

第一章 随机变量的抽样方法

均匀分布随机变量的抽样

- 对随机变量抽样是统计模拟的基本工具
 - ▶ 物理方法 (抛硬币、抽签): 真随机, 但数量少
 - ▶ 计算机抽样: 伪随机数, 但可以做到与目标分布真正的随机数无法通过统计检验区分

均匀分布随机变量的抽样

- 对随机变量抽样是统计模拟的基本工具
 - ▶ 物理方法 (抛硬币、抽签): 真随机, 但数量少
 - ▶ 计算机抽样: 伪随机数, 但可以做到与目标分布真正的随机数无法通过统计检验区分
- 对某分布抽样, 一般先产生 $U(0, 1)$ 的随机数, 然后再转换为服从目标分布的随机数

均匀分布随机变量的抽样

- 对随机变量抽样是统计模拟的基本工具
 - ▶ 物理方法 (抛硬币、抽签): 真随机, 但数量少
 - ▶ 计算机抽样: 伪随机数, 但可以做到与目标分布真正的随机数无法通过统计检验区分
- 对某分布抽样, 一般先产生 $U(0, 1)$ 的随机数, 然后再转换为服从目标分布的随机数
- 伪随机数序列的生成: $x_n = g(x_{n-1}, x_{n-2}, \dots, x_{n-p})$

均匀分布随机变量的抽样

- 对随机变量抽样是统计模拟的基本工具
 - ▶ 物理方法 (抛硬币、抽签): 真随机, 但数量少
 - ▶ 计算机抽样: 伪随机数, 但可以做到与目标分布真正的随机数无法通过统计检验区分
- 对某分布抽样, 一般先产生 $U(0, 1)$ 的随机数, 然后再转换为服从目标分布的随机数
- 伪随机数序列的生成: $x_n = g(x_{n-1}, x_{n-2}, \dots, x_{n-p})$
 - ▶ $U(0, 1)$ 随机数发生器: 集合 $\{0, 1, \dots, M\}$ 或 $\{1, 2, \dots, M\}$ 上的离散均匀分布, 再除以 M 或 $M+1$

均匀分布随机变量的抽样

- 对随机变量抽样是统计模拟的基本工具
 - ▶ 物理方法 (抛硬币、抽签): 真随机, 但数量少
 - ▶ 计算机抽样: 伪随机数, 但可以做到与目标分布真正的随机数无法通过统计检验区分
- 对某分布抽样, 一般先产生 $U(0, 1)$ 的随机数, 然后再转换为服从目标分布的随机数
- 伪随机数序列的生成: $x_n = g(x_{n-1}, x_{n-2}, \dots, x_{n-p})$
 - ▶ $U(0, 1)$ 随机数发生器: 集合 $\{0, 1, \dots, M\}$ 或 $\{1, 2, \dots, M\}$ 上的离散均匀分布, 再除以 M 或 $M+1$
 - ▶ 取值个数有限, 序列一定在某个时间后发生重复

均匀分布随机变量的抽样

- 对随机变量抽样是统计模拟的基本工具
 - ▶ 物理方法 (抛硬币、抽签): 真随机, 但数量少
 - ▶ 计算机抽样: 伪随机数, 但可以做到与目标分布真正的随机数无法通过统计检验区分
- 对某分布抽样, 一般先产生 $U(0, 1)$ 的随机数, 然后再转换为服从目标分布的随机数
- 伪随机数序列的生成: $x_n = g(x_{n-1}, x_{n-2}, \dots, x_{n-p})$
 - ▶ $U(0, 1)$ 随机数发生器: 集合 $\{0, 1, \dots, M\}$ 或 $\{1, 2, \dots, M\}$ 上的离散均匀分布, 再除以 M 或 $M+1$
 - ▶ 取值个数有限, 序列一定在某个时间后发生重复
 - ▶ 随机数发生器的周期: 序列发生重复的间隔 T

线性同余发生器

定义 (同余)

设 i, j 为整数, M 为正整数, 若 $j - i$ 为 M 的倍数, 则称 i 与 j 关于 M 同余, 记为 $i \equiv j \pmod{M}$ 。否则称 i 与 j 关于 M 不同余。

线性同余发生器

定义 (同余)

设 i, j 为整数, M 为正整数, 若 $j - i$ 为 M 的倍数, 则称 i 与 j 关于 M 同余, 记为 $i \equiv j \pmod{M}$ 。否则称 i 与 j 关于 M 不同余。

例如

$$7 \equiv 2 \pmod{5}, \quad -1 \equiv 4 \pmod{5}.$$

线性同余发生器

定义 (同余)

设 i, j 为整数, M 为正整数, 若 $j - i$ 为 M 的倍数, 则称 i 与 j 关于 M 同余, 记为 $i \equiv j \pmod{M}$ 。否则称 i 与 j 关于 M 不同余。

例如

$$7 \equiv 2 \pmod{5}, \quad -1 \equiv 4 \pmod{5}.$$

● 线性同余发生器递推公式

$$x_n = ax_{n-1} + c \pmod{M}, \quad n = 1, 2, \dots$$

- ▶ a, c : 事先设定的整数
- ▶ $0 \leq x_n < M$, 令 $R_n = x_n/M \in [0, 1)$

线性同余发生器

定义 (同余)

设 i, j 为整数, M 为正整数, 若 $j - i$ 为 M 的倍数, 则称 i 与 j 关于 M 同余, 记为 $i \equiv j \pmod{M}$ 。否则称 i 与 j 关于 M 不同余。

例如

$$7 \equiv 2 \pmod{5}, -1 \equiv 4 \pmod{5}.$$

- 线性同余发生器递推公式

$$x_n = ax_{n-1} + c \pmod{M}, n = 1, 2, \dots$$

- a, c : 事先设定的整数
 - $0 \leq x_n < M$, 令 $R_n = x_n/M \in [0, 1)$
- x_n 只有 M 个取值, 序列 x_0, x_1, \dots 一定会重复。使得 $x_n = x_m (n > m)$ 的 $n - m$ 的最小值 T 为随机数发生器在初值 x_0 下的周期 ($T \leq M$)

线性同余发生器

定义 (同余)

设 i, j 为整数, M 为正整数, 若 $j - i$ 为 M 的倍数, 则称 i 与 j 关于 M **同余**, 记为 $i \equiv j \pmod{M}$ 。否则称 i 与 j 关于 M 不同余。

例如

$$7 \equiv 2 \pmod{5}, \quad -1 \equiv 4 \pmod{5}.$$

● 线性同余发生器递推公式

$$x_n = ax_{n-1} + c \pmod{M}, n = 1, 2, \dots$$

- ▶ a, c : 事先设定的整数
- ▶ $0 \leq x_n < M$, 令 $R_n = x_n/M \in [0, 1)$

- x_n 只有 M 个取值, 序列 x_0, x_1, \dots 一定会重复。使得 $x_n = x_m (n > m)$ 的 $n - m$ 的最小值 T 为**随机数发生器在初值 x_0 下的周期** ($T \leq M$)

- ▶ 练习 1: 计算线性同余发生器

$$x_n = 3x_{n-1} + 3 \pmod{10}, n = 1, 2, \dots$$

取初值 $x_0 = 3$ 的周期。

线性同余发生器

- **满周期**: 从某个初值 x_0 出发达到最大周期 M

线性同余发生器

- **满周期**: 从某个初值 x_0 出发达到最大周期 M

定理

当下列三个条件都满足时，线性同余发生器可以达到满周期：

- ① c 与 M 互素
- ② 对 M 的任一个素因子 P , $a - 1$ 被 P 整除
- ③ 如果 4 是 M 的因子, 则 $a - 1$ 被 4 整除

线性同余发生器

- **满周期**: 从某个初值 x_0 出发达到最大周期 M

定理

当下列三个条件都满足时，线性同余发生器可以达到满周期：

- ① c 与 M 互素
- ② 对 M 的任一个素因子 P , $a - 1$ 被 P 整除
- ③ 如果 4 是 M 的因子, 则 $a - 1$ 被 4 整除

- 取 $a = 4m + 1$, $c = 2n + 1$ ($m, n \in \mathbb{N}$) 可达满周期

线性同余发生器

- **满周期**: 从某个初值 x_0 出发达到最大周期 M

定理

当下列三个条件都满足时，线性同余发生器可以达到满周期：

- ① c 与 M 互素
- ② 对 M 的任一个素因子 P , $a - 1$ 被 P 整除
- ③ 如果 4 是 M 的因子, 则 $a - 1$ 被 4 整除

- 取 $a = 4m + 1$, $c = 2n + 1$ ($m, n \in \mathbb{N}$) 可达满周期
- Kobayashi 提出了如下满周期 2^{31} 的线性同余发生器

$$x_n = 314159269x_{n-1} + 453806245 \pmod{2^{31}}.$$

线性同余发生器

好的均匀分布随机数发生器应满足:

- 周期足够长, 统计性质符合均匀分布

线性同余发生器

好的均匀分布随机数发生器应满足:

- 周期足够长, 统计性质符合均匀分布
- 有很好的随机性, 序列元素之间独立性好
 - ▶ 满周期的线性同余发生器, 序列中前后两项自相关系数的近似公式为

$$\rho(1) \approx \frac{1}{a} - \frac{6c}{aM}(1 - \frac{c}{M})$$

所以应将 a 选为较大的值 ($a < M$)。

FSR 发生器

线性同余发生器缺点:

- 产生的多维随机向量相关性大, 分布不均匀
- 周期不可能超过 2^L (L 为计算机整数位数)

FSR 发生器

线性同余发生器缺点:

- 产生的多维随机向量相关性大, 分布不均匀
- 周期不可能超过 2^L (L 为计算机整数位数)

Tausworthe (1965) 提出反馈位移寄存器法 (FSR):

$$\alpha_k = c_p \alpha_{k-p} + c_{p-1} \alpha_{k-p+1} + \cdots + c_1 \alpha_{k-1} \pmod{2}$$

其中 $c_i \in \{0, 1\}$.

FSR 发生器

线性同余发生器缺点:

- 产生的多维随机向量相关性大, 分布不均匀
- 周期不可能超过 2^L (L 为计算机整数位数)

Tausworthe (1965) 提出反馈位移寄存器法 (FSR):

$$\alpha_k = c_p \alpha_{k-p} + c_{p-1} \alpha_{k-p+1} + \cdots + c_1 \alpha_{k-1} \pmod{2}$$

其中 $c_i \in \{0, 1\}$.

- 递推可以利用逻辑运算快速实现

FSR 发生器

线性同余发生器缺点:

- 产生的多维随机向量相关性大, 分布不均匀
- 周期不可能超过 2^L (L 为计算机整数位数)

Tausworthe (1965) 提出反馈位移寄存器法 (FSR):

$$\alpha_k = c_p \alpha_{k-p} + c_{p-1} \alpha_{k-p+1} + \cdots + c_1 \alpha_{k-1} \pmod{2}$$

其中 $c_i \in \{0, 1\}$.

- 递推可以利用逻辑运算快速实现
- 对二进制序列 $\{\alpha_k : k = 1, 2, \dots\}$ 后, 依次截取长度 M 组合成整数 x_n , 再令 $R_n = x_n / 2^M \in [0, 1]$
- 作为多维均匀分布随机向量的发生器性质较好

FSR 发生器

线性同余发生器缺点:

- 产生的多维随机向量相关性大, 分布不均匀
- 周期不可能超过 2^L (L 为计算机整数位数)

Tausworthe (1965) 提出反馈位移寄存器法 (FSR):

$$\alpha_k = c_p \alpha_{k-p} + c_{p-1} \alpha_{k-p+1} + \cdots + c_1 \alpha_{k-1} \pmod{2}$$

其中 $c_i \in \{0, 1\}$.

- 递推可以利用逻辑运算快速实现
- 对二进制序列 $\{\alpha_k : k = 1, 2, \dots\}$ 后, 依次截取长度 M 组合成整数 x_n , 再令 $R_n = x_n / 2^M \in [0, 1]$
- 作为多维均匀分布随机向量的发生器性质较好
- 通过选择递推系数和初值可以达到很长的周期, 不受计算机字长限制

组合发生器

把若干个发生器组合利用，产生的随机数比单个发生器具有更长的周期和更好的随机性

- MacLaren and Marsaglia (1965) 提出组合同余法，组合两个同余发生器，其中一个用来“搅乱”次序

组合发生器

把若干个发生器组合利用，产生的随机数比单个发生器具有更长的周期和更好的随机性

- MacLaren and Marsaglia (1965) 提出组合同余法，组合两个同余发生器，其中一个用来“搅乱”次序
- Wichman and Hill (1982) 设计了如下的线性组合发生器。利用三个同余发生器：

$$U_n = 171 U_{n-1} \pmod{30269}$$

$$V_n = 172 V_{n-1} \pmod{30307}$$

$$W_n = 170 W_{n-1} \pmod{30323}$$

做线性组合并求余：

$$R_n = (U_n/30269 + V_n/30307 + W_n/30323) \pmod{1}$$

周期约为 7×10^{12} ，超过 $2^{31} \approx 2 \times 10^9$

随机数的检验

对 $U(0, 1)$ 随机数发生器产生的序列 $\{R_i : i = 1, 2, \dots, n\}$, 可以进行各种检验确认其均匀性:

- 把 $[0, 1]$ 等分成 K 段, 用 Pearson's χ^2 test 检验 $\{R_i : i = 1, 2, \dots, n\}$ 落在每一段的概率是否近似为 $1/K$

随机数的检验

对 $U(0, 1)$ 随机数发生器产生的序列 $\{R_i : i = 1, 2, \dots, n\}$, 可以进行各种检验确认其均匀性:

- 把 $[0, 1]$ 等分成 K 段, 用 Pearson's χ^2 test 检验 $\{R_i : i = 1, 2, \dots, n\}$ 落在每一段的概率是否近似为 $1/K$
- 用 Kolmogorov-Smirnov (K-S) test 检验 $\{R_i : i = 1, 2, \dots, n\}$ 是否近似服从 $U[0, 1]$ 分布

随机数的检验

对 $U(0, 1)$ 随机数发生器产生的序列 $\{R_i : i = 1, 2, \dots, n\}$, 可以进行各种检验确认其均匀性:

- 把 $[0, 1]$ 等分成 K 段, 用 Pearson's χ^2 test 检验 $\{R_i : i = 1, 2, \dots, n\}$ 落在每一段的概率是否近似为 $1/K$
- 用 Kolmogorov-Smirnov (K-S) test 检验 $\{R_i : i = 1, 2, \dots, n\}$ 是否近似服从 $U[0, 1]$ 分布
- 把 $\{R_i : i = 1, 2, \dots, n\}$ 每 d 个组合在一起成为 \mathbb{R}^d 向量, 把超立方体 $[0, 1]^d$ 每一维均匀分为 K 份, 得到 K^d 个子集, 用 Pearson's χ^2 test 检验这些 \mathbb{R}^d 向量落在每个子集的概率是否近似为 $1/K^d$
- ...

非均匀分布随机变量的抽样

- 均匀分布随机数的产生方法是很多非均匀分布抽样方法的基础
- 通用抽样方法:
 - ▶ CDF 逆变换
 - ▶ Acceptance-Rejection (A-R) 抽样

非均匀分布随机变量的抽样

- 均匀分布随机数的产生方法是很多非均匀分布抽样方法的基础
- 通用抽样方法:
 - ▶ CDF 逆变换
 - ▶ Acceptance-Rejection (A-R) 抽样
- 其他常用抽样方法:
 - ▶ 单调变换, 加和
 - ▶ Bootstrap
 - ▶ 混合分布抽样

CDF 逆变换

适用于对任何 CDF 逆函数已知的分布抽样

定义 (Cumulative distribution function (CDF))

一个随机变量 X 的 CDF $F(x)$ 定义为

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

CDF 逆变换

适用于对任何 CDF 逆函数已知的分布抽样

定义 (Cumulative distribution function (CDF))

一个随机变量 X 的 CDF $F(x)$ 定义为

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

- **CDF 逆变换的基本想法:** 假设随机变量 X 的 PDF $f(x) > 0, \forall x \in \mathbb{R}$, 其 CDF $F(x)$ 的逆函数 F^{-1} 存在, 抽取 $U \sim \mathbf{U}[0, 1]$, 令 $Y = F^{-1}(U)$, 则 $Y \sim F$.

CDF 逆变换

适用于对任何 CDF 逆函数已知的分布抽样

定义 (Cumulative distribution function (CDF))

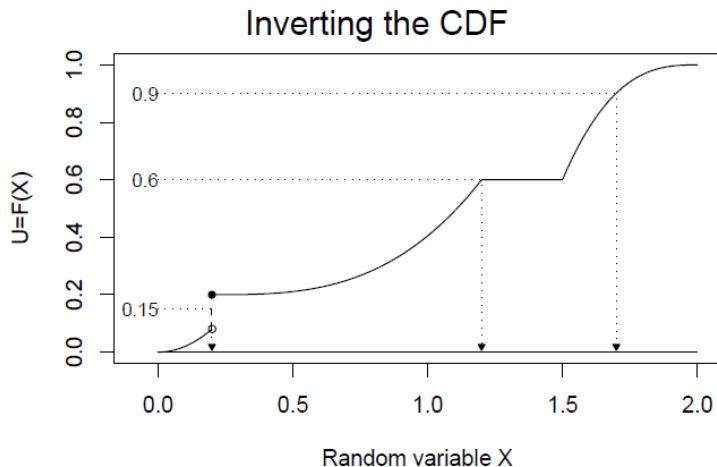
一个随机变量 X 的 CDF $F(x)$ 定义为

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

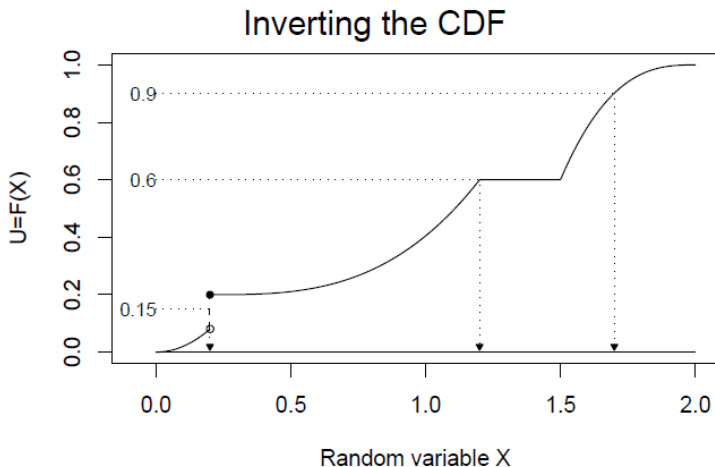
- **CDF 逆变换的基本想法:** 假设随机变量 X 的 PDF $f(x) > 0, \forall x \in \mathbb{R}$, 其 CDF $F(x)$ 的逆函数 F^{-1} 存在, 抽取 $U \sim \mathbf{U}[0, 1]$, 令 $Y = F^{-1}(U)$, 则 $Y \sim F$.
- 如果 CDF $F(x)$ 不连续 (离散分布) 或不可逆, 可以定义如下的**广义逆**:

$$F^{-1}(u) = \inf\{x \mid F(x) \geq u\}, \quad 0 < u < 1 \quad (1)$$

CDF 逆变换



CDF 逆变换



$$F^{-1}(0.15) = 0.2, F^{-1}(0.6) = 1.2, F^{-1}(0.9) = 1.7$$

CDF 逆变换

CDF 逆变换: 设 F 是一个 CDF, F^{-1} 是由(1)定义的逆函数, 令 $U \sim \mathbf{U}[0, 1]$, 则 $X = F^{-1}(U) \sim F$

CDF 逆变换

CDF 逆变换: 设 F 是一个 CDF, F^{-1} 是由(1)定义的逆函数, 令 $U \sim \mathbf{U}[0, 1]$, 则 $X = F^{-1}(U) \sim F$

- 如果 $U \sim \mathbf{U}[0, 1]$, $1 - U \sim \mathbf{U}[0, 1]$, 因此 $F^{-1}(1 - U) \sim F$

CDF 逆变换

CDF 逆变换: 设 F 是一个 CDF, F^{-1} 是由(1)定义的逆函数, 令 $U \sim \mathbf{U}[0, 1]$, 则 $X = F^{-1}(U) \sim F$

- 如果 $U \sim \mathbf{U}[0, 1]$, $1 - U \sim \mathbf{U}[0, 1]$, 因此 $F^{-1}(1 - U) \sim F$
- 如果 F 是一个连续的 CDF, $X \sim F$, G 是任意分布的 CDF, 则随机变量

$$Y = G^{-1}(F(X)) \sim G \quad (2)$$

称函数 $G^{-1}(F(\cdot))$ 为**QQ 变换**

CDF 逆变换

CDF 逆变换: 设 F 是一个 CDF, F^{-1} 是由(1)定义的逆函数, 令 $U \sim \mathbf{U}[0, 1]$, 则 $X = F^{-1}(U) \sim F$

- 如果 $U \sim \mathbf{U}[0, 1]$, $1 - U \sim \mathbf{U}[0, 1]$, 因此 $F^{-1}(1 - U) \sim F$
- 如果 F 是一个连续的 CDF, $X \sim F$, G 是任意分布的 CDF, 则随机变量

$$Y = G^{-1}(F(X)) \sim G \quad (2)$$

称函数 $G^{-1}(F(\cdot))$ 为**QQ 变换**

- ▶ QQ 变换将分布 F 的分位数转换为分布 G 下相应的分位数

CDF 逆变换举例

- **指数分布**. 标准指数分布 $\text{Exp}(1)$ 的 PDF:

$$f(x) = e^{-x}, x > 0$$

CDF 的逆函数:

$$F^{-1}(u) = -\log(1 - u)$$

先抽取 $U \sim \mathbf{U}(0, 1)$, 令 $X = -\log(1 - U)$ 或 $X = -\log(U)$ 可得服从 $\text{Exp}(1)$ 的样本

CDF 逆变换举例

- **指数分布**. 标准指数分布 $\text{Exp}(1)$ 的 PDF:

$$f(x) = e^{-x}, x > 0$$

CDF 的逆函数:

$$F^{-1}(u) = -\log(1 - u)$$

先抽取 $U \sim \mathbf{U}(0, 1)$, 令 $X = -\log(1 - U)$ 或 $X = -\log(U)$ 可得服从 $\text{Exp}(1)$ 的样本

- ▶ $Y \sim \text{Exp}(\lambda)$ 的 PDF:

$$f(y) = \lambda e^{-\lambda y}, y > 0$$

如果 $X \sim \text{Exp}(1)$, 则 $X/\lambda \sim \text{Exp}(\lambda)$, 因此 $Y = -\log(U)/\lambda \sim \text{Exp}(\lambda)$

CDF 逆变换举例

- **指数分布**. 标准指数分布 $\text{Exp}(1)$ 的 PDF:

$$f(x) = e^{-x}, x > 0$$

CDF 的逆函数:

$$F^{-1}(u) = -\log(1 - u)$$

先抽取 $U \sim \mathbf{U}(0, 1)$, 令 $X = -\log(1 - U)$ 或 $X = -\log(U)$ 可得服从 $\text{Exp}(1)$ 的样本

- ▶ $Y \sim \text{Exp}(\lambda)$ 的 PDF:

$$f(y) = \lambda e^{-\lambda y}, y > 0$$

如果 $X \sim \text{Exp}(1)$, 则 $X/\lambda \sim \text{Exp}(\lambda)$, 因此 $Y = -\log(U)/\lambda \sim \text{Exp}(\lambda)$

- ▶ 指数分布常用于描述一段时间的分布, 它的一个重要特性是**无记忆性**: 如果 $X \sim \text{Exp}(\lambda)$, 则

$$P(X \geq x + \Delta \mid X \geq x) = \frac{e^{-\lambda(x+\Delta)}}{e^{-\lambda x}} = e^{-\lambda \Delta}$$

CDF 逆变换举例

- **Bernoulli 分布.** 从 Bernoulli 分布 ($P(X=1) = p$, $P(X=0) = 1 - p$) 抽样可以令 $X = 1(1 - U \leq p)$ 或 $X = 1(U \leq p)$ 实现

CDF 逆变换举例

- **Bernoulli 分布**. 从 Bernoulli 分布 ($P(X=1) = p$, $P(X=0) = 1-p$) 抽样可以令 $X = 1(1-U \leq p)$ 或 $X = 1(U \leq p)$ 实现
- **Cauchy 分布**. Cauchy 分布是 t 分布的一个特例, 具有**厚尾**的特性, PDF:

$$f(x) = \frac{1}{\pi \cdot (1+x^2)}, \quad x \in \mathbb{R}$$

CDF:

$$F(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}$$

可利用如下 CDF 逆变换对 Cauchy 分布抽样:

$$X = \tan(\pi \cdot (U - 1/2)), \quad U \sim \mathbf{U}(0, 1)$$

几何解释: Cauchy 变量是一个在 $(-\pi/2, \pi/2)$ 上均匀分布的随机角的正切

CDF 逆变换举例

- **Poisson** 分布. $X \sim \text{Po}(\lambda)$ 的 PMF:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

Devroye (1986) 基于 CDF 逆变换提出以下算法从 Poisson 分布抽样 (考察算法输出 0 和 1 的概率)

Algorithm 1 Sample from Poisson distribution $\text{Po}(\lambda)$

Initialize $X = 0$, $p = q = e^{-\lambda}$, generate $U \sim \mathbf{U}(0, 1)$.

while $U > q$ **do**

$X = X + 1$

$p = p\lambda / X$

$q = q + p$

return X

CDF 逆变换举例

- **Poisson** 分布. $X \sim \text{Po}(\lambda)$ 的 PMF:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

Devroye (1986) 基于 CDF 逆变换提出以下算法从 Poisson 分布抽样 (考察算法输出 0 和 1 的概率)

Algorithm 1 Sample from Poisson distribution $\text{Po}(\lambda)$

Initialize $X = 0$, $p = q = e^{-\lambda}$, generate $U \sim \mathbf{U}(0, 1)$.

while $U > q$ **do**

$X = X + 1$

$p = p\lambda / X$

$q = q + p$

return X

- ▶ 算法从 $X = 0$ 开始寻找 U 落在哪两个累计概率所夹的区间

CDF 逆变换举例

- **Poisson** 分布. $X \sim \text{Po}(\lambda)$ 的 PMF:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

Devroye (1986) 基于 CDF 逆变换提出以下算法从 Poisson 分布抽样 (考察算法输出 0 和 1 的概率)

Algorithm 1 Sample from Poisson distribution $\text{Po}(\lambda)$

Initialize $X = 0$, $p = q = e^{-\lambda}$, generate $U \sim \mathbf{U}(0, 1)$.

while $U > q$ **do**

$X = X + 1$

$p = p\lambda/X$

$q = q + p$

return X

- ▶ 算法从 $X = 0$ 开始寻找 U 落在哪两个累计概率所夹的区间
- ▶ 条件 $U > q$ 会被检验 $X + 1$ 次, λ 较大时使用该方法抽样会很慢

CDF 逆变换举例

- **Weibull 分布**. Weibull 分布是对指数分布的推广, PDF:

$$f(x) = \frac{k}{\sigma} \left(\frac{x}{\sigma} \right)^{k-1} e^{-(x/\sigma)^k}, \quad x > 0$$

参数 $\sigma > 0, k > 0$, CDF:

$$F(x) = 1 - \exp \left(- (x/\sigma)^k \right), \quad x > 0$$

因此对 $U \sim \mathbf{U}(0, 1)$ 做变换 $X = \sigma (-\log(1 - U))^{1/k}$ 可得到 Weibull 分布的样本

CDF 逆变换举例

- **Weibull 分布**. Weibull 分布是对指数分布的推广, PDF:

$$f(x) = \frac{k}{\sigma} \left(\frac{x}{\sigma}\right)^{k-1} e^{-(x/\sigma)^k}, \quad x > 0$$

参数 $\sigma > 0, k > 0$, CDF:

$$F(x) = 1 - \exp\left(- (x/\sigma)^k\right), \quad x > 0$$

因此对 $U \sim \mathbf{U}(0, 1)$ 做变换 $X = \sigma (-\log(1 - U))^{1/k}$ 可得到 Weibull 分布的样本

定义 (Hazard function)

对一个随机变量 $X > 0$, 它的 hazard function $h(x)$ 定义为

$$h(x) = \lim_{t \rightarrow 0^+} \frac{1}{t} P(X \leq x + t \mid X \geq x), \quad x > 0.$$

Hazard function 也被称为瞬时失败概率

CDF 逆变换举例

- Weibull 分布的 hazard function:

$$h(x) = k \cdot \frac{x^{k-1}}{\sigma^k}$$

- ▶ $k < 1$ 时, $h(x)$ 随时间 x 递减, 比如婴儿的死亡概率
- ▶ $k = 1$ 时, Weibull 退化为指数分布, $h(x)$ 是常数
- ▶ $k > 1$ 时, $h(x)$ 随时间 x 递增, 比如非耐用品的生命周期

CDF 逆变换举例

- Weibull 分布的 hazard function:

$$h(x) = k \cdot \frac{x^{k-1}}{\sigma^k}$$

- ▶ $k < 1$ 时, $h(x)$ 随时间 x 递减, 比如婴儿的死亡概率
- ▶ $k = 1$ 时, Weibull 退化为指数分布, $h(x)$ 是常数
- ▶ $k > 1$ 时, $h(x)$ 随时间 x 递增, 比如非耐用品的生命周期

- **双指数分布.** 标准双指数分布的 PDF:

$$f(x) = \frac{1}{2} \exp(-|x|), \quad x \in \mathbb{R}$$

练习: 求标准双指数分布的 CDF 及其逆函数

CDF 逆变换举例

- Weibull 分布的 hazard function:

$$h(x) = k \cdot \frac{x^{k-1}}{\sigma^k}$$

- ▶ $k < 1$ 时, $h(x)$ 随时间 x 递减, 比如婴儿的死亡概率
- ▶ $k = 1$ 时, Weibull 退化为指数分布, $h(x)$ 是常数
- ▶ $k > 1$ 时, $h(x)$ 随时间 x 递增, 比如非耐用品的生命周期

- **双指数分布.** 标准双指数分布的 PDF:

$$f(x) = \frac{1}{2} \exp(-|x|), \quad x \in \mathbb{R}$$

练习: 求标准双指数分布的 CDF 及其逆函数

$$F(x) = \begin{cases} \frac{1}{2}e^x, & x < 0 \\ 1 - \frac{1}{2}e^{-x}, & x \geq 0 \end{cases} \quad F^{-1}(u) = \begin{cases} \log(2u), & 0 < u \leq 1/2 \\ -\log(2(1-u)), & 1/2 < u < 1 \end{cases}$$

CDF 逆变换举例

- **Triangular and power densities.** Triangular 和 power 分布都是 Beta 分布的特例，后面会介绍更高效的抽样方法

- ▶ Triangular density:

$$f(x) = 2x, \quad 0 < x < 1$$

$$F^{-1}(U) = \sqrt{U}, \quad U \sim \mathbf{U}(0, 1)$$

- ▶ Power density ($\alpha > 0$):

$$f(x) = \alpha x^{\alpha-1}, \quad 0 < x < 1.$$

$$F^{-1}(U) = U^{1/\alpha}, \quad U \sim \mathbf{U}(0, 1)$$

CDF 逆变换举例

- **Triangular and power densities.** Triangular 和 power 分布都是 Beta 分布的特例, 后面会介绍更高效的抽样方法

- ▶ Triangular density:

$$f(x) = 2x, \quad 0 < x < 1$$

$$F^{-1}(U) = \sqrt{U}, \quad U \sim \mathbf{U}(0, 1)$$

- ▶ Power density ($\alpha > 0$):

$$f(x) = \alpha x^{\alpha-1}, \quad 0 < x < 1.$$

$$F^{-1}(U) = U^{1/\alpha}, \quad U \sim \mathbf{U}(0, 1)$$

- **截断分布抽样.** 随机变量 $Y \sim F$ (F 是一个连续分布的 CDF), X 服从 F 在 (a, b) 上的截断分布 G , 则 X 的 CDF 为

$$G(x) = \frac{F(x) - F(a)}{F(b) - F(a)}$$

CDF 逆变换举例

- **Triangular and power densities.** Triangular 和 power 分布都是 Beta 分布的特例，后面会介绍更高效的抽样方法

- ▶ Triangular density:

$$f(x) = 2x, \quad 0 < x < 1$$

$$F^{-1}(U) = \sqrt{U}, \quad U \sim \mathbf{U}(0, 1)$$

- ▶ Power density ($\alpha > 0$):

$$f(x) = \alpha x^{\alpha-1}, \quad 0 < x < 1.$$

$$F^{-1}(U) = U^{1/\alpha}, \quad U \sim \mathbf{U}(0, 1)$$

- **截断分布抽样.** 随机变量 $Y \sim F$ (F 是一个连续分布的 CDF), X 服从 F 在 (a, b) 上的截断分布 G , 则 X 的 CDF 为

$$G(x) = \frac{F(x) - F(a)}{F(b) - F(a)}$$

因此可通过以下变换从截断分布 G 抽样:

$$X = F^{-1}(F(a) + (F(b) - F(a))U), \quad U \sim \mathbf{U}(0, 1)$$

单调变换

- CDF 逆变换通用但不总是很高效，有时利用分布之间的特殊关系可以构造更简洁的抽样变换
- 假设随机变量 X 的抽样方法已知，如果随机变量 $Y = \tau(X)$ ， $\tau(\cdot)$ 是一个单调递增函数，则 Y 的样本可通过对 X 的样本做单调变换 τ 得到

单调变换

- CDF 逆变换通用但不总是很高效, 有时利用分布之间的特殊关系可以构造更简洁的抽样变换
- 假设随机变量 X 的抽样方法已知, 如果随机变量 $Y = \tau(X)$, $\tau(\cdot)$ 是一个单调递增函数, 则 Y 的样本可通过对 X 的样本做单调变换 τ 得到
 - ▶ 对数正态分布: 如果 $X \sim N(\mu, \sigma^2)$, $Y = \exp(X)$ 服从对数正态分布
 - ▶ 当 $X \sim N(0, 1)$, $Y = \mu + \sigma X \sim N(\mu, \sigma^2)$

单调变换

- CDF 逆变换通用但不总是很高效, 有时利用分布之间的特殊关系可以构造更简洁的抽样变换
- 假设随机变量 X 的抽样方法已知, 如果随机变量 $Y = \tau(X)$, $\tau(\cdot)$ 是一个单调递增函数, 则 Y 的样本可通过对 X 的样本做单调变换 τ 得到
 - ▶ 对数正态分布: 如果 $X \sim N(\mu, \sigma^2)$, $Y = \exp(X)$ 服从对数正态分布
 - ▶ 当 $X \sim N(0, 1)$, $Y = \mu + \sigma X \sim N(\mu, \sigma^2)$
 - ▶ 任何一个单调变换 $Y = \tau(X)$ 都对应一个 QQ 变换 $F_Y^{-1}(F_X(\cdot))$

Box-Muller 变换

- Box-Muller 变换用两个独立的 $U(0, 1)$ 变量产生两个独立的 $N(0, 1)$ 变量:

$$\begin{aligned} Z_1 &= \sqrt{-2 \log U_1} \cos(2\pi U_2) \\ Z_2 &= \sqrt{-2 \log U_1} \sin(2\pi U_2) \end{aligned} \tag{3}$$

其中 $U_1, U_2 \sim \mathbf{U}(0, 1)$ 且独立

Box-Muller 变换

- Box-Muller 变换用两个独立的 $U(0, 1)$ 变量产生两个独立的 $N(0, 1)$ 变量:

$$\begin{aligned}Z_1 &= \sqrt{-2 \log U_1} \cos(2\pi U_2) \\Z_2 &= \sqrt{-2 \log U_1} \sin(2\pi U_2)\end{aligned}\tag{3}$$

其中 $U_1, U_2 \sim \mathbf{U}(0, 1)$ 且独立

- 原理: 将二元随机向量 $(Z_1, Z_2) \sim N(\mathbf{0}, I_2)$ 用极坐标表示

$$\begin{aligned}Z_1 &= R \cos(\theta) \\Z_2 &= R \sin(\theta)\end{aligned}\tag{4}$$

可以证明极角 $\theta \sim \mathbf{U}[0, 2\pi)$, 且独立于半径 R ; 半径 $R^2 = Z_1^2 + Z_2^2 \sim \chi_{(2)}^2$

Box-Muller 变换

- Box-Muller 变换用两个独立的 $U(0, 1)$ 变量产生两个独立的 $N(0, 1)$ 变量:

$$\begin{aligned}Z_1 &= \sqrt{-2 \log U_1} \cos(2\pi U_2) \\Z_2 &= \sqrt{-2 \log U_1} \sin(2\pi U_2)\end{aligned}\tag{3}$$

其中 $U_1, U_2 \sim \mathbf{U}(0, 1)$ 且独立

- 原理: 将二元随机向量 $(Z_1, Z_2) \sim N(\mathbf{0}, I_2)$ 用极坐标表示

$$\begin{aligned}Z_1 &= R \cos(\theta) \\Z_2 &= R \sin(\theta)\end{aligned}\tag{4}$$

可以证明极角 $\theta \sim \mathbf{U}[0, 2\pi)$, 且独立于半径 R ; 半径 $R^2 = Z_1^2 + Z_2^2 \sim \chi_{(2)}^2$

- Box-Muller 不是最快从 $N(0, 1)$ 抽样的方法

最大、最小统计量

- 假设 X_1, \dots, X_r 独立同分布且 CDF 是 F , $Y = \max(X_1, \dots, X_r)$ 的 CDF:

$$P(Y \leq y) = (F(y))^r$$

最大、最小统计量

- 假设 X_1, \dots, X_r 独立同分布且 CDF 是 F , $Y = \max(X_1, \dots, X_r)$ 的 CDF:

$$P(Y \leq y) = (F(y))^r$$

因此对 CDF 为 $G = F^r$ 的分布抽样, 可以先从分布 F 独立抽取 r 个样本, 然后只保留最大样本

- ▶ 三角分布的 CDF $G(y) = y^2$, $0 < y < 1$. 用 $\max(U_1, U_2)$ 抽样比逆变换法 $\sqrt{U_1}$ 快

最大、最小统计量

- 假设 X_1, \dots, X_r 独立同分布且 CDF 是 F , $Y = \max(X_1, \dots, X_r)$ 的 CDF:

$$P(Y \leq y) = (F(y))^r$$

因此对 CDF 为 $G = F^r$ 的分布抽样, 可以先从分布 F 独立抽取 r 个样本, 然后只保留最大样本

- ▶ 三角分布的 CDF $G(y) = y^2$, $0 < y < 1$. 用 $\max(U_1, U_2)$ 抽样比逆变换法 $\sqrt{U_1}$ 快

- $Y = \min(X_1, \dots, X_r)$ 的 CDF:

$$G(y) = 1 - (1 - F(y))^r$$

最大、最小统计量

- 假设 X_1, \dots, X_r 独立同分布且 CDF 是 F , $Y = \max(X_1, \dots, X_r)$ 的 CDF:

$$P(Y \leq y) = (F(y))^r$$

因此对 CDF 为 $G = F^r$ 的分布抽样, 可以先从分布 F 独立抽取 r 个样本, 然后只保留最大样本

- ▶ 三角分布的 CDF $G(y) = y^2$, $0 < y < 1$. 用 $\max(U_1, U_2)$ 抽样比逆变换法 $\sqrt{U_1}$ 快

- $Y = \min(X_1, \dots, X_r)$ 的 CDF:

$$G(y) = 1 - (1 - F(y))^r$$

- ▶ 考察 $Y = \min(U_1, U_2)$ 的分布, 其中 $U_1, U_2 \sim \mathbf{U}(0, 1)$. Y 的 CDF:

$$G(y) = 1 - (1 - y)^2, \quad 0 < y < 1$$

如果用 CDF 逆变换对该分布抽样, 需计算 $Y = 1 - \sqrt{1 - U_1}$

顺序统计量

最大、最小统计量都是**顺序统计量**的特例

定义 (顺序统计量)

对于 n 个独立同分布的随机变量, X_1, \dots, X_n . 它们的顺序统计量是将这 n 个变量的取值按从小到大的顺序排列, 记为

$$X_{(1)} \leq X_{(2)} \leq \dots, \leq X_{(n)}.$$

顺序统计量

最大、最小统计量都是**顺序统计量**的特例

定义 (顺序统计量)

对于 n 个独立同分布的随机变量, X_1, \dots, X_n . 它们的顺序统计量是将这 n 个变量的取值按从小到大的顺序排列, 记为

$$X_{(1)} \leq X_{(2)} \leq \dots, \leq X_{(n)}.$$

- 对于 n 个独立的 $U(0, 1)$ 随机变量 U_1, \dots, U_n , 证明 $U_{(r)} \sim \text{Beta}(r, n - r + 1)$

顺序统计量

最大、最小统计量都是**顺序统计量**的特例

定义 (顺序统计量)

对于 n 个独立同分布的随机变量, X_1, \dots, X_n . 它们的顺序统计量是将这 n 个变量的取值按从小到大的顺序排列, 记为

$$X_{(1)} \leq X_{(2)} \leq \dots, \leq X_{(n)}.$$

- 对于 n 个独立的 $U(0, 1)$ 随机变量 U_1, \dots, U_n , 证明 $U_{(r)} \sim \text{Beta}(r, n - r + 1)$
- n 个独立同分布的 $Y_i \sim F$, n 很大时, 一个快速得到 $Y_{(r)}$ 样本的方法: 先抽取 $X \sim \text{Beta}(r, n - r + 1)$, 则 $Y_{(r)} = F^{-1}(X)$

顺序统计量

练习：一个系统由 n 个独立的元件组成，每个元件或者工作或者不工作。至少需要 k 个元件工作才能保证系统正常运行。假设在 0 时刻，所有元件都正常工作，用 Y_i 表示元件 i 不工作的时刻， $Y_i > 0$ 且 Y_i 独立服从 Weibull($\sigma = 1, k = 2$) 分布（元件会老化）， $i = 1, \dots, n$ 。

Weibull($\sigma = 1, k = 2$) 的 CDF 为

$$F(x) = 1 - \exp(-x^2), \quad x > 0.$$

用 S 表示系统停止运行的时刻，如何得到 S 的样本？

加和

如果随机变量 Y 可写为 n 个独立同分布的随机变量之和, $Y = X_1 + \cdots + X_n$, 且 X_i 的分布较简单, 对 Y 抽样可以先从 X_i 的分布中独立抽取 n 个样本, 再加和

- **二项分布.** $Y \sim \text{Bin}(n, p)$ 等于 n 个 Bernoulli 变量的和
- **χ^2 分布.** 如果 $X_i \stackrel{iid}{\sim} \chi^2_{(\alpha)}$, $i = 1, 2, \dots, n$. 则

$$Y = \sum_{i=1}^n X_i \sim \chi^2_{(n\alpha)}.$$

加和

如果随机变量 Y 可写为 n 个独立同分布的随机变量之和, $Y = X_1 + \cdots + X_n$, 且 X_i 的分布较简单, 对 Y 抽样可以先从 X_i 的分布中独立抽取 n 个样本, 再加和

- **二项分布.** $Y \sim \text{Bin}(n, p)$ 等于 n 个 Bernoulli 变量的和
- **χ^2 分布.** 如果 $X_i \stackrel{iid}{\sim} \chi^2_{(\alpha)}$, $i = 1, 2, \dots, n$. 则

$$Y = \sum_{i=1}^n X_i \sim \chi^2_{(n\alpha)}.$$

- **Noncentral χ^2 分布.** 分布有两个参数, 自由度 n 和参数 $\lambda \geq 0$, 记为 $\chi'^2_{(n)}(\lambda)$. 可如下生成:

$$Y = \sum_{i=1}^n X_i^2, \quad X_i \stackrel{ind}{\sim} N(a_i, 1), \quad \lambda = \sqrt{\sum_{i=1}^n a_i^2}$$

Bootstrap

- 统计推断中, 如果无法给出统计量 $\hat{T} = T(\mathbf{X}_1, \dots, \mathbf{X}_n)$ 的具体分布 $(\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{iid}{\sim} F)$, 可以用 Bootstrap 方法得到 \hat{T} 的近似分布

Bootstrap

- 统计推断中, 如果无法给出统计量 $\hat{T} = T(\mathbf{X}_1, \dots, \mathbf{X}_n)$ 的具体分布 $(\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{iid}{\sim} F)$, 可以用 Bootstrap 方法得到 \hat{T} 的近似分布
- Bootstrap 需要对观察值的**经验分布**进行重抽样

定义 (经验分布)

观察值 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 的经验分布的 CDF 定义为

$$\hat{F}_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\mathbf{x}_i \leq \mathbf{x}).$$

- 经验分布是一个在所有观察值处取值的均匀离散分布
- 对经验分布进行重抽样时, 每个观察值 \mathbf{x}_i 被抽到的概率都是 $1/n$

Bootstrap

- Bootstrap 先对数据的经验分布重抽样生成 B 组数据, 每组数据 b 仍然包含 n 个样本, 即

$$\mathbf{X}_i^{*b} \stackrel{iid}{\sim} \hat{F}_n, \quad i = 1, \dots, n; \quad b = 1, \dots, B$$

- 每组重抽样数据 b 可以计算 \hat{T} 的一个值

$$\hat{T}^{*b} = T(\mathbf{X}_1^{*b}, \dots, \mathbf{X}_n^{*b}), \quad b = 1, \dots, B$$

Bootstrap

- Bootstrap 先对数据的经验分布重抽样生成 B 组数据, 每组数据 b 仍然包含 n 个样本, 即

$$\mathbf{X}_i^{*b} \stackrel{iid}{\sim} \hat{F}_n, \quad i = 1, \dots, n; \quad b = 1, \dots, B$$

- 每组重抽样数据 b 可以计算 \hat{T} 的一个值

$$\hat{T}^{*b} = T(\mathbf{X}_1^{*b}, \dots, \mathbf{X}_n^{*b}), \quad b = 1, \dots, B$$

- $\{\hat{T}^{*b} : b = 1, \dots, B\}$ 一般可以很好地近似 \hat{T} 服从的分布
 - ▶ $\text{Var}(\hat{T})$ 可以如下估计:

$$\text{Var}(\hat{T}) \approx \frac{1}{B-1} \sum_{b=1}^B (\hat{T}^{*b} - \bar{T})^2$$

- ▶ \hat{T} 的 95% 置信区间近似为 $\{\hat{T}^{*b} : b = 1, \dots, B\}$ 的 2.5% 分位数与 97.5% 分位数所夹的区间

Acceptance-Rejection

- 当 CDF 逆变换不可行, 可以尝试 **A-R 抽样**: 先从一个容易抽样的分布 G 中抽样, 然后按照某种规则去掉一些样本, 使得最后保留下来的样本服从目标分布 F
- 假设 f 和 g 是两个连续分布的 PDF, 使用 A-R 抽样需满足:
 - ① 可以从分布 g 中抽样
 - ② 函数 f/g 可计算
 - ③ 存在常数 $c > 0$ 使得 $f(x) \leq cg(x), \forall x \in \mathbb{R}$

Acceptance-Rejection

- 当 CDF 逆变换不可行, 可以尝试 **A-R 抽样**: 先从一个容易抽样的分布 G 中抽样, 然后按照某种规则去掉一些样本, 使得最后保留下来的样本服从目标分布 F
- 假设 f 和 g 是两个连续分布的 PDF, 使用 A-R 抽样需满足:
 - ① 可以从分布 g 中抽样
 - ② 函数 f/g 可计算
 - ③ 存在常数 $c > 0$ 使得 $f(x) \leq cg(x), \forall x \in \mathbb{R}$

Algorithm 2 Acceptance-rejection sampling

```
given  $c$  with  $f(x) \leq c \cdot g(x), \forall x \in \mathbb{R}$   
repeat  
     $Y \sim g$   
     $U \sim \mathbf{U}(0, 1)$   
until  $U \leq f(Y)/(cg(Y))$   
 $X = Y$   
return  $X$ 
```

Acceptance-Rejection

- 条件 $f(x) \leq cg(x)$ 保证了每个候选样本 y 被接受的概率 $A(y) \leq 1$, 该条件也要求 g 的 support 必须覆盖 f 的 support

Acceptance-Rejection

- 条件 $f(x) \leq cg(x)$ 保证了每个候选样本 y 被接受的概率 $A(y) \leq 1$, 该条件也要求 g 的 support 必须覆盖 f 的 support
- A-R 抽样的基本想法: 如果候选样本 y 来自分布 g 且以概率 $A(y)$ 被接受, 它出现的概率密度 $\propto g(y)A(y)$, 因此令 $A(y) \propto f(y)/g(y)$ 则有 $g(y)A(y) \propto f(y)$

Acceptance-Rejection

- 条件 $f(x) \leq cg(x)$ 保证了每个候选样本 y 被接受的概率 $A(y) \leq 1$, 该条件也要求 g 的 support 必须覆盖 f 的 support
- A-R 抽样的基本想法: 如果候选样本 y 来自分布 g 且以概率 $A(y)$ 被接受, 它出现的概率密度 $\propto g(y)A(y)$, 因此令 $A(y) \propto f(y)/g(y)$ 则有 $g(y)A(y) \propto f(y)$

定理

如果两个 PDF $f(x)$ 和 $g(x)$ 满足

$$f(x) \leq cg(x), \forall x \in \mathbb{R}$$

则 Algorithm 2 产生的 $X \sim f$.

Acceptance-Rejection

- A-R 方法中 g 分布产生的样本期望被接受的概率是 $1/c$ ($c \geq 1$)

Acceptance-Rejection

- A-R 方法中 g 分布产生的样本期望被接受的概率是 $1/c$ ($c \geq 1$)
- 为了减少样本损失, 提高算法效率, c 越小越好, 一般可将 c 取为

$$c = \sup_x \frac{f(x)}{g(x)}$$

Acceptance-Rejection

- A-R 方法中 g 分布产生的样本期望被接受的概率是 $1/c$ ($c \geq 1$)
- 为了减少样本损失, 提高算法效率, c 越小越好, 一般可将 c 取为

$$c = \sup_x \frac{f(x)}{g(x)}$$

- ▶ **从 Cauchy 分布生成正态分布的样本.** $N(0, 1)$ 的 PDF 为 $f(x) = 1/\sqrt{2\pi} \cdot \exp(-x^2/2)$. Cauchy 分布的 PDF 为 $g(x) = \pi^{-1}(1+x^2)^{-1}$. 计算

$$c = \sup_x \sqrt{\frac{\pi}{2}}(1+x^2) \exp(-x^2/2)$$

A-R 的几何解释

从几何角度理解 A-R 方法需要用到以下两个定理

定理

如果 $Y \sim g$ (g 是 PDF) 且 $Z|Y \sim \mathbf{U}(0, cg(Y))$ (常数 $c > 0$), 则 (Y, Z) 的联合分布是区域

$$S_c(g) = \{(y, z) \mid 0 \leq z \leq cg(y), y \in \mathbb{R}\} \subset \mathbb{R}^2$$

上的均匀分布。

A-R 的几何解释

从几何角度理解 A-R 方法需要用到以下两个定理

定理

如果 $Y \sim g$ (g 是 PDF) 且 $Z|Y \sim \mathbf{U}(0, cg(Y))$ (常数 $c > 0$), 则 (Y, Z) 的联合分布是区域

$$S_c(g) = \{(y, z) \mid 0 \leq z \leq cg(y), y \in \mathbb{R}\} \subset \mathbb{R}^2$$

上的均匀分布。

定理

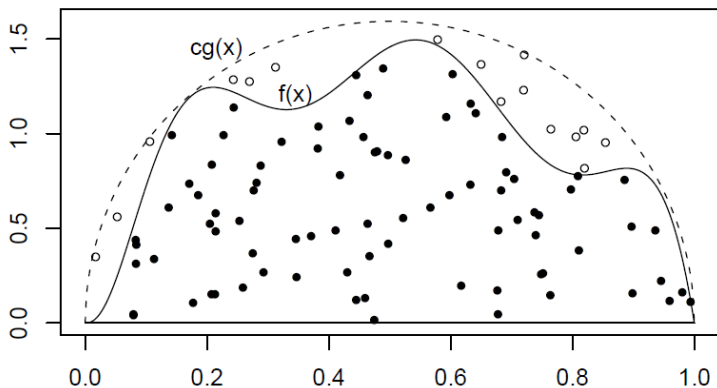
如果 $(X, Z) \sim \mathbf{U}(S_M(f))$, 其中

$$S_M(f) = \{(x, z) \mid 0 \leq z \leq Mf(x), x \in \mathbb{R}\}$$

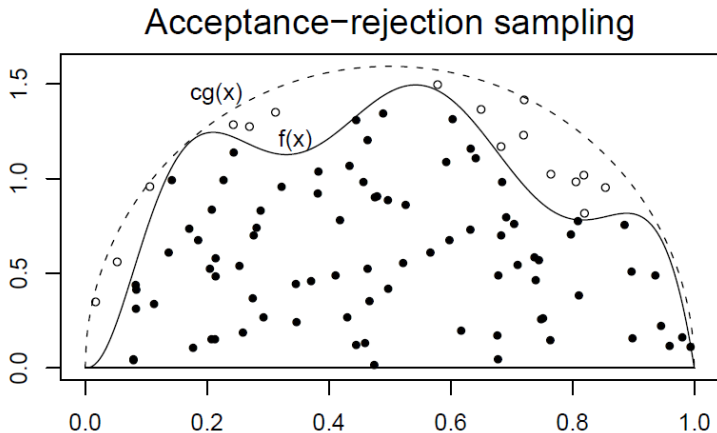
$M > 0$, f 是 \mathbb{R} 上的一个 PDF, 则 $X \sim f$.

A-R 的几何解释

Acceptance-rejection sampling



A-R 的几何解释



- A-R 的几何解释表明 Algorithm 2 可以使用未归一化的 PDF $\tilde{f} \propto f$ 和 $\tilde{g} \propto g$, 此时相当于在正比于 $g(x)$ 的曲线下均匀撒点, 然后只保留正比于 $f(x)$ 曲线下的点, 而这些被接受的点的横坐标仍服从分布 f

A-R 方法的应用

- 如何获得 $N(0, 1)$ 在区间 $[5, +\infty)$ 上的样本?

A-R 抽样的思路: 将 g 选为一个经过平移和缩放的指数分布 $Y \sim 5 + \text{Exp}(1)/5$, 则

$$g(y) \propto \tilde{g}(y) = \exp(-5(y-5)), y \geq 5$$

目标分布的 PDF 为

$$f(x) \propto \tilde{f}(x) = \exp(-x^2/2), x \geq 5$$

选取 \tilde{c} 使得 $\tilde{f}(x) \leq \tilde{c}\tilde{g}(x), \forall x \geq 5$, 如

$$\tilde{c} = \sup_{x \geq 5} \frac{\tilde{f}(x)}{\tilde{g}(x)} = \frac{\tilde{f}(5)}{\tilde{g}(5)} = \exp(-5^2/2)$$

即可使用 Algorithm 2 得到目标分布的样本

A-R 方法的应用

- **Normal polar**

在 Box-Muller 变换中使用一些 A-R 方法可以避免 \sin, \cos 运算

Algorithm 3 Box-Muller with Acceptance-rejection

repeat

$$V_j \stackrel{iid}{\sim} \mathbf{U}(-1, 1), j = 1, 2.$$

$$S = V_1^2 + V_2^2$$

until $S < 1$

return two independent $N(0, 1)$ random variables

$$Z_1 = V_1 \sqrt{-2 \log(S)/S}$$

$$Z_2 = V_2 \sqrt{-2 \log(S)/S}$$

混合分布抽样

让一个分布 $f(\cdot|\theta)$ 中的参数 θ 由另一个分布 G 随机产生，得到一个新的混合分布，这样定义的混合分布可如下抽样：

$$\theta \sim G, X | \theta \sim f(\cdot | \theta)$$

混合分布抽样

让一个分布 $f(\cdot|\theta)$ 中的参数 θ 由另一个分布 G 随机产生, 得到一个新的混合分布, 这样定义的混合分布可如下抽样:

$$\theta \sim G, X | \theta \sim f(\cdot | \theta)$$

一些常见的混合分布:

- **Beta-binomial 分布.** 有三个参数 $\alpha > 0, \beta > 0$ 和正整数 $n \geq 1$.
 $X \sim \text{Beta-binomial}(\alpha, \beta, n)$ 可以按如下方式抽样:

$$\theta \sim \text{Beta}(\alpha, \beta), X | \theta \sim \text{Bin}(n, \theta).$$

混合分布抽样

让一个分布 $f(\cdot|\theta)$ 中的参数 θ 由另一个分布 G 随机产生, 得到一个新的混合分布, 这样定义的混合分布可如下抽样:

$$\theta \sim G, X | \theta \sim f(\cdot | \theta)$$

一些常见的混合分布:

- **Beta-binomial 分布.** 有三个参数 $\alpha > 0, \beta > 0$ 和正整数 $n \geq 1$.
 $X \sim \text{Beta-binomial}(\alpha, \beta, n)$ 可以按如下方式抽样:

$$\theta \sim \text{Beta}(\alpha, \beta), X | \theta \sim \text{Bin}(n, \theta).$$

- **复合 Poisson 分布.** $X_i \stackrel{iid}{\sim} F, i = 1, \dots, N, N \sim \text{Po}(\lambda)$, 称随机变量 $S = \sum_{i=1}^N X_i$ 服从复合 Poisson 分布, 该分布有两个参数: $\lambda > 0$ 和分布 F

混合分布抽样

- **Negative binomial 分布.** 对一系列成功概率为 p 的独立 Bernoulli 试验, 在取得第 r 次成功时经历的失败次数 $X \sim \text{Negbin}(r, p)$:

$$P(X = k) = \binom{k+r-1}{r-1} p^r (1-p)^k, \quad k = 0, 1, \dots$$

$\text{Negbin}(r, p)$ 有如下等价的 Poisson-Gamma 混合分布:

$$X \sim \text{Po}(\lambda) \text{ for } \lambda \sim \text{Gam}\left(r, \frac{p}{1-p}\right)$$

混合分布抽样

- **Negative binomial 分布**. 对一系列成功概率为 p 的独立 Bernoulli 试验, 在取得第 r 次成功时经历的失败次数 $X \sim \text{Negbin}(r, p)$:

$$P(X = k) = \binom{k+r-1}{r-1} p^r (1-p)^k, \quad k = 0, 1, \dots$$

$\text{Negbin}(r, p)$ 有如下等价的 Poisson-Gamma 混合分布:

$$X \sim \text{Po}(\lambda) \text{ for } \lambda \sim \text{Gam}\left(r, \frac{p}{1-p}\right)$$

- **Noncentral χ^2 分布**. 如果 $\chi^2_{(\nu)}(\lambda)$ 的自由度 ν 不是整数 ($\nu > 0$), 基于 n 个正态随机变量平方和的抽样失效, 此时可从如下等价的 χ^2 -Poisson 混合分布抽样:

$$K \sim \text{Po}(\lambda/2), \quad Y | K \sim \chi^2_{(\nu+2K)}$$