随机变量的产生方法

王璐

生成随机变量是统计模拟的一个基本工具。我们可以用物理方法得到一组真实的随机数，比如反复抛掷硬币、骰子、抽签、摇号等，这些方法得到的随机数质量好，但是数量不能满足随机模拟的需要。主流的方法是使用计算机产生**伪随机数**。伪随机数是由计算机算法生成的序列 $\{x_i, i = 1, 2, \dots\}$, 因为计算机算法的结果是固定的，所以伪随机数不是真正的随机数，但是好的伪随机数序列可以做到与理论上真正的分布 F 无法通过统计检验区分开，所以我们也把计算机生成的伪随机数视为随机数。

需要生成某种分布的随机数时，一般先产生服从均匀分布的随机数，然后再将其转换为服从其它分布的随机数。

1 均匀分布随机变量的产生

计算机中伪随机数序列是迭代生成的，即 $x_n = g(x_{n-1}, x_{n-2}, \dots, x_{n-p})$, g 是确定的函数。均匀分布随机数发生器首先生成的是在集合 $\{0, 1, \dots, M\}$ 或 $\{1, 2, \dots, M\}$ 上离散取值的服从离散均匀分布的随机数，然后除以 M 或 $M + 1$ 变成 $[0, 1]$ 内的值当作服从连续均匀分布的随机数。这种方法实际上只取了有限个值，因为取值个数有限，根据算法 $x_n = g(x_{n-1}, x_{n-2}, \dots, x_{n-p})$ 可知序列一定在某个时间后发生重复，使得序列发生重复的间隔 T 叫做随机数发生器的周期。好的随机数发生器可以保证 M 很大且周期很长。现在常用的均匀分布随机数发生器由线性同余法、反馈位寄存器法以及随机数发生器的组合。这部分内容主要参考李东风 (2016) 第二章 2.1 节。

1.1 线性同余发生器

Definition 1.1 (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** (congruential)，记为 $i \equiv j \pmod{M}$。否则称 i 与 j 关于 M 不同余。

例如

$$11 \equiv 1 \pmod{10},\ -9 \equiv 1 \pmod{10}.$$

对于整数 A, 用 $A \pmod M$ 表示 A 除以 M 的余数，显然 A 和 $A \pmod M$ 同余，且 $0 \leq A \pmod M < M$。

线性同余发生器利用求余运算生成随机数，其递推公式为

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

其中 a 和 c 是事先设定的整数。取某个整数初值 x_0 后可以往下递推得到序列 $\{x_n\}$。注意到 $0 \leq x_n < M$, 令 $R_n = x_n/M$, 则 $R_n \in [0, 1)$, 最后把序列 $\{R_n\}$ 作为均匀分布的随机数序列输出。

因为线性同余法的递推算法仅依赖于前一项，序列元素取值只有 M 个可能值，所以产生的序列 $x_0, x_1, \ldots$ 一定会重复。若存在正整数 n 和 m 使得 $x_n = x_m (n > m)$, 则必有 $x_{n+k} = x_{m+k}$, $k = 1, 2, \ldots$, 即 $x_n, x_{n+1}, x_{n+2}, \ldots$ 重复了 $x_m, x_{m+1}, x_{m+2}, \ldots$, 称这样的 $n - m$ 的最小值 T 为此随机数发生器在初值 x_0 下的周期。由序列取值的有限性可知 $T \leq M$。

练习 1: 计算线性同余发生器

$$x_n = 7x_{n-1} + 7 \pmod{10}, n = 1, 2, \ldots$$

取初值 $x_0 = 7$ 的周期。(数列为 $7, 6, 9, 0, 7, 6, 9, 0, 7, \ldots$, 周期为 $T = 4$)

练习 2: 计算线性同余发生器

$$x_n = 5x_{n-1} + 1 \pmod 8, n = 1, 2, \ldots$$

取初值 $x_0 = 1$ 的周期。(数列为 $1, 6, 7, 4, 5, 2, 3, 0, 1, 6, 7, \ldots$, 周期为 $T = 8 = M$, 达最大周期)

当线性同余发生器从某个初值 x_0 出发达到最大周期 M, 也称**满周期**，则初值 x_0 取任意整数产生的序列都会达到满周期，序列总是从 x_M 开始重复。如果发生器从 x_0 出发不是满周期的，那么它从任何整数出发都不是满周期的。适当选取 M, a, c 可以使产生的随机数序列和真正的 $U[0, 1]$ 随机数表现接近。

Theorem 1. *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*

常取 $M = 2^L$, L 为计算机中整数的位数。根据定理1, 可取 $a = 4m + 1$, $c = 2n + 1$ (m 和 n 是任意正整数)，这样的线性同余发生器是满周期的。例如 Kobayashi 提出了如下的满周期 2^{31} 的线性同余发生器

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

其周期较长，统计性质比较好。

- 好的均匀分布随机数发生器应该周期足够长，统计性质符合均匀分布。把同余法生成的数列看成随机变量序列 $\{X_n\}$，在满周期时，可认为 X_n 是从 $\{0,1,\ldots,M-1\}$ 中随机等可能选取的，即

$$P(X_n = i) = 1/M,\ i = 0,1,\ldots,M-1$$

此时

$$E(X_n) = \sum_{i=0}^{M-1} i\frac{1}{M} = \frac{M-1}{2}$$

$$Var(X_n) = E(X_n^2) - [E(X_n)]^2 = \sum_{i=0}^{M-1} i^2\frac{1}{M} - \frac{(M-1)^2}{4} = \frac{1}{12}(M^2-1)$$

于是当 M 很大时

$$E(R_n) = E(X_n/M) = \frac{1}{2} - \frac{1}{2M} \approx \frac{1}{2}$$

$$Var(R_n) = Var(X_n/M) = \frac{1}{12} - \frac{1}{12M^2} \approx \frac{1}{12}$$

可见生成数列的期望和方差很接近均匀分布。

- 好的随机数发生器还应该有很好的随机性，产生的序列不应该有规律，序列之间独立性好。但是随机数发生器产生的序列是由确定的公式生成，不可能做到真正独立，至少我们要求序列的自相关性较弱。对于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

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

所以应该将 a 选为较大的值 ($a < M$)。

1.2 FSR 发生器

线性同余发生器产生一维均匀分布随机数效果很好，但产生的多维随机向量相关性大，分布不均匀。而且线性同余法的周期不可能超过 2^L。Tausworthe (1965) 提出一种新的做法——反馈位移寄存器法 (FSR), 对这些方面有改进。

FSR 按照如下递推法则生成一系列取值为 0 或 1 的数 $\alpha_1, \alpha_2, \ldots$, 每个 α_k 由前面若干个 $\{\alpha_i\}$ 的线性组合除以 2 的余数产生：

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

其中每个系数 c_i 只取 0 或 1，这样的递推可以利用程序语言中的逻辑运算快速实现。比如，如果 FSR 算法中的系数 $(c_1, c_2, \ldots, c_p)$ 仅有两个为 1，e.g. $c_p = c_{p-q} = 1 (1 < q < p)$, 递推法则可写为：

$$\begin{aligned}\alpha_k &= \alpha_{k-p} + \alpha_{k-p+q} \pmod 2 \\ &= \begin{cases} 0 & \text{if } \alpha_{k-p} = \alpha_{k-p+q} \\ 1 & \text{if } \alpha_{k-p} \neq \alpha_{k-p+q}. \end{cases}\end{aligned}$$

这可以用计算机的异或运算 $\oplus$ 进行快速计算：

$$\alpha_k = \alpha_{k-p} \oplus \alpha_{k-p+q},\ k = 1, 2, \ldots$$

给定初值 $(\alpha_{-p+1}, \alpha_{-p+2}, \ldots, \alpha_0)$ 递推得到序列 $\{\alpha_k : k = 1, 2, \ldots\}$ 后，依次截取长度为 M 的二进制序列组合成整数 x_n, 再令 $R_n = x_n/2^M$。巧妙选择递推系数和初值（种子）可以得到很长的周期，且作为多维均匀分布随机向量的发生器性质较好。在上述 $c_p = c_{p-q} = 1 (1 < q < p)$ 的例子中，递推算法只需要异或运算，不受计算机字长限制，适当选取 p, q 后周期可以达到 $2^p - 1$ (如取 $p = 98$)。

1.3 组合发生器法

随机数设计中比较困难的是独立性和多维的分布。可以考虑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) 提出了组合同余法，组合两个同余发生器，一个用来 “搅乱” 次序。将两个同余发生器记为 A 和 B。用 A 产生 m 个随机数 (e.g. m=128)，存放在数组 $T = (t_1, t_2, \ldots, t_m)$。需要产生 x_n 时，从 B 中生成一个随机下标 $j \in \{1, 2, \ldots, m\}$，取 $x_n = t_j$, 然后从 A 再生成一个新随机数替代 T 中的 t_j, 如此重复。这样组合可以增强随机性，加大周期 (可超过 2^L)。也可以只使用一个发生器，用 x_{n-1} 来选择下标。

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：

$$\begin{aligned}U_n &= 171 U_{n-1} \pmod{30269} \\ V_n &= 172 V_{n-1} \pmod{30307} \\ W_n &= 170 W_{n-1} \pmod{30323}\end{aligned}$$

做线性组合并求余：

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

这个组合发生器的周期约有 7×10^{12}, 超过 $2^{31} \approx 2 \times 10^9$.

在 R 软件中，用 `runif(n)` 产生 n 个 $U(0,1)$ 均匀分布的随机数。R 提供了若干种随机数发生器，可以用 `RNGkind()` 函数切换。在使用随机数进行模拟时，如果希望模拟的结果可重复，就需要在模拟开始时设置固定的随机数种子。在 R 中，可以用函数 `set.seed(m)` 来设置种子，其中 m 是任意整数。

1.4 随机数的检验

对均匀分布随机数发生器产生的序列 $\{R_i, i = 1, 2, \ldots, n\}$, 可以进行各种检验确认其均匀性。一些检验的想法有：

- 把 [0,1] 等分成 K 段，用 Pearson's χ^2 test 检验 $\{R_i, i = 1, 2, \ldots, n\}$ 落在每一段的概率是否近似为 $1/K$.
  - Pearson's χ^2 test 可以检验样本落入若干互斥分类的概率分布是否等于某个特定的离散分布 (reference distribution)。其原理是通过每个分类的实际观察次数与理论期望次数之差构造统计量。
- 用 Kolmogorov-Smirnov (K-S) test 检验 $\{R_i, i = 1, 2, \ldots, n\}$ 是否近似服从 $U[0, 1]$ 分布。
  - K-S test 可以检验样本是否服从某个特定的一维连续分布 (reference distribution)，其原理是通过定义样本的 empirical CDF 与 reference CDF 的距离构造统计量。
- 把 $\{R_i, i = 1, 2, \ldots, n\}$ 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 检验这些组合得到的 $\mathbb{R}^d$ 向量落在每个子集的概率是否近似为 $1/K^d$.
- $\cdots$

2 非均匀分布随机变量的产生

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石。常用的科学计算软件, 如 R、Matlab，都提供了很多常见的非均匀分布的抽样函数，如 normal, Poisson, binomial, exponential, gamma, beta, etc. 但是有时候我们可能需要从某个特殊分布中抽样，而这些软件没有提供现成的抽样函数。因此我们需要理解这些非均匀分布的随机数是如何生成的，以便在必要的时候 make a custom solution. 在这部分我们会学习一些通用型的方法，如 CDF 逆变换, acceptance-rejection sampling 等。这部分内容主要参考Owen (2013) Chapter 4.

2.1 CDF 逆变换

将均匀分布的随机变量转化为非均匀分布的随机变量最直接的方法是 CDF 逆变换 (inverse CDF transform). 理论上这个方法适用于任何 CDF 的逆函数已知的分布。

Definition 2.1 (Cumulative distribution function (CDF))**.** 一个随机变量 X 的 CDF $F(x)$ 定义为

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

任一分布可由它的 CDF 完全刻画。CDF 有如下性质:

- $F(+\infty) = 1$
- $F(-\infty) = 0$
- 右连续: $\lim\limits_{x \to y^+} F(x) = F(y)$

对于一个连续分布，CDF 和它的 PDF (probability density function) $f(x) \geq 0$ 的关系是

$$P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a).$$

假设随机变量 X 的 PDF $f(x) > 0, \forall x \in \mathbb{R}$. 那么它的 CDF $F(x)$ 单调且连续，因此存在逆函数 F^{-1}. 由于 $0 \leq F(x) \leq 1$, 如果我们抽取 $U \sim \boldsymbol{U}[0,1]$, 然后考察 $Y = F^{-1}(U)$ 的分布会发现

$$\begin{aligned} P(Y \leq y) &= P\left(F^{-1}(U) \leq y\right) \\ &= P\left(F\left(F^{-1}(U)\right) \leq F(y)\right) \\ &= P(U \leq F(y)) \\ &= F(y) \end{aligned}$$

因此 Y 服从 CDF 为 $F(\cdot)$ 的分布，记为 $Y \sim F$. 这就是 **CDF 逆变换的基本想法**。然而在实践中，它还面临很多问题。比如对离散分布

$$P(X = x_k) = p_k,\ k = 1, 2, \ldots$$

它的 CDF $F(x)$ 不连续且不可逆; 有时我们需要抽样的分布既有离散又有连续的部分，设 F_d 是一个离散分布的 CDF, F_c 是一个连续分布的 CDF, $0 < \lambda < 1$, 则 $\lambda F_d + (1-\lambda)F_c$ 也是一个 CDF，如图1所示，它在一些点的逆函数也无法定义。

上述问题可以通过为 CDF 定义如下的**广义逆**得到解决。

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

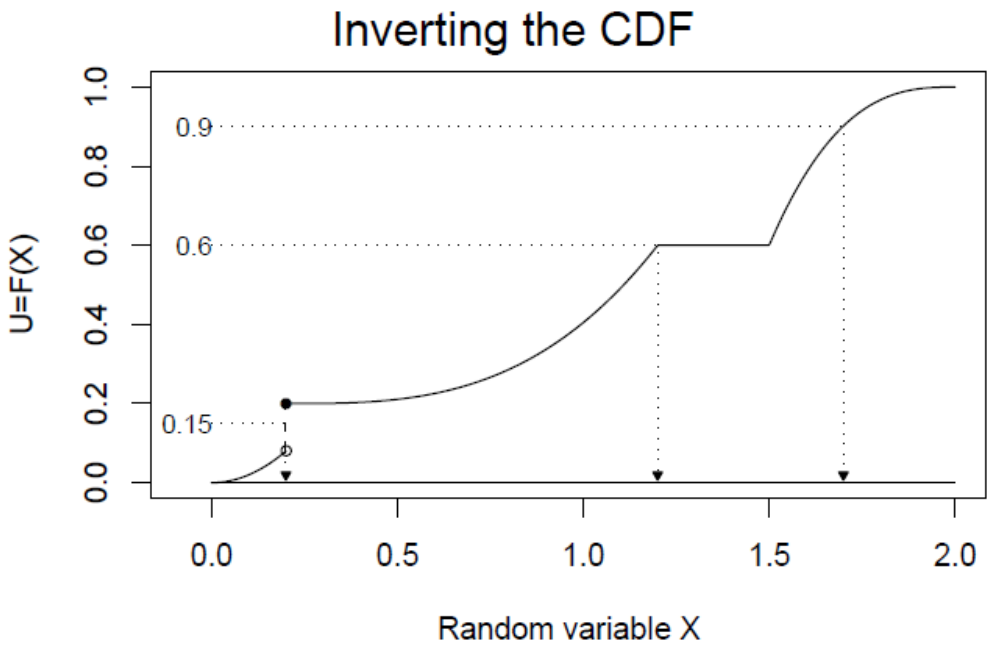

Figure 1: 随机变量 X 在 [0,2] 上取值，但它在区间 [1.2,1.5] 取值的概率为 0，图中的实线是 X 的 CDF, 它在 $x = 0.2$ 处有一个跳跃。

由于 $F(+\infty) = 1$，(1)中的集合总是非空的，因此总可以找到下确界。图1展示了 F 的广义逆在几个点的值：$F^{-1}(0.15) = 0.2$, $F^{-1}(0.6) = 1.2$, $F^{-1}(0.9) = 1.7$.

CDF 逆变换: 设 F 是一个 CDF, F^{-1} 是由(1)定义的逆函数。如果随机变量 $U \sim \boldsymbol{U}[0,1]$, 令 $X = F^{-1}(U)$, 则变换得到的随机变量 $X \sim F$.

Remarks

- 如果 $U \sim \boldsymbol{U}[0,1]$, 那么 $1 - U \sim \boldsymbol{U}[0,1]$, 因此 $F^{-1}(1-U) \sim F$. 有时候 $F^{-1}(1-U)$ 具有更简单的形式。
- 在 CDF 逆变换中，我们是对服从 $\boldsymbol{U}[0,1]$ 的随机变量做逆变换，上述想法可以进一步推广。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

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

  服从分布 G, 因为 $F(X) \sim \boldsymbol{U}[0,1]$. 函数 $G^{-1}(F(\cdot))$ 也被称为 **QQ 变换**，因为它将分布 F 的分位数 (quantile) 转换为分布 G 下相应的分位数。即如果 x 是 F 的 0.1-quantile ($P(X \le x) = F(x) = 0.1$), 则 $y = G^{-1}(F(x))$ 是 G 的 0.1-quantile ($P(Y \le y) = G(y) = 0.1$).
- 有时 $G^{-1}(F(\cdot))$ 的形式比 G^{-1} 或 F 都简单，我们可以直接将 $X \sim F$ 转化为 $Y = G^{-1}(F(X)) \sim G$，而不需要知道 F 或 G^{-1} 的具体形式。比如知道如何从 $N(0,1)$ 抽样

后，如果想得到 $N(\mu,\sigma^2)$ 的样本，可以做变换 $Y=\mu+\sigma Z$, $Z\sim N(0,1)$，这其实是一个 QQ 变换(2).

2.1.1 CDF 逆变换举例

很多重要的一元分布可以用逆变换的方法进行抽样，这里列举一些有用的例子。

- **指数分布 (Exponential distribution)**. 标准的指数分布 Exp(1) 的 PDF 是

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

它的 CDF 是

$$F(x)=P(X\le x)=\int_0^x f(t)dt=1-e^{-x},\ x>0.$$

CDF 的逆是

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

因此先抽取 $U\sim \boldsymbol{U}(0,1)$, 再令 $X=-\log(1-U)$ 可以产生服从 Exp(1) 的样本。或者可直接令 $X=-\log(U)$ ~Exp(1), 因为 $1-U\sim \boldsymbol{U}(0,1)$, $F^{-1}(1-U)=-\log(U)$.

一般的指数分布有一个 rate parameter $\lambda>0$, 记为 Exp(λ), 有时人们会用 scale parameter $\theta=1/\lambda$ 取代 λ 描述指数分布。$Y\sim$ Exp(λ) 的期望是 $E(Y)=1/\lambda$, PDF 是

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

如果 $X\sim$ Exp(1), 则 $X/\lambda\sim$ Exp(λ). 因此可以通过变换 $Y=-\log(U)/\lambda$ 获得 Exp(λ) 的样本。

  - 指数分布常用于描述一段时间的分布，但它的一个重要特性是**无记忆性** (memoryless). 如果 $X\sim$ Exp(λ), 则

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

它与 x 无关。因此指数分布不适合描述一个非耐用品的生命周期，比如灯泡的寿命等。后面介绍的 Weibull 分布更适合描述这种情况。

- **Bernoulli 分布**。如果 $X\sim$ Bern(p), 则 $P(X=1)=p$, $P(X=0)=1-p$. 从 Bernoulli 分布抽样可以利用变换 $X=\mathbf{1}(1-U\le p)$ 或 $X=\mathbf{1}(U\le p)$ 实现。

- **Cauchy 分布**。Cauchy 分布是 t 分布的一个特例，具有**厚尾** (heavy tails) 的特性。Cauchy 分布的 PDF 是

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

密度函数 $f(x)$ 在 $x \to \pm\infty$ 时下降地很慢以至于 $E(|X|) = \infty$. Cauchy 的 CDF 如下

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

对 CDF 求逆，我们可以利用如下变换从 Cauchy 分布抽样，

$$X = \tan\left(\pi \cdot (U - 1/2)\right),\ U \sim \boldsymbol{U}(0, 1).$$

从几何角度看，Cauchy 变量是一个在 $(-\pi/2, \pi/2)$ 上均匀分布的随机角的正切 (tangent). Cauchy 分布可用于描述看似像正态分布，又存在一些极端大或小的观察值的情形。

- **离散均匀分布**。对于在 $\{0, 1, \ldots, k-1\}$ 上均匀分布的离散随机变量，可以令

$$X = \lfloor kU \rfloor,\ U \sim \boldsymbol{U}(0, 1)$$

其中 $\lfloor kU \rfloor$ 表示 $\leq kU$ 的最大整数。如果我们想从 $\boldsymbol{U}\{1, 2, \ldots, k\}$ 上抽样，可令 $X = \lceil kU \rceil$, 其中 $\lceil kU \rceil$ 表示 $\geq kU$ 的最小整数。

- **Poisson** 分布。如果 $X \sim \text{Po}(\lambda)$, 则

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

且 $E(X) = \lambda$. Algorithm 1 是 Devroye (1986) 基于 CDF 逆变换提出的从 Poisson 分布抽样的方法。算法中条件 $U > q$ 会被检验 $X + 1$ 次，因此算法平均需要的迭代步数是 $E(X + 1) = \lambda + 1$. 显然对 λ 较大的情形，使用该方法抽样会很慢。

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

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

while $U > q$ **do**

$X = X + 1$

$p = p\lambda / X$

$q = q + p$

return X

- **正态分布**。用 Φ 表示 $N(0,1)$ 的 CDF，如果 $U \sim \boldsymbol{U}(0,1)$, 那么 $Z = \Phi^{-1}(U) \sim N(0,1)$. 但是 Φ 和 Φ^{-1} 都没有解析形式，这使得用 CDF 逆变换从正态分布抽样变得十分困难。不过通过对 Φ^{-1} 做精确的数值近似，使用逆变换抽样仍然是可行的，比如 Wichura (1988) 提出的算法 AS241, 其中使用的近似函数的精度非常高

$$\frac{|\hat{\Phi}_W^{-1}(u) - \Phi^{-1}(u)|}{|\Phi^{-1}(u)|} < 10^{-15}, \quad \min(u, 1-u) > 10^{-316}.$$

后面会介绍另一种更简便的方法——Box-Muller 变换。

- **Weibull 分布**。Weibull 分布是对指数分布的推广，它的 PDF

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

有两个参数 $\sigma > 0$, $k > 0$. $k = 1$ 时, Weibull 分布退化为指数分布 Exp$(1/\sigma)$. Weibull 分布的 CDF 是

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

因此对 $U \sim \boldsymbol{U}(0,1)$ 做变换 $X = \sigma\left(-\log(1-U)\right)^{1/k}$ 可实现从 Weibull 分布抽样。

之前我们讨论过指数分布的无记忆性，为了探讨 Weibull 分布的基本特性，引入一个新的概念——**hazard function**.

Definition 2.2 (Hazard function)**.** 对一个随机变量 $X > 0$, 它的 hazard function $h(x)$ 定义为

$$h(x) = \lim_{t \to 0^+} \frac{1}{t} P(X \le x + t \mid X \ge x), \ x > 0.$$

如果用 X 表示一个灯泡的生命周期，hazard function $h(x)$ 表示给定灯泡在 x 时刻正常工作，它瞬间出现故障的概率（密度），也称瞬时失败概率（instantaneous probability of failure）。

练习: 求指数分布 Exp(λ) 和 Weibull 分布 Weibull(k, σ) 的 hazard function. (一个有用的近似: $e^x \approx 1 + x$, $x \to 0$)

- 指数分布 Exp(λ) 的 hazard function

$$h(x) = \lim_{t \to 0^+} \frac{1}{t}(1 - e^{-\lambda t}) = \lim_{t \to 0^+} \frac{\lambda t}{t} = \lambda$$

可以看到指数分布的瞬时失败概率是常数，与当前时刻 x 无关。

– Weibull 分布的 hazard function

$$
\begin{aligned}
h(x) &= \lim_{t\to 0^+} \frac{1}{t} \cdot \frac{P(x \le X \le x+t)}{P(X \ge x)} \\
&= \lim_{t\to 0^+} \frac{1}{t} \cdot \frac{F(x+t) - F(x)}{1 - F(x)} \\
&= \lim_{t\to 0^+} \frac{1}{t} \cdot \frac{-\exp\left[-\left(\frac{x+t}{\sigma}\right)^k\right] + \exp\left[-\left(\frac{x}{\sigma}\right)^k\right]}{\exp\left[-\left(\frac{x}{\sigma}\right)^k\right]} \\
&= \lim_{t\to 0^+} \frac{1}{t} \cdot \left\{-\exp\left[-\left(\frac{x+t}{\sigma}\right)^k + \left(\frac{x}{\sigma}\right)^k\right] + 1\right\} \\
&= \lim_{t\to 0^+} \frac{1}{t} \left[\left(\frac{x+t}{\sigma}\right)^k - \left(\frac{x}{\sigma}\right)^k\right] \\
&= \frac{d}{dx}\left(\frac{x}{\sigma}\right)^k \\
&= k \cdot \frac{x^{k-1}}{\sigma^k}
\end{aligned}
$$

可以看到

* $k < 1$ 时, 瞬时失败概率 $h(x)$ 随时间 x 递减。这种事件有可能发生，比如婴儿在刚出生时死亡概率很高，但随时间推移死亡概率在下降。
* $k = 1$ 时，Weibull 退化为指数分布，瞬时失败概率是常数。
* $k > 1$ 时，瞬时失败概率 $h(x)$ 随时间 x 递增。这种事件很常见，比如任何会老化的产品的生命周期。

References

李东风 (2016). 统计计算. 高等教育出版社.

Owen, A. B. (2013). *Monte Carlo theory, methods and examples.*

