表 4.1 标准正态分布表 $\Phi(x)=\int_{-\infty}^{x}e^{-t^2/2}/\sqrt{2\pi}dt.$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

4.5 连续随机变量函数的分布

当知道一个随机变量的概率分布后, 经常会考虑它的一些函数的分布, 例如已知一个圆的直径 X 服从均匀分布 $U(a,b)$, 则可以考虑圆的面积 $Y=\pi(X/2)^2$ 的分布. 一般地, 若已知随机变量 X 的分布函数, 以及 $g(x)$ 是定义在随机变量 X 所有可能取值的函数, 则称 $Y=g(X)$ 为随机变量 X 的函数, 很显然 Y 也是随机变量. 研究的问题可以归纳为: 若已知随机变量 X 的概率分布和函数 $g(x)$, 如何求解随机变量 $Y=g(X)$ 的概率分布.

设离散型随机变量 X 的分布列为 $p_k = P(X = x_k)$ $(k = 1, 2, \cdots)$, 求解随机变量 $Y = g(X)$ 的分布列较为简单, 将相等的项 $g(x_i) = g(x_j)$ 合并, 相应的概率相加即可. 例如

例 4.12 若随机变量 X 的概率分布列为 $P(X = k) = 1/2^k$ $(k = 1, 2, \cdots)$, 求随机变量 $Y = \cos(\pi X/2)$ 的分布列.

解 当 $n = 1, 2, \cdots$ 时有

$$\cos(k\pi/2) = \begin{cases} 1 & k = 4n \\ 0 & k = 2n-1 \\ -1 & k = 4n-2\ , \end{cases}$$

所以 $\cos(\pi X/2)$ 的取值为 $\{-1, 0, +1\}$, 其概率分别为

$$\begin{aligned} P(Y=1) &= \sum_{n=1}^{\infty} P(X=4n) = \sum_{n=1}^{\infty} \frac{1}{2^{4n}} = \frac{1}{15}\ , \\ P(Y=0) &= \sum_{n=1}^{\infty} P(X=2n-1) = \sum_{n=1}^{\infty} \frac{1}{2^{2n-1}} = \frac{2}{3}\ , \\ P(Y=-1) &= \sum_{n=1}^{\infty} P(X=4n-2) = \sum_{n=1}^{\infty} \frac{1}{2^{4n-2}} = \frac{4}{15}\ , \end{aligned}$$

由此给出随机变量 Y 的分布列.

针对连续型随机变量, 一般采用概率密度函数来刻画其概率分布. 若已知函数 $g(x)$ 和随机变量 X 的概率密度函数为 $f_X(x)$, 求解随机变量 $Y = g(X)$ 的概率密度 $f_Y(y)$, 通常分为以下两步:

1) 求解随机变量 Y 的分布函数 $F_Y(y) = P(Y \leqslant y) = P(g(X) \leqslant y) = \int_{g(x)\leqslant y} f_X(x)dx$;
2) 求解随机变量 Y 的密度函数 $f_Y(y) = F'_Y(y)$.

求解此类问题常用到的数学工具是积分求导公式: 设函数 $F(y) = \int_{\psi(y)}^{\varphi(y)} f(x)dx$, 则有

$$F'(y) = f(\varphi(y))\varphi'(y) - f(\psi(y))\psi'(y)\ . \tag{4.2}$$

下面来看两个例子.

例 4.13 设随机变量 X 的概率密度为 $f_X(x)$, 求 $Y = X^2$ 的概率密度.

解 根据分布函数的定义可知

$$F_Y(y) = P(Y \leqslant y) = P(X^2 \leqslant y)\ .$$

当 $y \leqslant 0$ 时有 $F_Y(y)=0$; 当 $y>0$ 时有

$$F_Y(y)=P(X^2\leqslant y)=P(-\sqrt{y}\leqslant x\leqslant\sqrt{y})=\int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx\ ,$$

根据 (4.2) 得到密度函数

$$f_Y(y)=F_Y'(y)=f_X(\sqrt{y})/2\sqrt{y}+f_X(-\sqrt{y})/2\sqrt{y}\ .$$

综上所述有

$$f_Y(y)=\begin{cases}\left(f_X(-\sqrt{y})+f_X(\sqrt{y})\right)/2\sqrt{y} & y>0\ ,\\ 0 & y\leqslant 0\ .\end{cases}$$

例 4.14 设随机变量 X 的密度函数为

$$f(x)=\begin{cases}x/8 & 0<x<4\\ 0 & \text{其它},\end{cases}$$

求 $Y=2X+8$ 的密度函数.

解 首先求解分布函数

$$F_Y(y)=P(Y\leqslant y)=P(2X+8\leqslant y)=P(X\leqslant(y-8)/2)=F_X((y-8)/2)\ ,$$

求导可得密度函数

$$f_Y(y)=f_X((y-8)/2)/2=\begin{cases}(y-8)/32 & (y-8)/2\in[0,4]\\ 0 & \text{其它}\end{cases}=\begin{cases}(y-8)/32 & y\in[8,16]\\ 0 & \text{其它}\ .\end{cases}$$

针对一般情况有如下定理:

定理 4.8 设随机变量 X 的概率密度是定义在实数域上的函数 $f_X(x)$, 函数 $y=g(x)$ 处处可导且严格单调 (即 $g'(x)>0$ 或 $g'(x)<0$), 令其反函数 $x=g^{-1}(y)=h(y)$, 则 $Y=g(X)$ 的概率密度为

$$f_Y(y)=\begin{cases}f_X(h(y))|h'(y)| & y\in(\alpha,\beta)\\ 0 & \text{其它},\end{cases}$$

其中 $\alpha=\min\{g(-\infty),g(+\infty)\}$ 和 $\beta=\max\{g(-\infty),g(+\infty)\}$.

可将上述定理推广至区间函数 $x\in[a,b]$, 上述定理依旧成立, 此时有 $\alpha=\min\{g(a),g(b)\}$ 和 $\beta=\max\{g(a),g(b)\}$.

证明 证明思路与前面的例题类似, 这里不妨假设 $g'(x) > 0$, 同理可以考虑 $g'(x) < 0)$ 的情况. 根据 $g'(x) > 0$ 可知其反函数 $x = h(y)$ 也严格单调, 且 $g(x) \in [\alpha, \beta]$. 因此, 当 $y \leqslant \alpha$ 时, 有 $F_Y(y) = 0$; 当 $y \geqslant \beta$ 时有 $F_Y(y) = 1$; 当 $\alpha < y < \beta$ 时,

$$F_Y(y) = P(g(X) < y) = P(X \leqslant h(y)) = F(h(y)).$$

于是有随机变量 Y 的概率密度

$$f_Y(y) = F'(h(y)) \cdot h'(y) = f_X(h(y)) \cdot h'(y).$$

根据$x = h(y)$ 严格单调可知 $h'(y) > 0$.

定理 4.9 设 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则 $Y = aX + b$ $(a > 0)$ 服从正太分布 $\mathcal{N}(a\mu + b, a^2\sigma^2)$.

证明 设函数 $g(x) = ax + b$, 可得 $\alpha = -\infty$, $\beta = +\infty$, 以及 $y = g(x)$ 的反函数为

$$x = h(y) = (y - b)/a,$$

且有 $h'(y) = 1/a$. 根据定理 4.8 可知

$$f_Y(y) = \frac{1}{a} f_X(\frac{y-b}{a}) = \frac{1}{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-(\frac{y-b}{a} - \mu)^2)^2/2\sigma^2} = \frac{1}{\sqrt{2\pi}a\sigma} e^{-(y-b-a\mu)^2/2a^2\sigma^2},$$

由此证明了 $Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

例 4.15 设随机变量 X 的分布函数 $F_X(x)$ 是严格单调的连续函数, 则 $Y = F_X(X) \sim U(0, 1)$.

证明 令 $Y = F_X(X)$ 的分布函数为 $G(y)$, 则

$$G(y) = P(Y \leqslant y) = P(F_X(X) \leqslant y)\ .$$

因为分布函数 $F_X(x) \in [0, 1]$, 当 $y < 0$ 时有 $G(y) = 0$; 当 $y \geqslant 1$ 时有 $G(y) = 1$; 当 $y \in [0, 1]$ 时, 由于 $F_X(X)$ 严格单调, 所以 $F_X^{-1}(y)$ 存在且严格单调, 于是有

$$G(y) = P(F_X(X) \leqslant y) = P\left(X \leqslant F_X^{-1}(y)\right) = F_X\left(F_X^{-1}(y)\right) = y\ .$$

综上所述有密度函数

$$f_Y(y) = \begin{cases} 1 & y \in [0, 1] \\ 0 & \text{其它}\ . \end{cases}$$

4.6 常用分布的随机数*

随机数在计算机仿真学和密码学等领域具有广泛的应用, 通过产生大量的随机数据来实现对真实世界的模拟. 可以利用物理随机过程来产生真实的随机数, 例如反复抛掷硬币、骰子、抽签、摇号等, 这些方法可以得到质量很高的随机数, 但其数量和类型通常较少、难以满足实际的需求.

现在的主流方法是使用计算机产生伪随机数, 通过计算机的确定算法来生成“类似”真随机数的伪随机数. 由于算法给出的结果总是确定的, 所以伪随机数并不是真正的随机数, 但是好的伪随机数序列与真实随机数序列表现几乎相同, 很难进行区分.

本节首先介绍如何生成在 $(0,1)$ 上均匀分布的随机数, 然后根据此随机数构造其它常用分布的随机数. 这里仅给出具体的构造方法, 关于其中的原理, 有兴趣的读者可以参考相关书籍.

4.6.1 区间 $(0,1)$ 上均匀分布的随机数

目前有很多方法生成 $(0,1)$ 上均匀分布的随机数, 这里介绍最常用的线性同余法. 通过下面的迭代方式产生一系列随机数 $x_1, x_2, \cdots, x_k$,

$$x_i = ax_{i-1} + c \qquad \bmod \quad m\ ,$$

其中 $1 \leqslant i \leqslant k \leqslant m$, 常数 x_0 为初始给定的种子. 为了获得较好的随机性, 通常 m 的取值应足够大, 如 $m = 2^k$ $(k = 31, 63)$, 常数 c 与 m 互质, 常数 $a-1$ 被 m 的因子整除, 例如一种可行的选择是

$$x_i = 31415926x_{i-1} + 453806245 \qquad \bmod \quad 2^{31}.$$

最后得到在区间 $(0,1)$ 上均匀分布的随机数 $x_1/m, x_2/m, \cdots, x_k/m$. 很多时候会根据实际情况选择不同的初始种子 x_0. 有了在 $(0,1)$ 上均匀分布的随机数, 则可以构造一些服从常用分布的随机数.

4.6.2 常用离散型分布的随机数

定理 4.10 若随机变量 $X \sim U(0,1)$, 以及 $F(y)$ 是某个离散型随机变量的分布函数, 其可能的取值为 $\{y_1, y_2, \cdots\}$, 不妨假设 $y_1 < y_2 < \cdots$. 设随机变量

$$Y = \begin{cases} y_1 & X \leqslant F(y_1) \\ y_i & X \in (F(y_{i-1}), F(y_i)]\ (i \geqslant 2)\ , \end{cases}$$

则 Y 的分布函数 $F_Y(y) = F(y)$ (如图 4.8 所示).

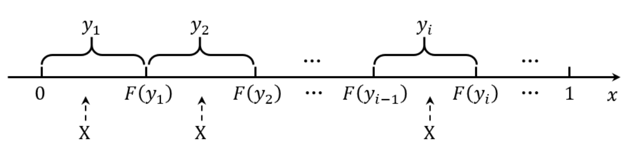

图 4.8 离散分布的随机数生成

证明 当 $y < y_1$ 时有

$$F_Y(y) = P(Y \leqslant y) \leqslant P(Y < y_1) = 0 = F(y)\ .$$

当 $y_1 \leqslant y < y_2$ 时根据 $X \sim U(0,1)$ 有

$$F_Y(y) = P(Y \leqslant y) = P(Y = y_1) = P(X \leqslant F(y_1)) = F(y_1) = F(y)\ .$$

当 $y \geqslant y_2$ 时再次根据 $X \sim U(0,1)$ 有

$$F_Y(y) = P(Y \leqslant y) = \sum_{y_i:\, y_i \leqslant y} P(Y = y_i) = \sum_{y_i:\, y_i \leqslant y} P\left(X \in (F(y_{i-1}), F(y_i)]\right) = \max_{y_i:\, y_i \leqslant y} F(y_i) = F(y)\ .$$

定理 4.10 说明可以利用均匀分布的随机数来构造其它离散型随机变量的随机数, 以下以伯努利分布为例, 可类似生成其它常用分布的随机数.

例 4.16 已知在区间 $(0,1)$ 上均匀分布的随机数 $x_1, x_2, \cdots, x_k$ $(k \geqslant 1)$, 如何生成参数为 p 的伯努利分布的随机数.

解 设随机变量 Y 服从参数为 p 的伯努利分布, 则 Y 的分布函数 $F(0) = 1-p$ 和 $F(1) = 1$. 根据定理 4.10, 当 $1 \leqslant i \leqslant k$ 时构造随机数

$$y_i = \begin{cases} 0 & x_i \leqslant F(0) = 1-p \\ 1 & x_i \in (F(0), F(1)] = (1-p, 1]\ , \end{cases}$$

其服从参数 p 的伯努利分布.

4.6.3 常用连续型分布的随机数

定理 4.11 若随机变量 $X \sim U(0,1)$, 且 $F(y)$ 是某一个连续的分布函数, 很显然反函数 $F^{-1}(y)$ 存在, 则随机变量

$$Y = F^{-1}(X)$$

的分布函数为 $F_Y(y) = F(y)$.

证明 根据分布函数的定义和 $F(x)$ 的单调性有

$$F_Y(y) = P(Y \leqslant y) = P(F^{-1}(X) \leqslant y) = P(X \leqslant F(y))\ .$$

再根据分布函数 $F(y)$ 的非负性和随机变量 $X \sim U(0,1)$ 有

$$F_Y(y) = F(y)\ ,$$

由此完成证明.

例 4.17 若已知在区间 $(0,1)$ 上均匀分布的随机数 $x_1, x_2, \cdots, x_k$ $(k \geqslant 1)$, 如何生成参数为 $\lambda = 4$ 的指数分布的随机数.

解 若随机变量 Y 服从参数为 $\lambda = 4$ 的指数分布, 则有连续的分布函数 $F_Y(y) = 1 - e^{-4\lambda y}$, 以及反函数 $F_Y^{-1}(y) = -\ln(1-y)/4$. 根据定理 4.11 可构造随机数

$$y_i = -\ln(1-x_i)/4 \qquad (i \in [k])$$

服从参数为 $\lambda = 4$ 的指数分布.

由于正太分布的分布函数不存在显示表达式, 所以它的反函数也不存在显示表达式, 因此不能利用定理 4.11 来直接生成服从正太分布的随机数. 这里简要介绍一种生成标准正太分布的随机数的方法、以及一些相关的结论, 详细的证明可参考相关书籍.

设 X 和 Y 是相互独立的标准正太分布随机变量, 则它们的极坐标

$$\begin{cases} R = \sqrt{X^2 + Y^2} \\ \theta = \arctan(Y/X) \end{cases}$$

相互独立, 且 R^2 服从参数为 $1/2$ 的指数分布, 而 θ 服从在 $(0, 2\pi)$ 上的均匀分布. 设相互独立的随机变量 X_1 和 X_2 在区间 $(0,1)$ 上服从均匀分布, 则有

$$R = (-2\ln X_1)^{1/2} \qquad \text{和} \qquad \theta = 2\pi X_2\ .$$

由此可得相互独立的标准正太分布随机变量

$$X = R\cos(\theta) = (-2\ln X_1)^{1/2}\cos(2\pi X_2) \quad \text{和} \quad Y = R\sin(\theta) = (-2\ln X_1)^{1/2}\sin(2\pi X_2)\ ,$$

这种产生标准正太分布随机变量的方法被称为 **Box-Muller 方法**.