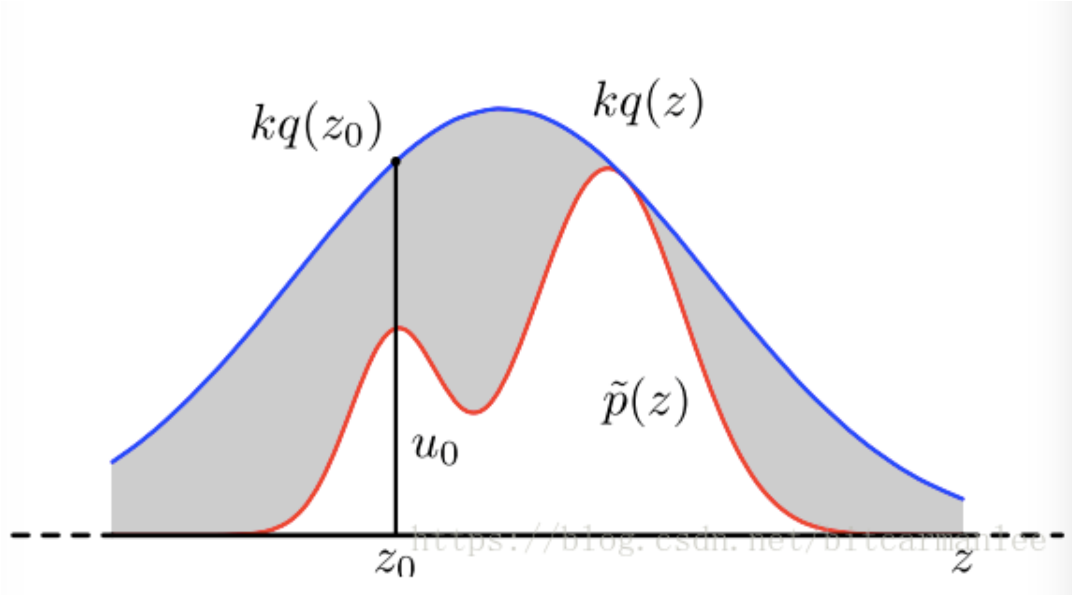
采样是因为数据太大是，为了对整体数据具有的性质进行估算而给出的方法，用于计算期望，方差，梯度都很适用。总而言之就是对全部数据都计算一遍的回避。

反函数采样
我们如果知道一个样本的分布函数，像根据这个分布函数进行采样。那么最最直接的方法就是算该分布函数的累积分布函数的反函数进行坐标的采样。采样出来的坐标用于对样本的选取，完成采样。证明思路如下
设指数分布f1，均匀0-1分布f2，两者的累积分布函数为F1，F2
很显然我们可以得到两个事实
1.因为F单调递增，所以反函数F-1也单调递增
2.F-1（a） < x，a < F（x）
所以对于P（F1-1(a) <= x）== P(a <= F1(x))；P为F2X
由于F1的值域满足0，1所以 P(a <= F1(x)）== F1（x）
即F-1（a）是符合exp分布的，即反函数可以使均匀分布转换为指数分布，利用这个关系就可以生成用于指数分布抽样的数据。

拒绝采样
有时，根据需要采样的分布p生成数据过于困难，所以在知道要生成的数据的分布函数的基础上，选择一个好生成数据的分布q，然后随机生成数据与p进行比较即
均匀分布生成x，计算q（x）-》比较p（x），q（x），如果在p（x） 的范围内，就接受，否则就拒绝



重要性采样

同样使用了一个更好生成数据的分布来进行计算，由于网上的博客都用期望做例子进行推导，下面直接粘贴，然后叙述

$$E[f(x)] = \int_x f(x)p(x)dx$$

如果符合p(x)分布的样本不太好生成，我们可以引入另一个分布q(x)，可以很方便地生成样本。使得

$$\int_x f(x)q(x)dx = \int_x f(x)\frac{p(x)}{q(x)}q(x)dx = \int_x g(x)q(x)dx,$$

$$where\, g(x) = f(x)\frac{p(x)}{q(x)} = f(x)w(x)$$

我们将问题转化为了求g(x)在q(x)分布下的期望！！！

我们称其中的 $w(x) = \dfrac{p(x)}{q(x)}$ 叫做Importance Weight.

https://blog.csdn.net/u011332699/article/details/74298555

看到这里我给出的理解是，为了近似计算该积分就得一一对应相乘，在无x，f的对应关系的情况下，直接计算p（x）不现实。但p分布的数据又难以得到。所以就转变为对f对应的样本的采样。而采样的方式就是带权重采样，权重代表了p的分布关系，而p的计算通过p/q近似得出。

所以算法流程变为了：

1.生成均匀分布的一连串x

2.已知p，设以分布q，两者的分布函数已知，计算p（x）/q（x），作为抽样权重

3.利用该权重，选取f中样本进行计算
所以这个算法的好用与否就取决于p/q能多大程度体现原分布了，同时一串数据方差越稳定数据结果越可信，所以p*f/q方差要尽可能小。而实际中找一个这样的q是难的。
https://www.cnblogs.com/xbinworld/p/4266146.html

MCMC
最有名的采样算法，理论基础为信息论中的马尔科夫的平稳分布。一个可以达到平稳状态的转移矩阵，在经过有限次转移后，其分布会达到稳定。
https://blog.csdn.net/bitcarmanlee/article/details/82795137
这里的博客说的很详细了，设要抽样的分布为未知的，那么就把它看做一连串转移矩阵过程中的收敛结果，所以我们需要根据细致平稳条件，将转移矩阵写好就行，得到了alpha*q，这里q为设的一个分布，alpha代表为了将alpha*q作为合适的转移矩阵，而加上的scale值。在q已知的情况下，alpha的计算就成了问题，所以，这个采样核心就是alpha的计算

> 由于一般情况下，目标平稳分布$\pi(x)$和某一个马尔科夫链状态转移矩阵Q不满足细致平稳条件
>
> $$\pi(i)Q(i,j) \neq \pi(j)Q(j,i)$$
>
> 为了使细致平稳条件成立，可以这样做：
>
> $$\pi(i)Q(i,j)\alpha(i,j) = \pi(j)Q(j,i)\alpha(j,i)$$
>
> 为了两边相等，可以取
>
> $$\alpha(i,j) = \pi(j)Q(j,i)$$
>
> $$\alpha(j,i) = \pi(i)Q(i,j)$$
>
> 这样，转移矩阵P最后就变成为:
>
> $$P(i,j) = Q(i,j)\alpha(i,j)$$

也就是说，目标矩阵P可以通过任意一个马尔科夫链状态转移矩阵Q乘以$\alpha(i,j)$得到，$\alpha(i,j)$一般称为接受率，取值在[0,1]之间，即目标矩阵P
可以通过任意一个马尔科夫链状态转移矩阵Q以一定的接受率获得。与前面的拒绝采样比较，拒绝采样是以一个常用的分布通过一定的拒绝率获得一个非常见的分布。而MCMC则是以一个常见的马尔科夫链状态转移矩阵Q通过一定的拒绝率得到目标转移矩阵P，两者思路类似。

总结一下上面的过程，MCMC采样算法的过程为：
1.初始化马尔科夫链初始状态为$X_0 = x_0$
2.循环采样过程如下：
2.1 第t个时刻马尔科夫链状态为$X_t = x_t$，由转移矩阵$Q(x)$中采样得到$y \sim q(x|x_t)$
2.2 从均匀分布采样$u \sim Uniform[0,1]$
2.3 如果$u < \alpha(x_t, y) = p(y)q(y, x_t)$，则接受转移$x_t \to y$, $X_{t+1} = y$
2.4 否则不接受转移，即$X_{t+1} = x_t$

而这种算法的一个弊端就是，alpha可能太小，导致遍历时间延长
给出改进

假设 α(i,j)=0.1,α(j,i)=0.2, 此时满足细致平稳条件，于是
p(i)q(i,j)×0.1=p(j)q(j,i)×0.2

上式两边扩大5倍，我们改写为
p(i)q(i,j)×0.5=p(j)q(j,i)×1

看，我们提高了接受率，而细致平稳条件并没有打破！这启发我们可以把细致平稳条件中的α(i,j),α(j,i) 同比例放大，使得两数中最大的一个放大到1，这样我们就提高了采样中的跳转接受率。所以我们可以取

$$\alpha(i,j) = min\{\frac{p(j)q(j,i)}{p(i)q(i,j)}, 1\}$$

经过以上细微的变化，我们就得到了如下教科书中最常见的 Metropolis-Hastings 算法。
1.初始化马尔科夫链初始状态为$X_0 = x_0$
2.循环采样过程如下：
2.1 第t个时刻马尔科夫链状态为$X_t = x_t$，由转移矩阵$Q(x)$中采样得到$y \sim q(x|x_t)$
2.2 从均匀分布采样$u \sim Uniform[0,1]$
2.3 如果$u < \alpha(x_t, y) = min\{\frac{p(j)q(j,i)}{p(i)q(i,j)}, 1\}$，则接受转移$x_t \to y$, $X_{t+1} = y$
2.4 否则不接受转移，即$X_{t+1} = x_t$