# 机器学习
# Machine Learning

北京航空航天大学计算机学院
School of Computer Science and Engineering, Beihang University
刘庆杰  陈佳鑫

2025年春季学期
Spring 2025

部分内容来源于C. Bishop和A. NG等人的课程以及互联网资源

# 11.1 什么是采样?

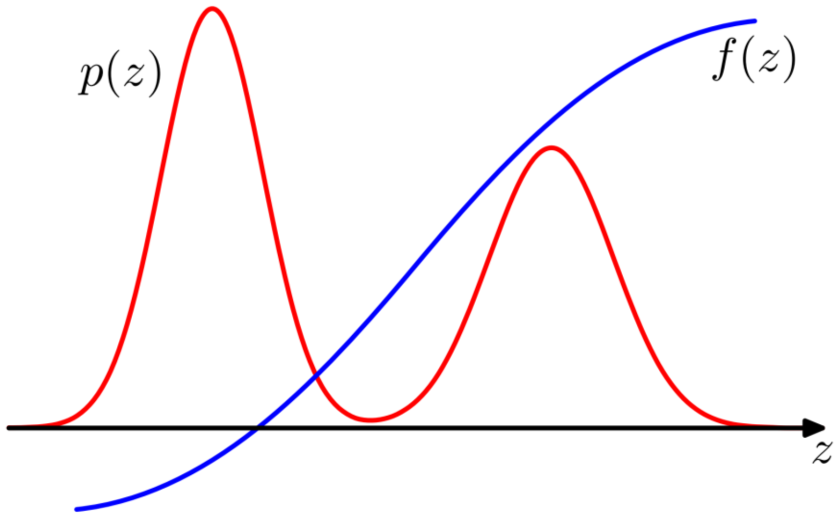
- 采样的定义
- 采样的应用

# 什么是采样?

## ● 采样的定义

- ➢ 采样是一种用有限个样本点来近似总体分布，并刻画总体分布中不确定性的方法
- ➢ **基本思想：对于满足分布$p(z)$的**随机变量$Z$，从分布$p(z)$中独立抽取一组样本$z(l)$(其中$l = 1, 2, \dots, L$)，其随机变量期望值可以**近似**为一个有限的和：

$$\mathbb{E}[f(z)] = \int f(z)p(z)\,dz \approx \frac{1}{L}\sum_{l=1}^{L} f(z(l))$$



# 采样的应用

- 布丰投针问题

  - 1777年法国科学家布丰（Buffon）提出一种利用**采样思想**计算圆周率π的方法—随机投针法。

1、在平面上画上间距为a的平行线

2、取一根长度为$l$（$l<a$）的针，

随机投掷于平面上，针与任一线相交的概率$p=\frac{2l}{\pi a}$

π的值可计算为: $\pi=\frac{2l}{ap}\approx\frac{2l}{a}(\frac{N}{n})$

$N$为总的投针次数，$n$为与平行线相交次数

  - 布丰投针问题是采样方法中**蒙特卡洛方法**的早期实例，它为后来的许多数值计算方法奠定了基础。

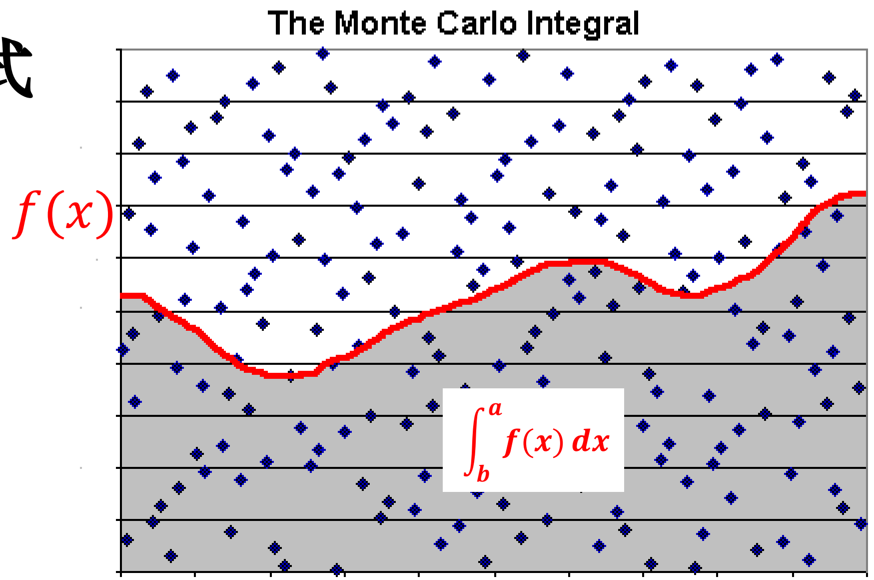
# 采样的应用

- 复杂函数定积分的计算

  - 对于一个解析形式那以求得的积分$\int_b^a f(x)\,dx$，可以通过采样方法中蒙特卡洛方法（Monte Carlo Method）求其近似值。
  - 蒙特卡洛方法基本思想是把要求解的问题转化为求解某一随机事件的概率，或者某个随机变量的期望
  - 对于积分$\int_b^a f(x)\,dx$，它的解析值可通过以下方式近似求得：

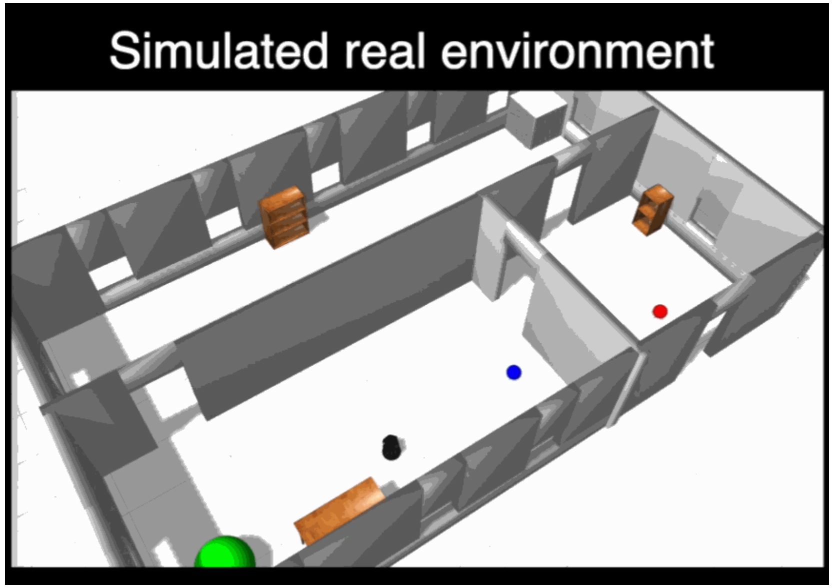
$$\int_b^a f(x)\,dx = \int_b^a \frac{f(x)}{q(x)} q(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{q(x_i)}$$

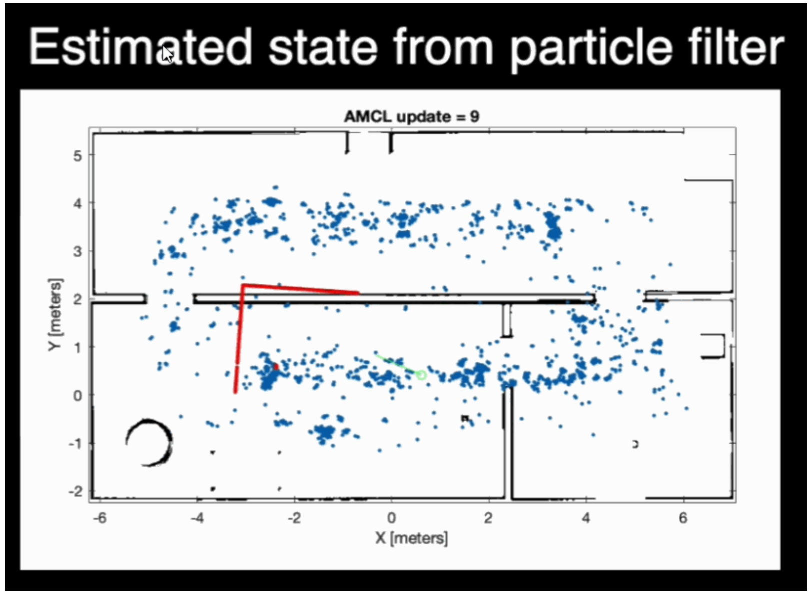
  - 其中$q(x)$是某种容易采样的分布，$x_i$是服从$q(x)$分布的随机样本



# 采样的应用

- 机器人导航：是帮助机器人确定自身位置、规划路径，并安全到达目的地的技术和方法。
  - ➢ 在机器人导航中广泛应用的**粒子滤波器**是一种基于**重要性采样**的算法。
  - ➢ 基本思想：通过为每个粒子分配**权重**来反映其对观测数据的适应程度，并通过**重采样**步骤保留高权重粒子，从而有效地估计系统的状态分布



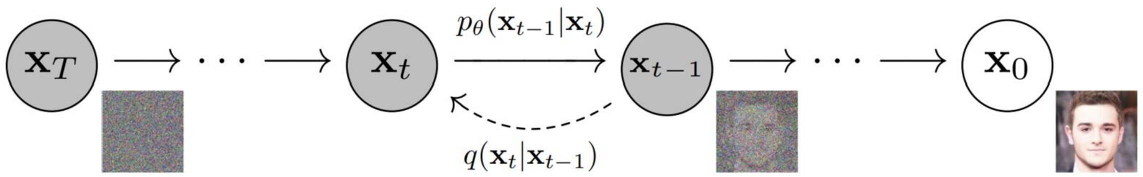

https://www.mathworks.com/videos/autonomous-navigation-part-2-understanding-the-particle-filter-1594903924427.html

# 采样的应用

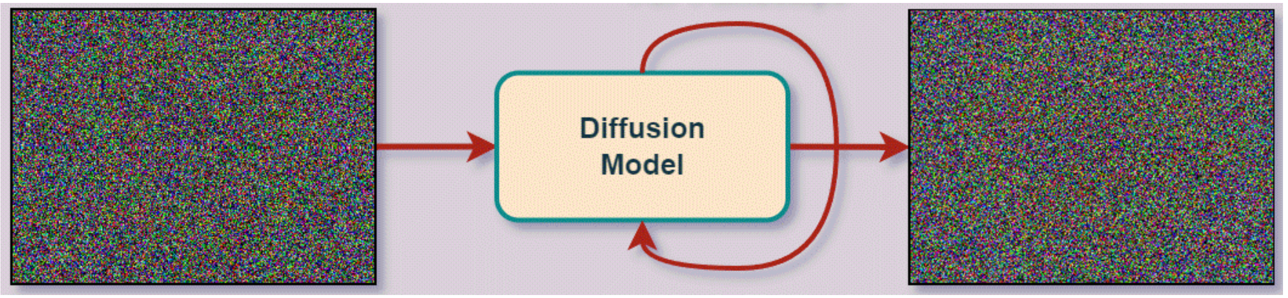
## ● 扩散模型中的噪声采样

- ➢ 扩散模型是一类用于文生图的新兴生成模型，它们通过逐步去噪过程生成图像。
- ➢ 基本思想：在扩散模型中，图像生成从一个**完全随机的噪声图像**开始，模型逐步**从一个噪声分布中利用马尔可夫蒙特卡洛采样**来决定去噪的程度和方向，最终产生高质量的图像。

➢ 去噪过程




➢ 应用动画



# 采样的方法

- 基本采样法
- 拒绝采样法
- 重要性采样法
- 马尔可夫蒙特卡洛采样法

# 11.2 基本采样法

- 思想
- 算法步骤
- 实例
- 局限性

# 基本采样法 (Basic Sampling)

- 思想：从均匀分布中通过函数变换产生目标分布
  - 均匀分布(Bootstrap Sampling)：

$$p(z)=1 \quad z\in(0,1) \tag{1}$$

  - 通过函数变换：

$$y=f(z)$$

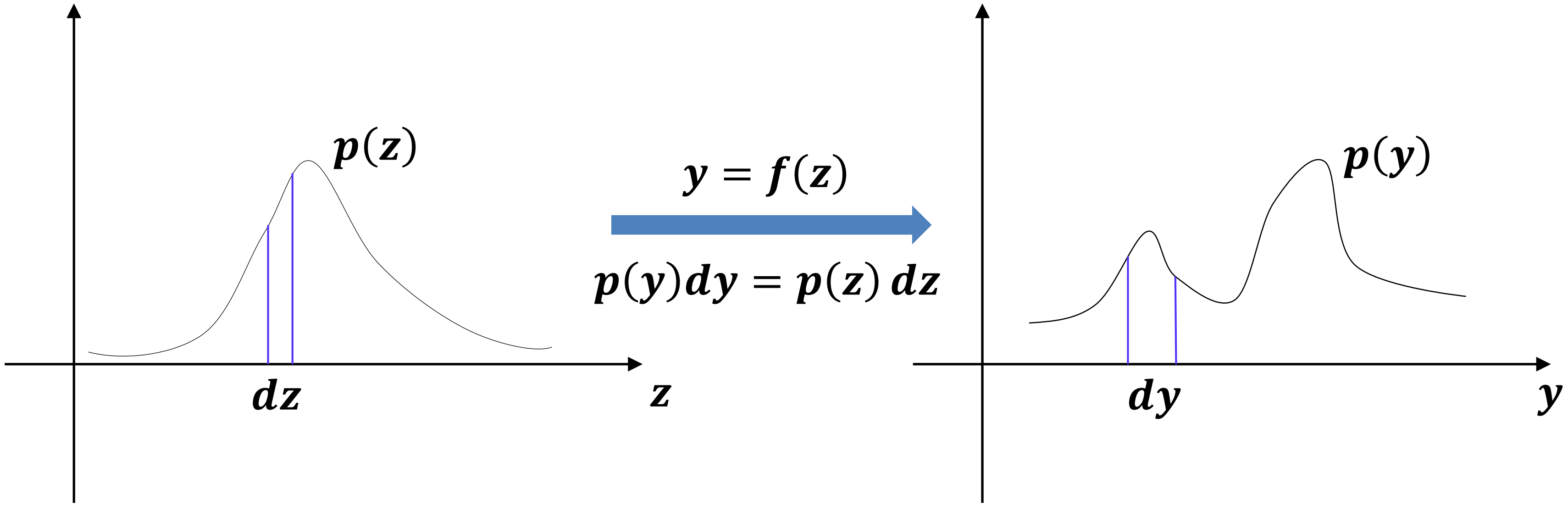
  - 产生目标分布：

$$p(y)$$

- 关键问题：如何构造函数变换$f$?

# 基本采样法 (Basic Sampling)

- **构造函数变换*f***
  - 设$z$有概率密度函数$p(z)$，而$y = f(z)$构成$z$到$y$的一一对应变换




  - 则$y$的概率密度函数为 $$p(y) = p(z)\left|\frac{dz}{dy}\right| \quad (2)$$

# 基本采样法 (Basic Sampling)

- **构造函数变换*f***
  - 设$z$有概率密度函数$p(z)$，而$y=f(z)$构成$z$到$y$的一一对应变换

    则$y$的概率密度函数为 $$p(y)=p(z)\left|\frac{dz}{dy}\right| \qquad (2)$$
  - 结合(1)，(2)有

$$z=\boxed{h(y)\equiv\int_{-\infty}^{y}p(\hat{y})\,d\hat{y}} \rightarrow \boxed{\text{累积分布函数（CDF）}}$$

则有：

$$y=\underbrace{\boxed{h^{-1}}}_{f}(z)$$

# 基本采样法 (Basic Sampling)

## 基本采样的算法流程

**输入：**

目标分布 $p(y)$

均匀分布 $z \sim U(0,1)$

**过程：**

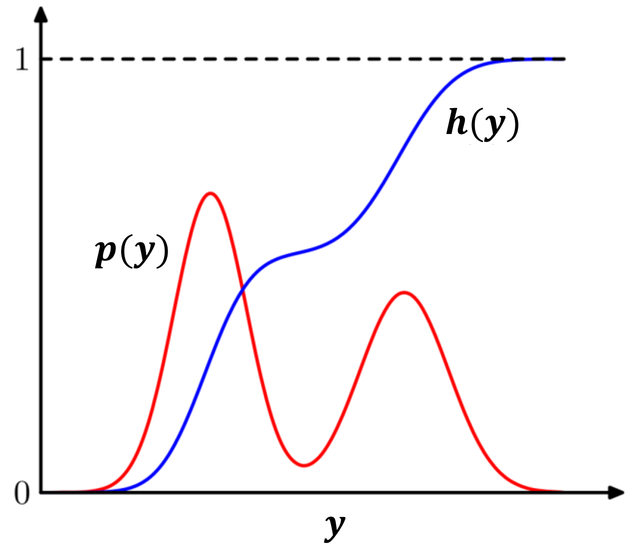
1: 计算$p(y)$的累积分布函数$h(y)$

2: 计算累积分布函数的逆函数$h^{-1}$

3: 从$(0,1)$均匀分布中产生$z_0$

4: 计算对应的$y_0 = h^{-1}(z_0)$

5: 重复步骤3-4至采样完成



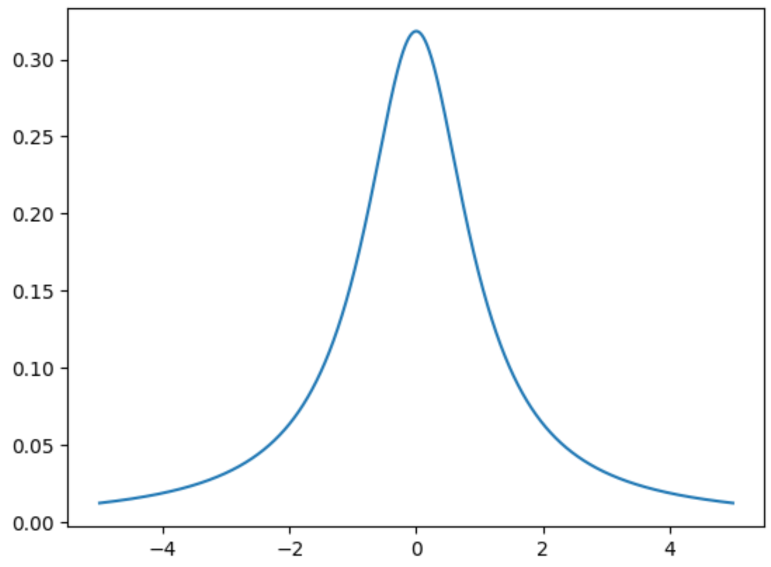
# 基本采样法 (Basic Sampling)

- 例子：标准柯西分布

已知 $z \sim U(0,1)$

求 $y = f(z)$

使 $p(y) = \frac{1}{\pi}\frac{1}{1+y^2}$




➡ 计算目标分布$p(y)$的累积分布函数$h(y)$

$$h(y) = \int_{-\infty}^{y} p(y)dy = \frac{1}{\pi}\arctan(y) + \frac{1}{2}$$

➡ 计算$h(y)$的逆函数

$$y = h^{-1}(z) = \pi\tan(z - \frac{1}{2})$$

# 基本采样法 (Basic Sampling)

- **例子：指数分布**

已知 $z \sim U(0,1)$

求 $y = f(z)$

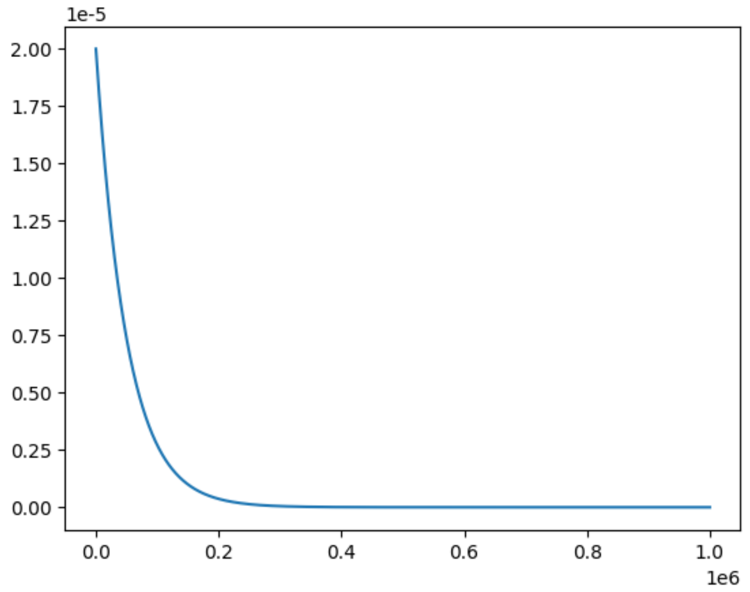
使 $p(y) = \lambda \exp(-\lambda y)$

➡ **计算目标分布$p(y)$的累积分布函数$h(y)$**

$$h(y) = \int_{-\infty}^{y} p(y)dy = 1 - \exp(-\lambda y)$$

➡ **计算$h(y)$的逆函数**

$$y = h^{-1}(z) = -\lambda^{-1}\ln(1-z)$$



# 基本采样法 (Basic Sampling)

- 基本采样法无法实现对累积分布函数及其逆函数难以求解的目标分布进行采样
- 例子：标准正态分布

$$p(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

➢ 计算其累积分布函数:

$$\begin{aligned} h(y) &= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \\ &= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy + \int_{0}^{y} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \\ &= \frac{1}{2} + \boxed{\frac{1}{2}\mathrm{erf}\left(\frac{y}{\sqrt{2}}\right)} \rightarrow \boxed{\text{高斯误差函数}} \end{aligned}$$

➢ 由于一维标准正态分布的累积分布函数中含有高斯误差函数$\mathrm{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-y^2}\, dy$，无法求其逆函数，不适用于基本采样法直接进行采样

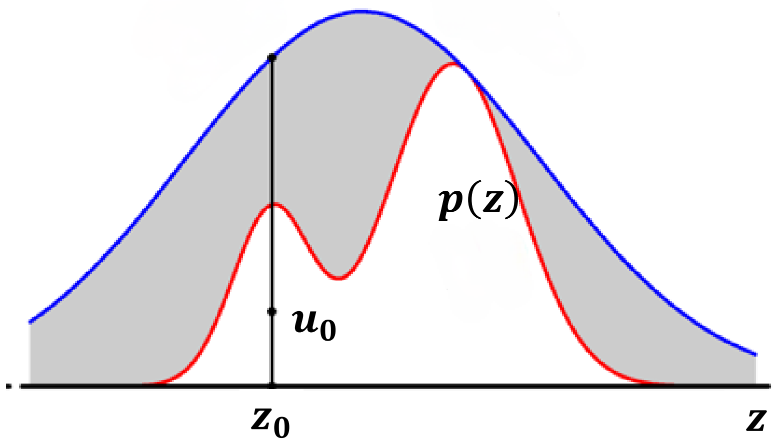
注：虽然基本采样法不能直接用于一维标准正态分布的采样，但可通过对二维正态分布映射变换来获得其样本集。而二维标准正态分布的采样可以通过Box-Muller变换结合基本采样法来实现。Bishop, Christopher M. "Pattern recognition and machine learning." *Springer google schola* 2 (2006): 1122-1128.

# 11.3 拒绝采样法

- 思想
- 算法步骤
- 实例
- 局限性

# 拒绝采样法 (Rejection Sampling)

- 拒绝采样法利用易采样的分布，通过**接受-拒绝**的思想来近似目标分布，避免了基本采样法所需的累积分布函数及其逆函数的求解
- 拒绝采样法思路
  - 设目标分布的概率密度函数为$p(z)$，借助一个容易采样的分布$q(z)$去逼近$p(z)$， $q(z)$称为**建议分布**
  - 从$q(z)$中采集产生点$(z_0, u_0)$
  - 若$u_0 < p(z_0)$，则返回采样点$(z_0, u_0)$，否则**拒绝**采样点

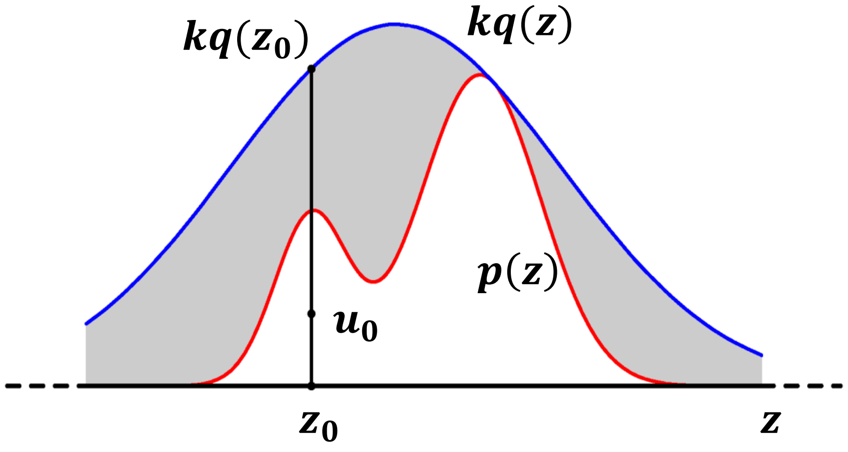


# 拒绝采样法 (Rejection Sampling)

- **关键问题：建议分布**
  - 应满足
    - □ 建议分布应覆盖目标分布的样本集，以确保能够生成符合目标的样本
    - □ 建议分布的形状应与目标分布相似并尽可能逼近目标
  - 为实现上述要求，找到一个常数$k$使得对任意的$z$有

$$\frac{p(z)}{q(z)} \leq k$$

此时，$kq(z)$称作对比函数



# 拒绝采样法 (Rejection Sampling)

- **拒绝采样法的算法流程**

---

**输入：**

目标分布 $p(z)$

**过程：**

1: 找到容易采样的建议分布$q(z)$

2: 找到$k$，使得对任意的$z$有 $\frac{p(z)}{q(z)} \leq k$

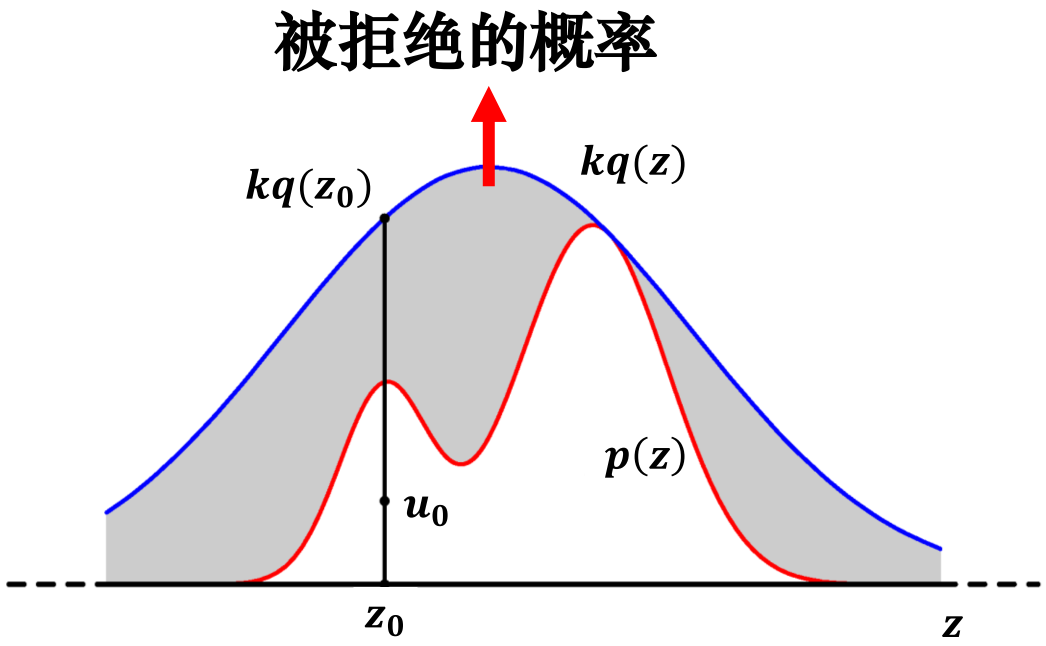
3: 从$q(z)$中采样$z_0$

4: 从$(0,1)$均匀分布中采样$u_0$

5: 若$u_0 \leq \frac{p(z_0)}{kq(z_0)}$，则返回$z_0$，否则拒绝采样

6: 重复步骤3-5至采样完成

---




$$p(accept) = \int \frac{p(z)}{kq(z)} q(z)\, dz = \frac{1}{k} \int p(z)\, dz$$

# 示例-标准正态分布

- 已知目标分布为标准正态分布，即

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

- 取建议分布为柯西分布

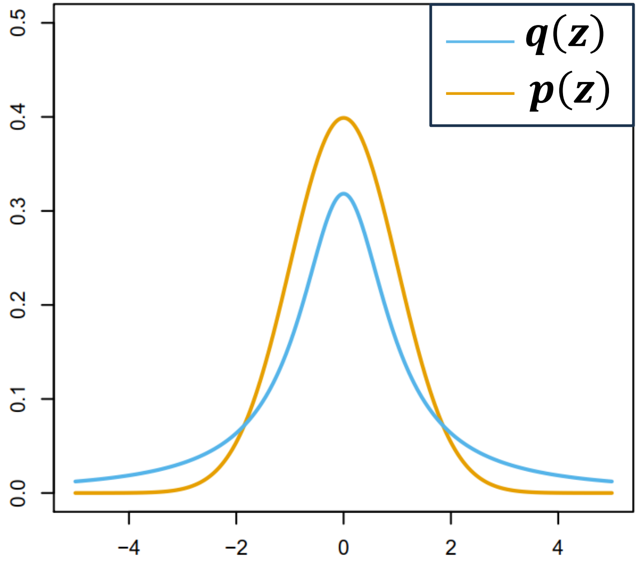
$$q(z) = \frac{1}{\pi(1+z^2)}$$

- 则可求得此时

$$k = \sqrt{2\pi}\, e^{-1/2}$$

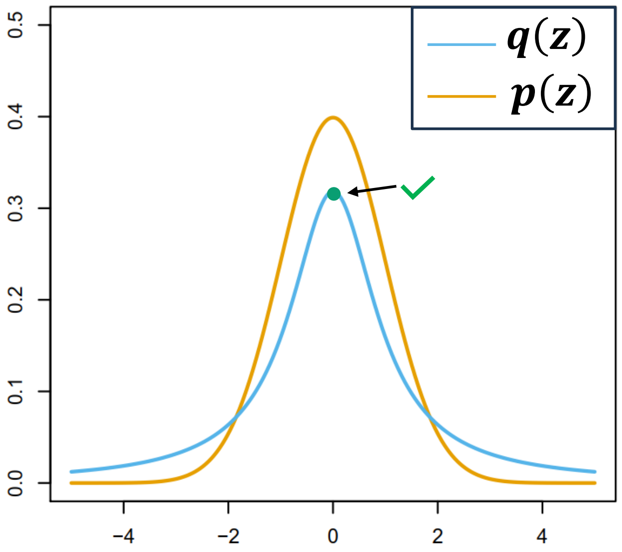
- 代入可得

$$\frac{p(z)}{kq(z)} = \frac{(1+z^2)e^{-z^2/2}}{2e^{-1/2}}$$

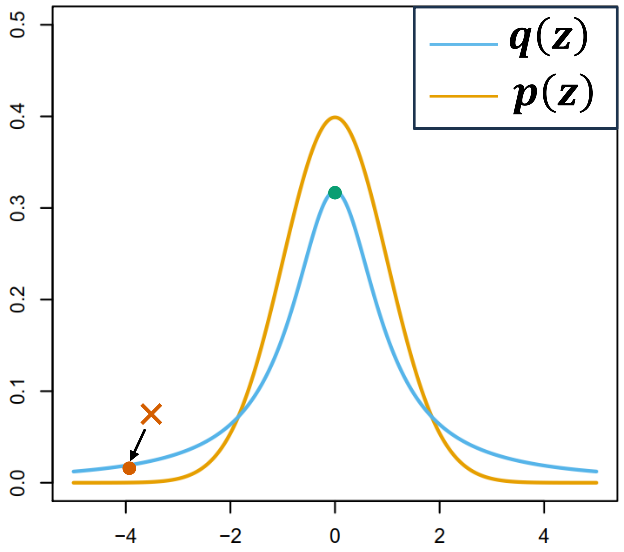


# 示例-标准正态分布

$$\frac{p(z)}{kq(z)}=\frac{(1+z^2)e^{-z^2/2}}{2e^{-1/2}}$$

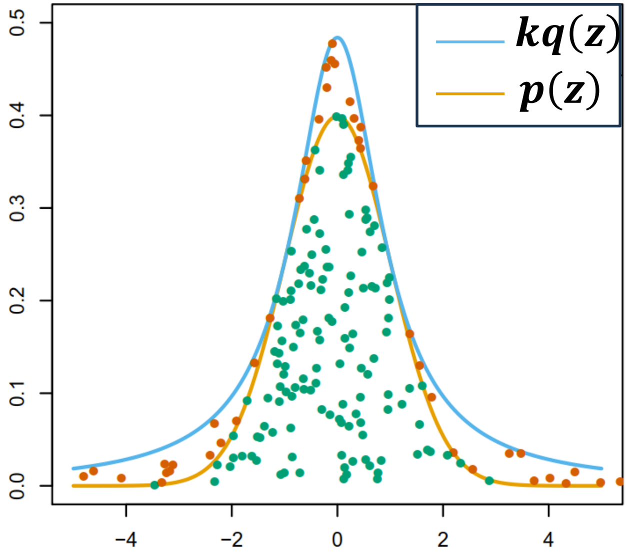



从$q(z)$中采样$z_0=0$，
从 $U\sim(0,1)$中采样$u_0=0.1$
$u_0<\frac{p(z_0)}{kq(z_0)}=0.824$，接受采样




从$q(z)$中采样$z_0=-4$，
从 $U\sim(0,1)$中采样$u_0=0.5$
$u_0>\frac{p(z_0)}{kq(z_0)}=0.004$，拒绝采样

…

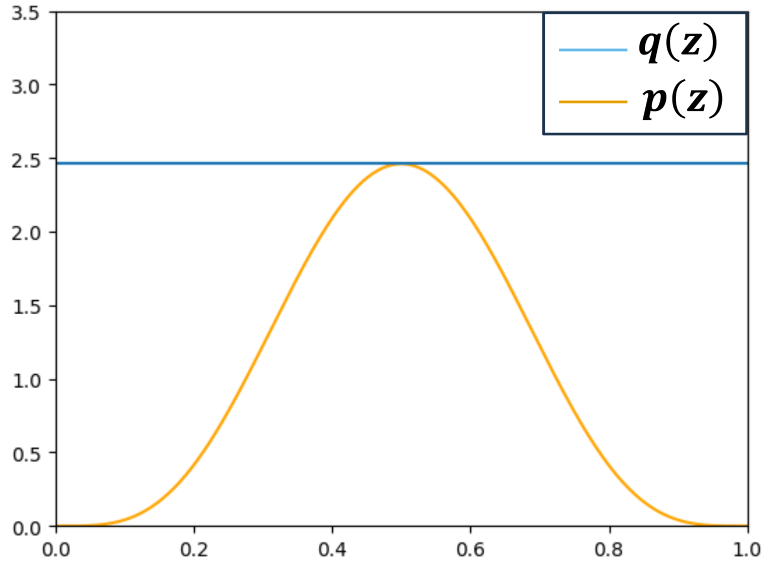



迭代至采样完成。
绿点代表所得样本集。

# 示例-Beta $(\alpha,\beta)$分布

- 已知目标分布为Beta$(\alpha,\beta)$分布， 即：

$$p(z)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}z^{\alpha-1}(1-z)^{\beta-1},\Gamma\text{为伽玛函数}$$




- 取建议分布为均匀分布$U(0,1)$
- 则可设$k=\max(p(z))$，得

$$k=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\underbrace{\boxed{\left(\frac{\alpha-1}{\alpha+\beta-2}\right)^{\alpha-1}\left(\frac{\beta-1}{\alpha+\beta-2}\right)^{\beta-1}}}_{k'}$$

- 代入可得

$$\frac{f(z)}{kq(z)}=\frac{z^{\alpha-1}(1-z)^{\beta-1}}{k'}$$

# 拒绝采样法 (Rejection Sampling)

- **拒绝采样法的局限性**
  - 当接受率较低时，导致被拒绝的样本较多，从而导致采样效率较低
  - 对于高维的复杂分布$p(z_1, z_2, \ldots, z_n)$，找到合适的建议分布$q(z)$和$k$非常困难
  - 对于高维的复杂分布，只能得到条件分布$p(z_{n-1}|z_n)$，但却无法得到高维分布的联合概率分布，无法采用拒绝采样法

# 11.4 重要性采样

- 重要性采样

# 重要性采样

- **重要性采样 (Importance Sampling)**
  - 用于在概率模型中**估计**某些**难以直接计算的期望值**
  - 特别适用于当目标分布**难以直接从中抽样**，或者抽样成本非常高时

**期望估计**：对于变量 $z \sim p(z)$，关于 $z$ 的函数 $f(z)$，预测 $f(z)$ 的关于 $Z$ 的期望：

$$\mathbb{E}(f) = \int f(z)p(z)\mathrm{d}z \quad \Rightarrow \quad \mathbb{E}(f) \simeq \frac{1}{L}\sum_{l=1}^{L} f\left(z^{(l)}\right) \quad z^{(l)} \sim p(z)$$

**依赖于能够从分布 $p(z)$ 中抽样，不具有普适性**

# 重要性采样

- **重要性采样 (Importance Sampling)**

  基本思想：与拒绝采样类似，重要性采样使用一个易于抽样的建议分布来代替目标分布，通过加权的方式调整样本，使其反映目标分布的特性，近似目标分布

  重要性采样：引入易于抽取样本的建议分布 $q(z)$，可以将期望表示为从 $q(z)$ 中抽取的样本集合 $\{z^{(l)}\}$ 的有限和形式：

重要性权重

$$\mathbb{E}(f) = \int f(z)p(z)\mathrm{d}z = \int f(z)\frac{p(z)}{q(z)}q(z)\mathrm{d}z \simeq \frac{1}{L}\sum_{l=1}^{L}\boxed{\frac{p(z^{(l)})}{q(z^{(l)})}}f(z^{(l)})$$

重要性权重矫正了从错误分布中采样而引入的偏差

# 重要性采样

## ● 重要性采样 – 算法流程

**输入：** 函数 $f(z)$; 目标分布 $p(z)$; 样本数 $L$.

**过程：**

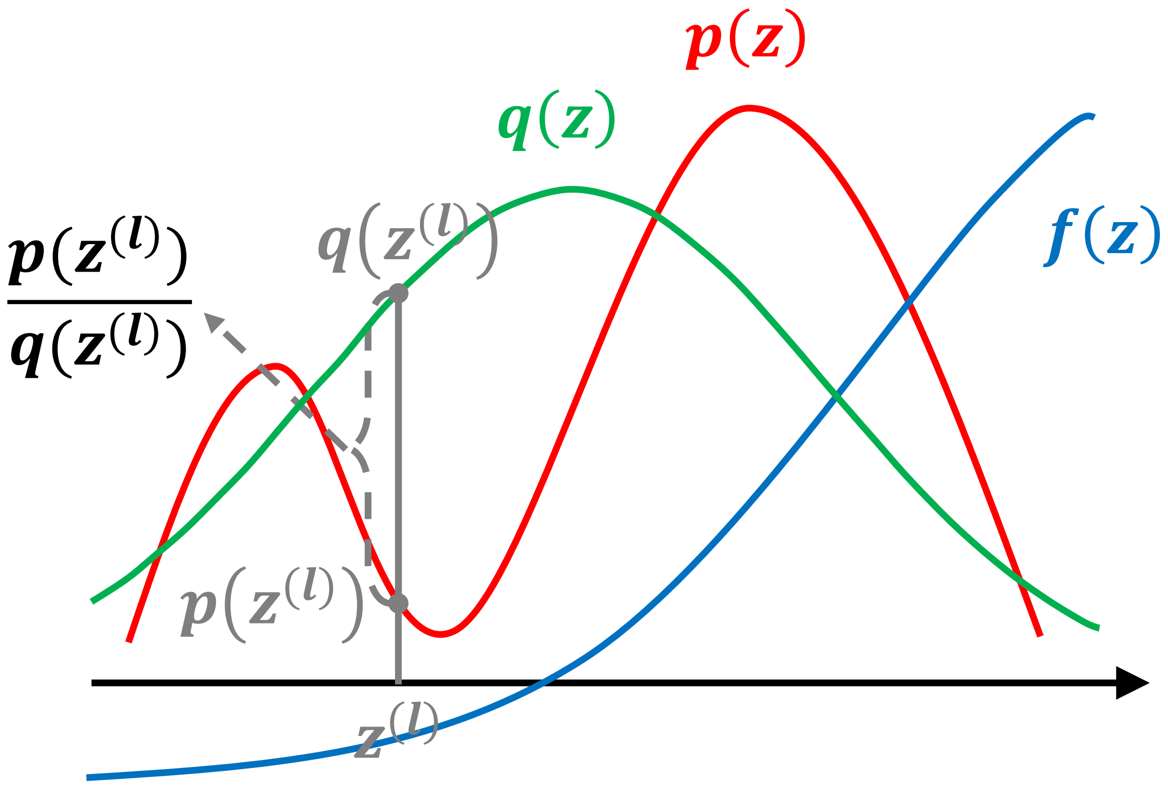
1: 找到容易采样的建议分布 $q(z)$

2: 从 $q(z)$ 中抽取样本集合 $\{z^{(l)}\}$

3: 计算期望：

$$\mathbb{E}(f) = \frac{1}{L}\sum_{l=1}^{L}\frac{p(z^{(l)})}{q(z^{(l)})}f(z^{(l)})$$

**输出：** 期望 $\mathbb{E}(f)$



# 重要性采样

- **更一般场景**，概率分布没有归一化：$p(z)=\frac{\tilde{p}(z)}{\boxed{Z_p}}$，$\tilde{p}(z)$已知，$Z_p$ 未知

归一化常量

- 对于 $q(z)$，有 $q(z)=\frac{\tilde{q}(z)}{Z_q}$，进一步再看 $\mathbb{E}(f)$：

$$\mathbb{E}(f)=\int f(z)p(z)\mathrm{d}z=\frac{Z_q}{Z_p}\int f(z)\frac{\tilde{p}(z)}{\tilde{q}(z)}q(z)\mathrm{d}z\simeq\frac{Z_q}{Z_p}\frac{1}{L}\sum_{l=1}^{L}\frac{\tilde{p}(z^{(l)})}{\tilde{q}(z^{(l)})}f\left(z^{(l)}\right)$$

- 也可以使用从 $q(z)$ 中抽取的样本集合 $\{z^{(l)}\}$ 计算 $\frac{Z_q}{Z_p}$

$$\frac{Z_p}{Z_q}=\frac{1}{Z_q}\int\tilde{p}(z)\,\mathrm{d}z=\int\frac{\tilde{p}(z)}{\tilde{q}(z)}q(z)\,\mathrm{d}z\simeq\frac{1}{L}\sum_{l=1}^{L}\frac{\tilde{p}(z^{(l)})}{\tilde{q}(z^{(l)})}$$

# 重要性采样

$$\mathbb{E}(f)=\int f(\mathbf{z})p(\mathbf{z})\mathrm{d}\mathbf{z}=\frac{Z_q}{Z_p}\int f(\mathbf{z})\frac{\widetilde{p}(\mathbf{z})}{\widetilde{q}(\mathbf{z})}q(\mathbf{z})\mathrm{d}\mathbf{z}\simeq\frac{Z_q}{Z_p}\frac{1}{L}\sum_{l=1}^{L}\frac{\widetilde{p}(\mathbf{z}^{(l)})}{\widetilde{q}(\mathbf{z}^{(l)})}f(\mathbf{z}^{(l)})$$

$$\frac{Z_p}{Z_q}=\frac{1}{Z_q}\int\widetilde{p}(\mathbf{z})\,\mathrm{d}\mathbf{z}=\int\frac{\widetilde{p}(\mathbf{z})}{\widetilde{q}(\mathbf{z})}q(\mathbf{z})\,\mathrm{d}\mathbf{z}\simeq\frac{1}{L}\sum_{l=1}^{L}\frac{\widetilde{p}(\mathbf{z}^{(l)})}{\widetilde{q}(\mathbf{z}^{(l)})}$$

代入

- 联立上式，将 $\frac{Z_p}{Z_q}$ 代入 $\mathbb{E}(f)$，有：

$$\mathbb{E}(f)\simeq\sum_{l=1}^{L}w_l f(\mathbf{z}^{(l)}),\qquad w_l=\frac{\widetilde{p}(\mathbf{z}^{(l)})/q(\mathbf{z}^{(l)})}{\sum_m\widetilde{p}(\mathbf{z}^{(m)})/q(\mathbf{z}^{(m)})}$$

# 重要性采样

## 特点总结

- 可用于多种复杂度的分布估计，尤其是难以直接抽样的情况
- 相较于简单随机采样效率更高
- 建议分布与目标分布差异较大时，会导致估计的方差很大，估计变得不稳定

# 11.5 马尔可夫蒙特卡罗采样

- 马尔可夫蒙特卡罗采样
- Metropolis-Hastings
- 吉布斯采样

# 马尔可夫蒙特卡罗法 (Markov Chain Monte Carlo)

- 蒙特卡罗（Monte Carlo）坐落于欧洲地中海之滨、法国的东南方，有一个版图很小的国家摩纳哥公国。蒙特卡罗是世界著名的赌城，是摩纳哥的标志，世界三大赌城之一。富丽堂皇的蒙地卡罗赌场，建成于一八六三年，是一幢古色古香以及巍峨的宫殿式建筑物。1856年，摩纳哥亲王Charles三世为了解决财政危机，才在市区北边开设了第一家赌场。
- 蒙特卡罗方法于20世纪40年代美国在第二次世界大战中研制原子弹的“曼哈顿计划”计划的成员S.M.乌拉姆和J.冯·诺伊曼首先提出。数学家冯·诺伊曼用驰名世界的赌城—摩纳哥的Monte Carlo—来命名这种方法，为它蒙上了一层神秘色彩。在这之前，蒙特卡罗方法就已经存在。1777年，法国Buffon提出用投针实验的方法求圆周率π。这被认为是蒙特卡罗方法的起源。


# 马尔可夫蒙特卡罗法 (Markov Chain Monte Carlo)

- 马尔可夫蒙特卡洛方法（MCMC）
  - 是一种从复杂概率分布中采集样本的方法
  - 不同于拒绝采样法需选取适当的建议分布，MCMC利用马尔可夫链的收敛性质，通过迭代生成样本，并最终收敛到目标分布
  - 在高维的联合概率分布难以求解时，仅需知道目标分布对应的各维间的条件概率分布，也可利用MCMC采样进行求解（基本采样法和拒绝采样法均不适用）
- 关键问题：马尔可夫链的构造

# 马尔可夫链

- **回顾：马尔可夫链**

  - **一阶马尔可夫链（First Order Markov Chain）**

$$p\left(z^{m+1} \middle| z^{1}, \dots, z^{m}\right) = p\left(z^{m+1} \middle| z^{m}\right)$$

  - **高阶马尔可夫链（High Order Markov Chain）**

$$p\left(z^{m+1} \middle| z^{1}, \dots, z^{m}\right) = p\left(z^{m+1} \middle| z^{m}, \dots, z^{m-n}\right)$$

  - **转移概率（Transition Probabilities）**

$$T\left(z^{m}, z^{m+1}\right) \equiv p\left(z^{m+1} \middle| z^{m}\right)$$

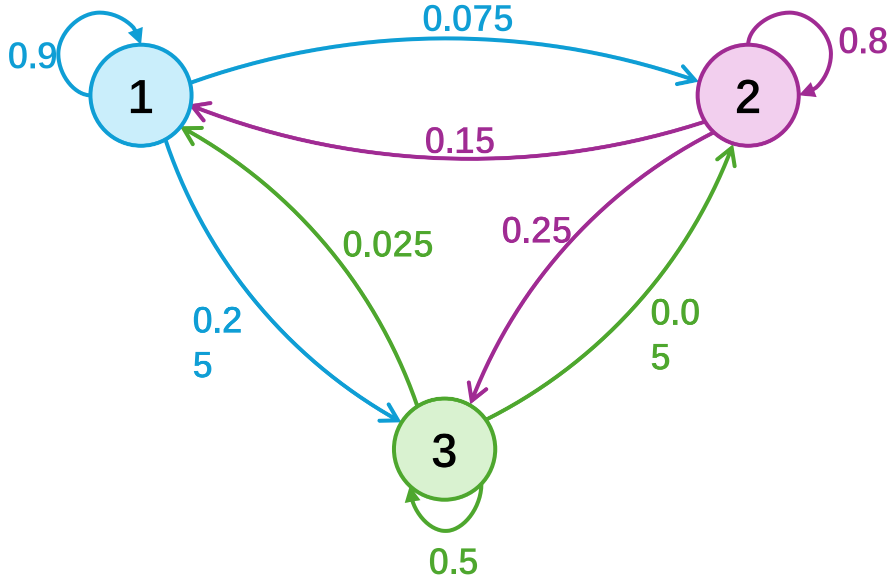
  - **转移概率矩阵**

$$T = \begin{bmatrix} T(1,1), & T(1,2), & \cdots & T(1,m) \\ T(2,1), & T(2,2), & & T(2,m) \\ \vdots & & \ddots & \vdots \\ T(m,1), & T(m,2), & \cdots & T(m,m) \end{bmatrix}$$

# 马尔可夫链

- **例子**

**股市模型中共有三种状态：牛市（Bull market），熊市（Bear market）和横盘（Stagnant market），用1，2，3 分别代表这三个状态。每一个状态都以一定的概率转化到下一个状态。比如，牛市以0.025的概率转化到横盘的状态。这个状态概率转化图可以以矩阵的形式表示。**




$$T = \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$

# 马尔可夫链

## 例子

将股市的初始状态记作：

$$p_0=[p_0(1),p_0(2),p_0(3)]$$

则第一个时刻的状态可表示为： $p_1=p_0T$

第二个时刻的状态将是： $p_2=p_1T=p_0T^2$

第$n$个时刻的状态将是： $p_n=p_{n-1}T=p_0T^n$

假设初始概率分布为： $p_0=[0.3,0.4,0.3]$

计算后代各阶层的分布为：

| 时刻 | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0.405 | 0.4175 | 0.1775 |
| 2 | 0.4715 | 0.4087 | 0.1198 |
| 3 | 0.5156 | 0.3293 | 0.0921 |
| … | … | … | … |
| 58 | 0.625 | 0.3125 | 0.0625 |
| 59 | 0.625 | 0.3125 | 0.0625 |
| … | … | … | … |
| 99 | 0.625 | 0.3125 | 0.0625 |
| 100 | 0.625 | 0.3125 | 0.0625 |

迭代到一定次数后状态保持不变！

# 马尔可夫链

- **收敛性质：当$m \to \infty$时，马尔可夫链各状态趋近于平稳，即**

$$p_m = p_{m-1}T \xrightarrow{m \to \infty} p = pT \qquad p = \boxed{[p(1), \dots, p(j), \dots]} \longrightarrow \boxed{\text{平稳分布}}$$

  - 且有

$$\lim_{m \to \infty} T^m = \begin{bmatrix} p(1), \dots, p(j), \dots \\ \dots \\ p(1), \dots, p(j), \dots \end{bmatrix}$$

- **细致平稳-充分条件**

**若马尔可夫链的状态转移矩阵$T$和概率分布$p(z)$对所有的$i,j$满足：**

$$p(i)T(i,j) = p(j)T(j,i)$$

**则称概率分布$p(z)$是状态转移矩阵$T$的平稳分布。**

# 马尔可夫蒙特卡罗法 (Markov Chain Monte Carlo)

- **MCMC采样法的思想**
  - 对于需采样的目标分布$p(z)$，构造转移矩阵为$T$的马尔可夫链，使它的平稳分布恰好为$p(z)$
- **MCMC采样法的步骤**
  - 从任意简单概率分布中采样初始状态
  - 利用目标平稳分布对应的状态转移矩阵$T$构造马尔可夫链
  - 迭代至一定次数后达到平稳分布对应的样本集即为采样结果

如何构造目标分布即马尔可夫链中的平稳分布所对应的状态转移矩阵$T$?

# Metropolis-Hastings 方法

- 一般情况下，目标平稳分布$p(z)$ 和某个马尔可夫链转移矩阵$T'$不满足细致平稳条件，即

$$p(i)T'(i,j) \neq p(j)T'(j,i)$$

- 思想：引入$\alpha(i,j)$使

$$p(i)T'(i,j)\alpha(i,j) = p(j)T'(j,i)\alpha(j,i)$$

➢ 欲使等式成立，则可令

$$\left.\begin{aligned}\alpha(i,j) &= \pi(j)T(j,i)\\ \alpha(j,i) &= \pi(i)T(i,j)\end{aligned}\right\}\text{接受率}$$

➢ 此时平稳分布对应转移矩阵为

$$T(i,j) = T'(i,j)\alpha(i,j)$$

➢ 从状态$i$以$T'(i,j)$转移到状态$j$时，以$\alpha(i,j)$的概率接受这个转移

# Metropolis-Hastings 方法

## ● M-H的算法流程

**输入:**

目标平稳分布$p(z)$, 任意选定的马尔可夫链状态转移矩阵$T$

状态转移次数阈值$n_1$, 需要的样本数$n_2$

**过程:**

1: 从任意简单概率分布采样得到初始状态值$z_0$

2: $for\ t = 0\ to\ n_1 + n_2$:

3: 从条件概率分布$T(z|z_t)$中产生$z_*$

4: 从(0，1)均匀分布中产生$u_0$

5: 若$u_0 < \boxed{\alpha(z_t, z_*) = p(z_*)T(z_t, z_*)}$ 则接受$z_{t+1} = z_*$，否则$z_{t+1} = z_t$

6: 样本集$\left(z_{n_1}, z_{n_1+1}, \dots, z_{n_1+n_2-1}\right)$即为所需平稳分布对应的样本集

若$\alpha(z_t, z_*)$过小，则采样效率较低！

# Metropolis-Hastings 方法

- 接受率$\alpha(i,j)$低，导致采样效率低
- 在细致平稳条件两边乘以因子$C$

$$p(i)T(i,j)\alpha(i,j)\cdot C=p(j)T(j,i)\alpha(j,i)\cdot C$$

细致平稳条件并没有打破！！！

- 同比例放大，使$\alpha(i,j),\alpha(j,i)$最大的为1，令

$$A(i,j)=min\left\{1,\frac{p(j)T(j,i)}{p(i)T(i,j)}\right\}$$

- 替换M-H算法流程中的$\alpha(i,j)$为$A(i,j)$

# Metropolis-Hastings 方法

- **M-H采样法的局限性**
  - ➢ 由于接受率计算式$\frac{p(j)T(j,i)}{p(i)T(i,j)}$在高维时计算量较大，导致算法效率较低；且接受率一般小于1，也会导致计算效率较低
  - ➢ 当特征维度较高时，往往很难求出各特征维度的联合分布
- **如何改进?**
  - ➢ 能否不拒绝状态转移?
  - ➢ 无法求出联合分布，能求出各维度间的条件概率分布，可以进行MCMC采样么?

城市空气质量研究涉及PM2.5浓度、温度、湿度、风速和气压等多维特征。获取这些特征的联合分布非常困难，需要综合分析多个地点、时间和气象条件。然而，计算各维度间的条件概率分布则相对容易，例如计算在特定风速下的PM2.5浓度的平均值，即可得到不同风速条件下的PM2.5浓度的条件概率分布。

# 吉布斯采样 (Gibbs Sampling)

- **重新寻找细致平稳条件**

  - 设目标分布为二元分布$p(x,y)$, 假设两个状态为$(x_1,y_1)$和$(x_1,y_2)$， 则有:

$$p(x_1,y_1)\cdot p(y_2|x_1)=p(x_1)\cdot p(y_1|x_1)\cdot p(y_2|x_1)$$
$$p(x_1,y_2)\cdot p(y_1|x_1)=p(x_1)\cdot p(y_2|x_1)\cdot p(y_1|x_1)$$

所以:

$$p(x_1,y_1)\cdot p(y_2|x_1)=p(x_1,y_2)\cdot p(y_1|x_1)$$

满足细致平稳条件 → 可进行MCMC采样!

- **Gibbs采样采用$p(y_2|x_1)$作为状态转移概率，可见其为M-H采样的特殊形式，且接受率$\alpha$为1.**

# 吉布斯采样 (Gibbs Sampling)

## 二维Gibbs采样算法流程

**输入：**

目标平稳分布对应的条件概率$p(x_1|x_2)$，

状态转移次数阈值$n_1$，需要的样本数$n_2$

**过程：**

1: 从任意简单概率分布采样得到初始状态值$x_1^{(0)}, x_2^{(0)}$

2: $for\ t = 0\ to\ n_1 + n_2:$

3: 从条件概率分布$p\left(x_2|x_1^{(t)}\right)$中产生$x_2^{(t+1)}$

4: 从条件概率分布$p\left(x_1|x_2^{(t+1)}\right)$中产生$x_1^{(t+1)}$

5: 样本集$\left((x_1^{(n_1)}, x_2^{(n_1)}), (x_1^{(n_1+1)}, x_2^{(n_1+1)}), \ldots, (x_1^{(n_1+n_2-1)}, x_2^{(n_1+n_2-1)})\right)$即为所需平稳分布对应的样本集

# 吉布斯采样 (Gibbs Sampling)

- **二维Gibbs采样例子**
  - 设目标分布为二维正态分布$Norm(\mu,\Sigma)$，其中$\mu=(\mu_1,\mu_2)=(5,-1)$,

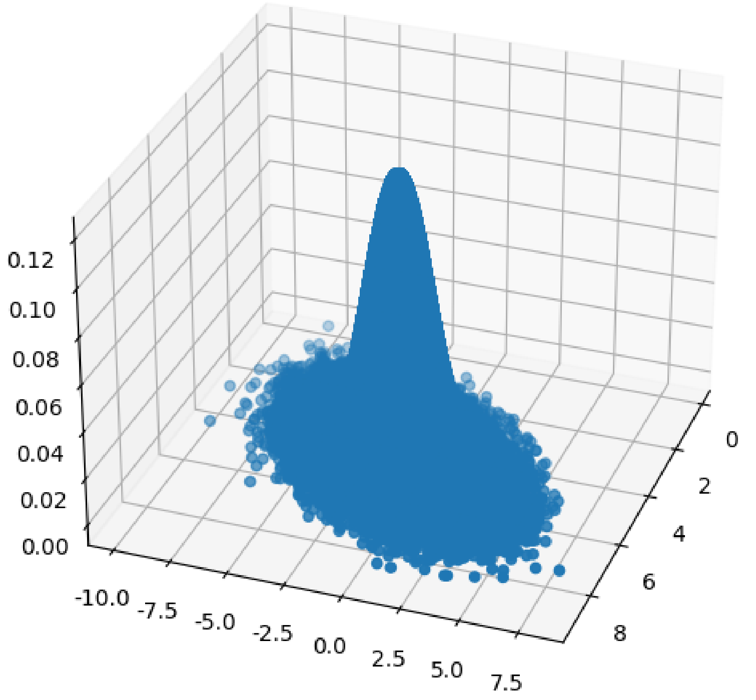
$$\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_1^2\end{pmatrix}=\begin{pmatrix}1 & 1\\ 1 & 4\end{pmatrix}$$

  - 二维正态分布的条件概率分布为:

$$p(x_1|x_2)=Norm(\mu_1+\frac{\rho\sigma_1}{\sigma_2(x_2-\mu_2)},(1-\rho$$

$$p(x_2|x_1)=Norm(\mu_2+\frac{\rho\sigma_2}{\sigma_1(x_1-\mu_1)},(1-\rho$$

  - 设共采样100000个点，转移次数为20次，则采样结果为



# 吉布斯采样 (Gibbs Sampling)

- **多维Gibbs采样算法：每次只改变一个维度的值，保持其他维度不变**

**输入：**

**目标平稳分布的所对应的条件概率分布**

**状态转移次数阈值$n_1$，需要的样本数$n_2$**

**过程：**

1: 从任意简单概率分布采样得到初始状态值$(x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$

2: $for\ t = 0\ to\ n_1 + n_2:$

3: 从条件概率分布$p\left(x_1 | x_2^{(t)}, \dots, x_n^{(t)}\right)$中产生$x_1^{(t+1)}$

4: 从条件概率分布$p\left(x_2 | x_1^{(t+1)}, x_3^{(t)}, \dots, x_n^{(t)}\right)$中产生$x_2^{(t+1)}$

…

从条件概率分布$p\left(x_n | x_1^{(t+1)}, x_2^{(t+1)}, \dots, x_{n-1}^{(t+1)}\right)$中产生$x_n^{(t+1)}$

5: 样本集$\left((x_1^{(n_1)}, x_2^{(n_1)}, \dots, x_n^{(n_1)}), \dots, (x_1^{(n_1+n_2-1)}, \dots, x_n^{(n_1+n_2-1)})\right)$即为所需平稳分布对应的样本集

# 马尔可夫蒙特卡洛采样法小结

- 马尔可夫蒙特卡罗采样法（MCMC）是基于马尔可夫链进行采样的一类方法
- Metropolis-Hasting（M-H）是MCMC的一种特定实现，通过引入接受率来构造合适的状态转移概率，以生成下一个状态
- 吉布斯方法是M-H方法中的一个特例，无需知道高维的联合分布，只需知道各维度间的条件概率分布，且接受率为1，保证了较高的采样效率

# 采样方法总结

<table>
<tr><th>采样目标</th><th colspan="2">采样方法</th><th>P</th><th>CDF<br>及逆</th><th>建议<br>分布</th><th>状态转移<br>矩阵T</th></tr>
<tr><td rowspan="4">采样服从目标分布的样本集</td><td colspan="2">基本采样法</td><td>√</td><td>√</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">拒绝采样法</td><td>√</td><td>-</td><td>√</td><td>-</td></tr>
<tr><td rowspan="2">MCMC<br>采样法</td><td>M-H</td><td>√</td><td>-</td><td>-</td><td>√</td></tr>
<tr><td>Gibbs</td><td>-</td><td>-</td><td>-</td><td>√</td></tr>
<tr><td>求目标函数的期望</td><td colspan="2">重要性采样法</td><td>√</td><td>-</td><td>√</td><td>-</td></tr>
</table>

注：P为目标分布对应的概率密度函数
CDF为目标分布对应的累积分布函数