**采样方法**

# 采样目的

一般来说采样的目的是评估一个随机变量函数在某个分布上的期望值。







获得服从指定概率分布的样本。

# 采样方法

## 反向采样（Inverse Sampling）

如果随机变量服从概率密度分布函数为的连续分布，且这个分布可以计算累积分布函数(Cumulative Distribution Function, CDF)，则可以通过计算CDF的反函数，获得采样。

指数分布：





[0,1]均匀分布：

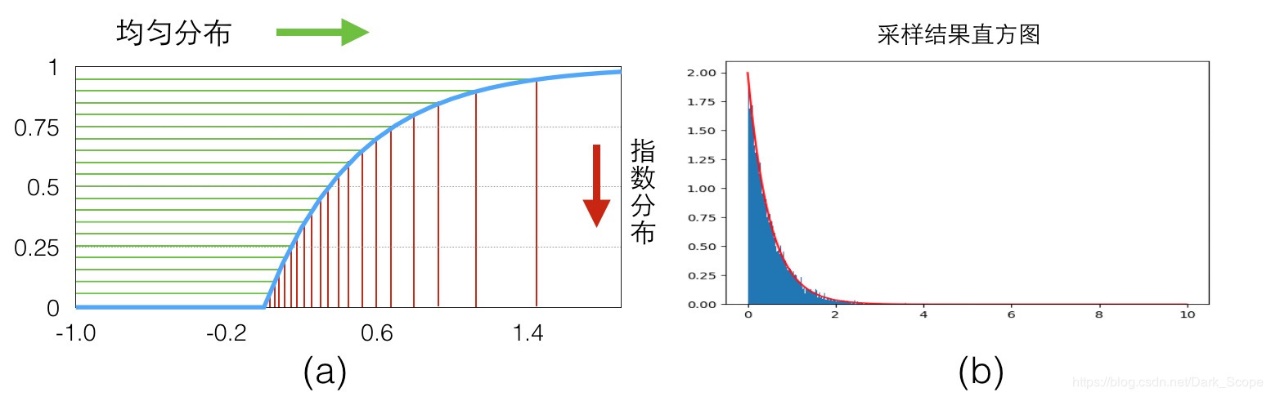


因为单调递增，所以单调递增，值域为[0,1)。假设



根据的定义，它是指数分布的累积分布函数，所以上面这个公式说明符合指数分布。通过的反函数将一个0到1均匀分布的随机数转换成了符合指数分布的随机数。





**Note：需要能求出要采样分布的累积分布函数的反函数**

## 接受-拒绝抽样（Acceptance-Rejection Sampling)）

假设想要采集出来的样本需要服从某一个分布，其实就是希望样本在其概率密度函数高的地方出现得更多。我们可以从均匀分布随机生成一个样本，按照一个正比于的概率接受这个样本，也就是说虽然是从均匀分布随机采样，但留下的样本更有可能是高的样本。

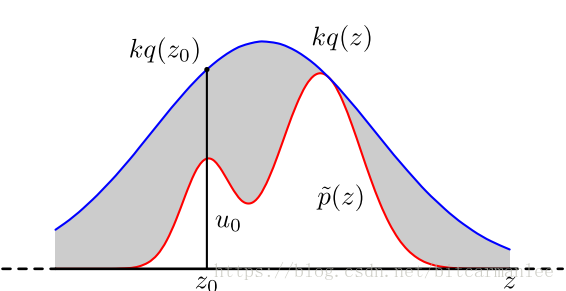
假设方便进行采样的分布函数为，满足：



依据产生的样本为，则这个样本被接受的概率为：



实际上和越接近，其接受概率越高，采样效率也越高。



**Note：找到一个合适的****往往很困难，接受概率有可能会很低。**

在**高维**的情况下，Rejection Sampling有两个问题：

* 合适的分布很难找
* 很难确定一个合理的值，导致拒绝率很高。

## 重要性采样（Importance Sampling）

符合分布的样本不太好生成，我们可以引入另一个分布，可以很方便地生成样本。





## 马尔可夫链蒙特卡罗采样

