^s$ and a distribution of the form $p(z_t|y_{1:t}, q_{1:t}^s)$.

Using this approach we can reduce the dimensionality of the space, reducing the variance of the estimate.

This method can be applied for example to track moving objects that have piecewise linear dynamics, like an airplane or a missile. Another application is in mobile robotics, in the problem of simultaneous localization and mapping or SLAM, which computation is cubic in the number of landmarks, but with Monte Carlo it becomes linear in the number of particles.

5 Conclusion

In this summary we could learn about Monte Carlo methods, and how it is applied in different kinds of context, from the simple distributions to high dimensional ones.

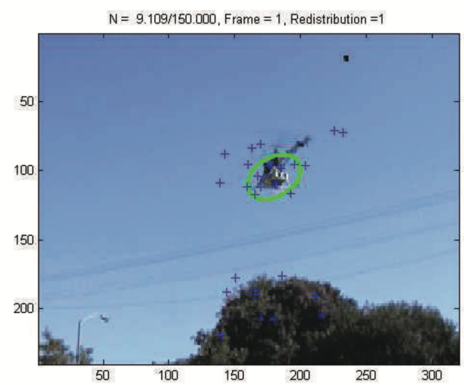

Figure 3. Example of used of Monte Carlo in Image Tracking.