# Importance sampling

Renhe W.

## Contents



## List of Figures



## List of Tables

# 1 Importance Sampling

重要性抽样（Importance Sampling）是一种用于估计概率分布性质的统计方法，特别是在计算期望值或概率密度函数的归一化常数时非常有用. 它通过从一个不同的提议分布中抽样，来估计原始分布的性质. 重要性抽样是最重要的方差减小技术之一。这种技术对于估计罕见事件的概率（见第 10 章）特别有用. 标准设置是估计一个量:

$$\ell=\mathbb{E}_f H(\mathbf{X})=\int H(\mathbf{x}) f(\mathbf{x}) \mathrm{d} \mathbf{x}, \tag{1.1}$$

其中 $H$ 是一个实值函数，$f$ 是随机向量 $\mathbf{X}$ 的概率密度，称为名义概率密度。下标 $f$ 被添加到期望算子以表示它是相对于密度 $f$ 进行的.

令 $g$ 是另一个概率密度，使得 $Hf$ 被 $g$ 所主导. 也就是说，$g(\mathbf{x})=0 \Rightarrow H(\mathbf{x}) f(\mathbf{x})=0$。使用密度 $g$，我们可以表示 $\ell$ 如下：

$$\ell=\int H(\mathbf{x}) \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) \mathrm{d} \mathbf{x}=\mathbb{E}_g H(\mathbf{X}) \frac{f(\mathbf{X})}{g(\mathbf{X})}. \tag{1.2}$$

因此，如果 $\mathbf{X}_1, \ldots, \mathbf{X}_N \sim_{\mathrm{idd}} g$，那么

$$\widehat{\ell}=\frac{1}{N} \sum_{k=1}^N H\left(\mathbf{X}_k\right) \frac{f\left(\mathbf{X}_k\right)}{g\left(\mathbf{X}_k\right)} \tag{1.3}$$

是 $\ell$ 的无偏估计。这个估计被称为重要性抽样估计，$g$ 被称为重要性抽样密度. 密度比值为：

$$W(\mathbf{x})=\frac{f(\mathbf{x})}{g(\mathbf{x})}, \tag{1.4}$$

被称为似然比（本书定义），因为似然通常在统计学中被视为参数的函数（见第 B.2 节）. 通常也可以称为重要性权重（importance weight）.

**算法 1.1** (重要性抽样估计). 重要性抽样估计的具体算法步骤为：

1. *选择一个主导 $Hf$ 的重要性抽样密度 $g$.*
2. *生成独立同分布的样本 $\mathbf{X}_1, \ldots, \mathbf{X}_N \stackrel{idd}{\sim} g$，并令 $Y_i=H\left(\mathbf{X}_i\right) f\left(\mathbf{X}_i\right) / g\left(\mathbf{X}_i\right), i=1, \ldots, N$.*

*3.* 通过 $\widehat{\ell}=\bar{Y}$ 估计 $\ell$，并确定一个近似的 $1-\alpha$ 置信区间，如下所示：

$$\left(\widehat{\ell}-z_{1-\alpha/2}\frac{S}{\sqrt{N}},\widehat{\ell}+z_{1-\alpha/2}\frac{S}{\sqrt{N}}\right),$$

其中 $z_\gamma$ 表示 $\mathrm{N}(0,1)$ 分布的 $\gamma$ 分位数，$S$ 是 $Y_1,\ldots,Y_N$ 的样本标准差。

【证明】〈无偏性〉

*Proof.* 假设我们从提议分布 $g$ 中抽取了 $N$ 个独立同分布的样本 $\mathbf{X}_1,\ldots,\mathbf{X}_N$ ，然后计算每个样本的加权值 $H\left(\mathbf{X}_k\right)\frac{f(\mathbf{X}_k)}{g(\mathbf{X}_k)}$ 。那么, $\ell$ 的估计值 $\hat{\ell}$ 为:

$$\hat{\ell}=\frac{1}{N}\sum_{k=1}^{N}H\left(\mathbf{X}_k\right)\frac{f\left(\mathbf{X}_k\right)}{g\left(\mathbf{X}_k\right)}$$

由于期望的线性性质，我们有:

$$\mathbb{E}_g[\hat{\ell}]=\mathbb{E}_g\left[\frac{1}{N}\sum_{k=1}^{N}H\left(\mathbf{X}_k\right)\frac{f\left(\mathbf{X}_k\right)}{g\left(\mathbf{X}_k\right)}\right]=\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}_g\left[H\left(\mathbf{X}_k\right)\frac{f\left(\mathbf{X}_k\right)}{g\left(\mathbf{X}_k\right)}\right]$$

由于每个 $\mathbf{X}_k$ 都是从 $g$ 中独立抽取的，它们有相同的分布，所以:

$$\mathbb{E}_g[\hat{\ell}]=\mathbb{E}_g\left[H(\mathbf{X})\frac{f(\mathbf{X})}{g(\mathbf{X})}\right]=\mathbb{E}_f[H(\mathbf{X})]$$

即 $\hat{\ell}$ 是 $\ell$ 的无偏估计。 □

## 2 一个正态分布简单的例子

**Example 2.1** (Toy Example)**.** 估计一个标准正态分布 *(*目标分布*)* 下的函数 $H(x)=x^2$ 的期望值. 概率密度函数 $f(x)$ 为 $\mathcal{N}(0,1)$ ，即均值为 *0*，标准差为 *1* 的正态分布.

求解过程:

- 估计目标

  我们的目标是估计 $\mathbb{E}_f[H(X)]=\mathbb{E}_f\left[X^2\right]$.

- 提议分布 $g(x)$
  为简单示范，这里选择了一个均值为 2 的正态分布作为提议分布 $g(x)$ ，即 $\mathcal{N}(2,1)$.

- 抽样和权重计算
  我们从提议分布 $g(x)$ 中生成了 $N$ 个样本。对于每个样本 $x_i$ ，我们计算了重要性权重 $w_i$ ，这是目标分布 $f(x_i)$ 和提议分布 $g(x_i)$ 的概率密度之比，即 $w_i = \frac{f(x_i)}{g(x_i)}$.

- 估计期望值
  重要性抽样估计器 $\hat{\ell}$ 是加权平均值，即 $\hat{\ell} = \frac{\sum_{i=1}^{N} w_i H(x_i)}{\sum_{i=1}^{N} w_i}$，这里，$\hat{\ell}$ 是 $H(X)$ 在目标分布 $f(x)$ 下的估计期望值.

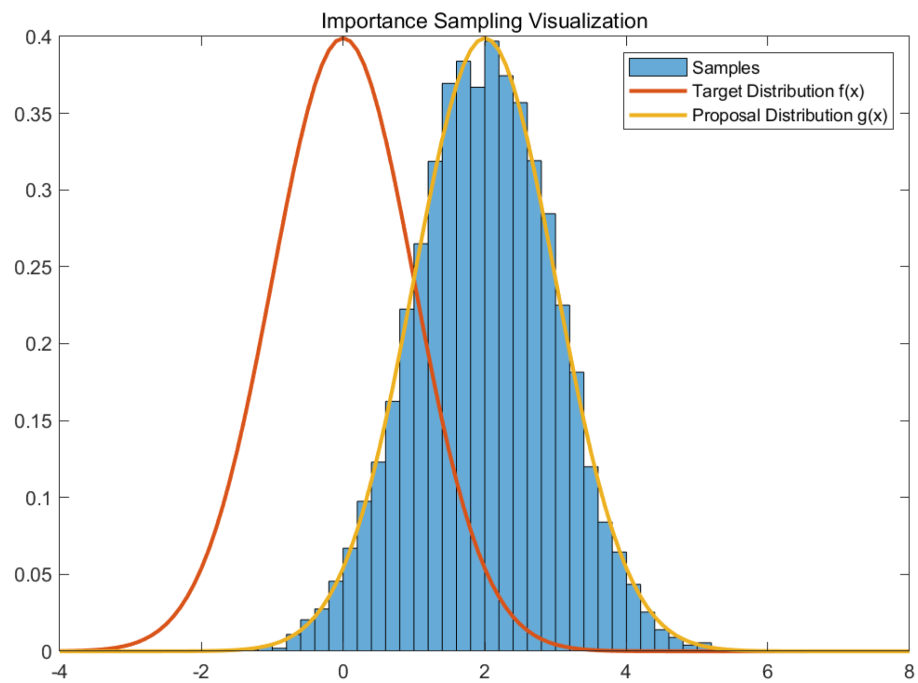


**图 1.** 一个正态分布简单的例子

# References