博学　审问　慎思　明辨　笃行

概率论与数理统计

中山大學
SUN YAT-SEN UNIVERSITY

主讲老师：邝东阳，kuangdy@mail.sysu.edu.cn

助教：黄培，huangp89@mail2.sysu.edu.cn

范志隆，fanzhlong@mail2.sysu.edu.cn

Part 1
连续型分布

博学 | 审问 | 慎思 | 明辨 | 笃行

cdf v.s. pdf

定义 2.2.3. 若存在非负函数$f(x)$使得随机变量X的分布函数

$$F(x) = \int_{-\infty}^{x} f(x)dx, \quad for\ all\ x \ \in \mathbb{R}$$

则称随机变量X为连续型的，并称$f(x)$为X的概率密度函数(*pdf, probability density function*)。

注 2.3.1 由于在有限个点上改变被积函数的值不会改变积分的值，故：

1）可以在有限个点上不定义$p(x)$的值.

2）若$p(x)$是随机变量X的密度，$\tilde{p}(x)$仅在有限个点上与$p(x)$有不同的值，则$\tilde{p}(x)$也是X的密度.

pdf

1. $f(x) \geq 0, forall\ x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.
3. 若$f(x)$在x点连续，则$F'(x) = f(x)$.
4. 对任意的Borel可测集合B, $P(X \in B) = \int_{x\in B} f(x)dx$

反之，若一个函数$f(x)$满足如上1-2两条，则存在某个连续型随机变量使得$f(x)$为其概率密度。

pdf

定理 2.3.1 设 X 为随机变量，$p_X(x)$ 为非负可积函数，下面的两个条件每个都是使得(2.3.1)式成立的充分必要条件(因而也是使得 $p_X(x)$ 是 X 的密度的充分必要条件).

1) 对任意实数 $a,b,a<b$ 都有

$$P(a < X \leqslant b) = \int_a^b p_X(x)\,\mathrm{d}x. \tag{2.3.2}$$

2) 对实直线的每个子集 D 都有

$$P(X \in D) = \int_D p_X(x)\,\mathrm{d}x.$$

注 2. 如果 f 只取某有限区间 $[a,b]$ 的值，令

$$\tilde{f}(x)=\begin{cases} f(x) & x\in[a,b], \\ 0 & \text{其它}. \end{cases}$$

则 $\tilde{f}$ 是定义在 $(-\infty,+\infty)$ 上的密度函数，且 $f(x)$ 和 $\tilde{f}(x)$ 给出相同的概率分布.

注 3. 假设有总共一个单位的质量连续地分布在 $a\leq x\leq b$ 上. 那么 $f(x)$ 表示在点 x 的质量密度且 $\int_c^d f(x)dx$ 表示在区间 $[c,d]$ 上的全部质量.

cdf and pdf and probability

注 4. $F(x)$ 表示的是随机变量的数值小于或等于 x 的概率，即

$$F(x) = P(X \leq x) \quad -\infty < x < +\infty. \tag{2.2}$$

由式 (2.2) 定义的 F 为 X 的 (累积) 分布函数的一般定义. 它适用于任意的随机变量. 设 X 为一离散型随机变量，它以概率 $\{p_1, ..., p_n, ..\}$ 取值 $\{a_1, ..., a_n, ...\}$. 则

$$F(x) = \sum_{a_i \leq x} p_i.$$

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} p(v)\,\mathrm{d}v.$$

cdf and pdf and probability

命题 2.3.2 设 X 是连续型随机变量，则 X 的分布函数连续且对任意实数 a，$P(X=a)=0$.

命题 2.3.3 设随机变量 X 的分布函数 $F(x)$ 是连续函数，并且除了有限个点外，$F(x)$ 的导数存在且连续.

均匀分布

设 $-\infty < a < b < +\infty$，如果随机变量 X 具有密度函数

$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b\,, \\ 0 & \text{其它}, \end{cases} \tag{2.8}$$

则称该随机变量为区间 $[a,b]$ 上的均匀分布, 记作 $U[a,b]$. 如此定义的 $f(x)$ 显然是一个概率密度函数. 容易算出其相应的分布函数为

$$F(x) = \begin{cases} 0, & x \le a, \\ \frac{x-a}{b-a}, & a < x \le b, \\ 1, & x > b. \end{cases}$$

$$p(x)=\frac{1}{b-a}I_{[a,b]}(x),\quad F(x)=\frac{x-a}{b-a}I_{[a,b)}(x)+I_{[b,+\infty)}(x).$$

中山大學 SUN YAT-SEN UNIVERSITY

例 2.3.2 设随机变量 X 服从[1,6]上的均匀分布，求关于 y 的二次方程

$$y^2+Xy+1=0$$

有实根的概率.

解 X 有密度 $(1/5)I_{[1,6]}(x)$，上面的方程有实根的充分必要条件是 $X^2-4\geqslant 0$，故方程有实根的概率是

$$\begin{aligned}P(X^2-4\geqslant 0)&=P\{X\in(-\infty,-2]\cup[2,+\infty)\}\\&=\int_{(-\infty,-2]\cup[2,+\infty)}(1/5)I_{[1,6]}(x)\,\mathrm{d}x\\&=\int_{-\infty}^{-2}(1/5)I_{[1,6]}(x)\,\mathrm{d}x+\int_{2}^{+\infty}(1/5)I_{[1,6]}(x)\,\mathrm{d}x\\&=\int_{2}^{6}(1/5)\,\mathrm{d}x=0.8.\end{aligned}$$

注意上式的第二步应用了定理 2.3.1.

指数分布

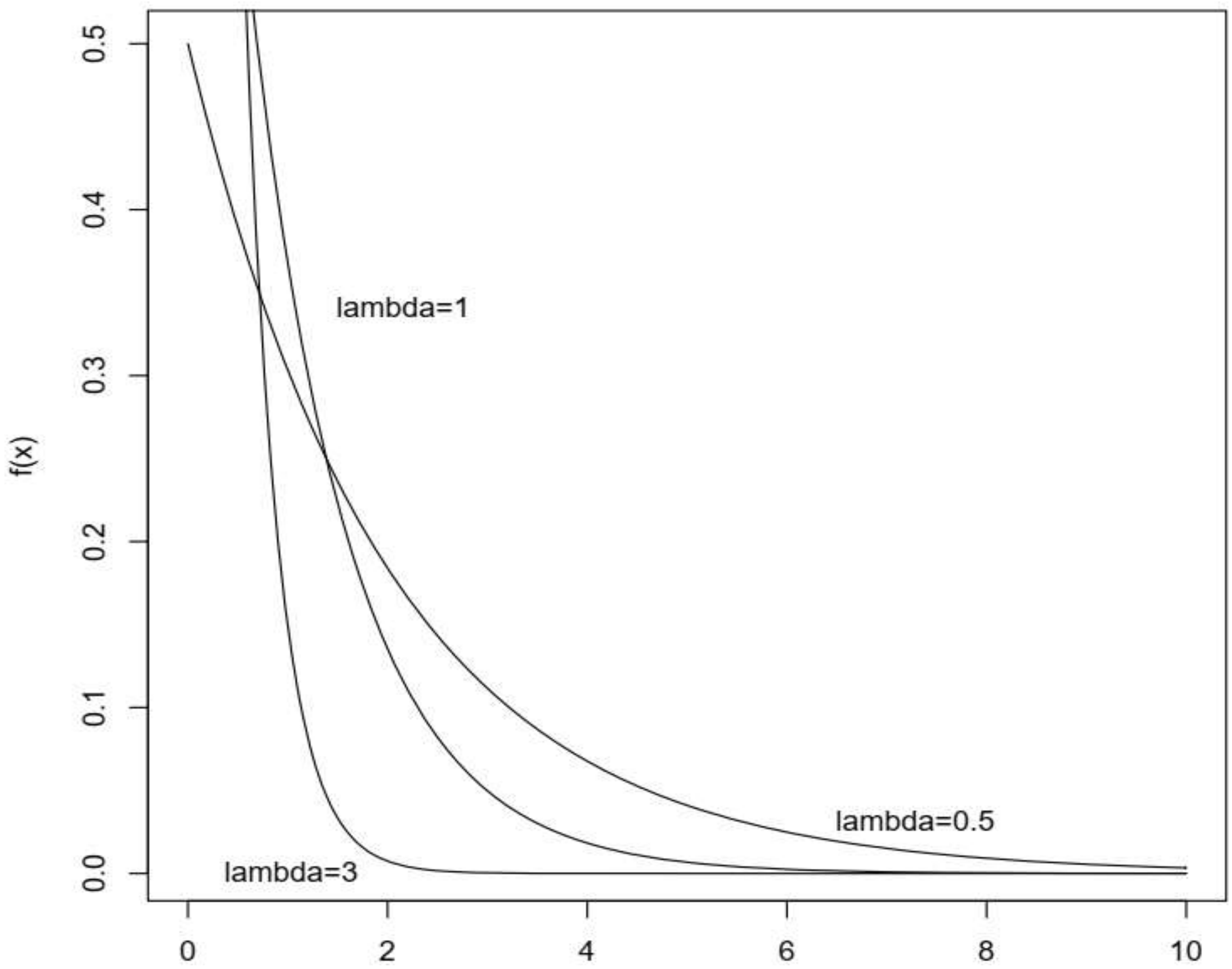
若随机变量 X 具有概率密度函数

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0,\\ 0 & x\le 0,\end{cases}$$

其中 $\lambda>0$ 为常数, 则称 X 服从参数为 λ 的指数分布.

指数分布的分布函数为

$$F(x)=\begin{cases}1-e^{-\lambda x} & x>0,\\ 0 & x\le 0.\end{cases}$$



lambda=1
lambda=0.5
lambda=3
f(x)
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
10

指数分布

指数分布经常用于作为各种”寿命”的分布的近似. 令 X 表示某元件的寿命. 我们引进 X 的失效率函数如下:

$$h(x) = \lim_{\Delta x \to 0} \frac{P(x \le X \le x + \Delta x | X > x)}{\Delta x}.$$

失效率表示了元件在时刻 x 尚能正常工作, 在时刻 x 以后, 单位时间内发生失效的概率. 则如果

$$h(x) \equiv \lambda \quad (\text{常数}), \quad 0 < x < +\infty,$$

X 服从指数分布. 即指数分布描述了无老化时的寿命分布.

设 X 表示某种电子元件的寿命，$F(x)$ 为其分布函数。若假设元件无老化，即元件在时刻 x 正常工作的条件下，其失效率保持为某个常数 λ，与 x 无关。试证明 X 服从指数分布。

比较课本例2.3.3

解：失效率即单位时间内失效的概率，因此由题设知

$$P(x \le X \le x+h|X>x)/h=\lambda, \quad h\to 0$$

因为

$$P(x \le X \le x+h|X>x)=\frac{P(\{x\le X\le x+h\}\{X>x\})}{P(X>x)}=\frac{F(x+h)-F(x)}{1-F(x)}$$

所以有

$$\lim_{h\to 0}P(x\le X\le x+h|X>x)/h=\frac{F'(x)}{1-F(x)}=\lambda$$

即得到微分方程 $\frac{F'(x)}{1-F(x)}=\lambda$,解此方程得到

$$F(x)=1-e^{-\lambda x}$$

指数分布

指数分布的最重要的特点是“无记忆性”. 即若 X 服从指数分布, 则对任意的 $s,t>0$ 有

$$P(X>s+t \mid X>s)=P(X>t).$$

可以证明指数分布是唯一具有上述性质的连续型分布

正态分布

如果一个随机变量 X 具有概率密度函数

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},\quad -\infty<x<+\infty \tag{2.3}$$

其中 $-\infty<\mu<+\infty$ $\sigma^2>0$，则称 X 为一正态随机变量，记为 $X\sim N(\mu,\sigma^2)$. 以 (2.3) 为密度的分布称为参数为 μ 和 σ^2 的正态分布.

具有参数 $\mu=0$，$\sigma=1$ 的正态分布称为标准正态分布. 用 $\Phi(x)$ 和 $\phi(x)$ 表示标准正态分布 $N(0,1)$ 的分布函数和密度函数.

正态分布

以 $F(x)$ 记正态分布 $N(\mu,\sigma^2)$ 的概率分布函数，则恒有 $F(x)=\Phi(\frac{x-\mu}{\sigma})$. 所以任一正态分布的概率分布函数都可通过标准正态分布的分布函数计算出来.

设 $X\sim N(\mu,\sigma^2)$, 则 $\frac{X-\mu}{\sigma}\sim N(0,1)$.

$$P\left(\frac{X-\mu}{\sigma}\le x\right)=P(X\le\sigma x+\mu)$$
$$=\int_{-\infty}^{\sigma x+\mu}\frac{1}{\sqrt{2\pi}\sigma}e^{-(t-\mu)^2/(2\sigma^2)}dt$$
$$\overset{t=\sigma z+\mu}{=}\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$$

正态分布

如果随机变量ξ服从正态分布$N(a,\sigma^2)$,则有

$$P(|\xi - a| < 3\sigma) = 0.9974\,,$$

我们把这个性质称为3σ原则.

$$\int_{-\infty}^{+\infty} e^{-x^2}\mathrm{d}x = \sqrt{\pi},$$

正态分布

例 2.3.4 已知 $X \sim N(0,1)$,求 $P(X \leqslant 2)$,$P(X \geqslant -1)$,$P(-2 < X < 1)$.

解 $P(X \leqslant 2) = \Phi(2) = 0.977\,3$,

$$P(X \geqslant -1) = 1 - P(X < -1) = 1 - P(X \leqslant -1)$$
$$= 1 - \Phi(-1) = \Phi(1) = 0.841\,3,$$

$$P(-2 < X < 1) = P(X < 1) - P(X \leqslant -2) = P(X \leqslant 1) - P(X \leqslant -2)$$
$$= \Phi(1) - \Phi(-2) = \Phi(1) + \Phi(2) - 1$$
$$= 0.841\,3 + 0.977\,3 - 1 = 0.818\,6.$$

Gamma 分布

如果随机变量 X 的密度函数为

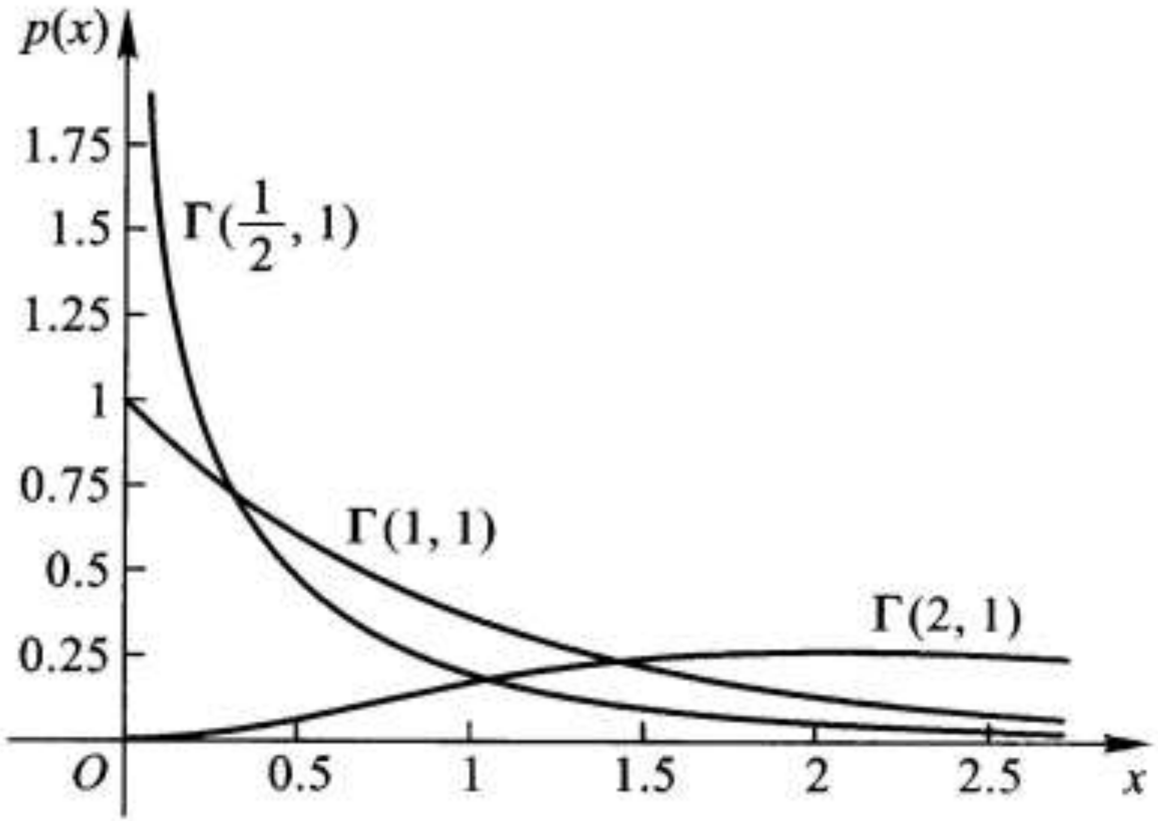
$$p(x)=\begin{cases}\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$$

其中 $\alpha,\beta>0$ 是参数，则称 X 服从 Γ 分布（或伽玛分布）$\Gamma(\alpha,\beta)$，记为 $X\sim\Gamma(\alpha,\beta)$. 其中

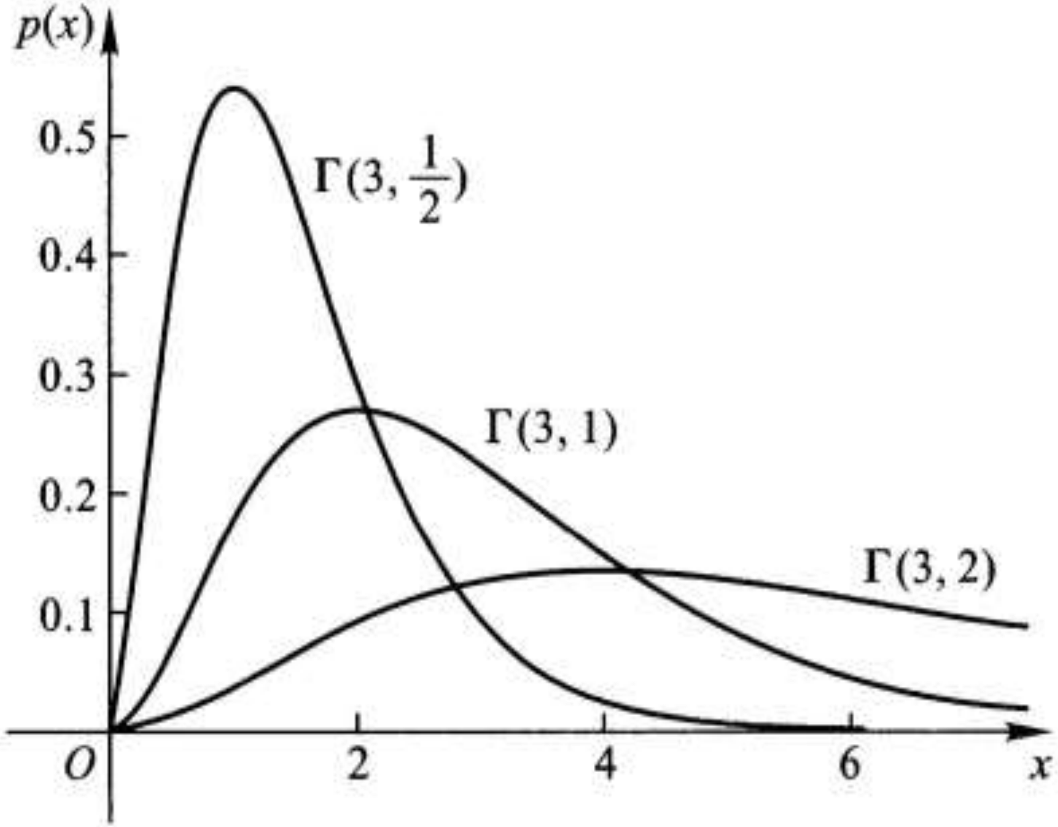
$$\Gamma(\alpha)=\int_0^{+\infty}t^{\alpha-1}e^{-t}\mathrm{d}t,\ \alpha>0$$

称为 Γ 函数（伽玛函数）.

Gamma 分布



$p(x)$
1.75
1.5
1.25
1
0.75
0.5
0.25
O
0.5
1
1.5
2
2.5
x
$\Gamma(\frac{1}{2}, 1)$
$\Gamma(1, 1)$
$\Gamma(2, 1)$



$p(x)$
0.5
0.4
0.3
0.2
0.1
O
2
4
6
x
$\Gamma(3, \frac{1}{2})$
$\Gamma(3, 1)$
$\Gamma(3, 2)$

注 2.3.5　Γ 函数有下列性质：

$$\Gamma(\alpha+1)=\alpha\Gamma(\alpha),\quad \Gamma(1)=1,\quad \Gamma(1/2)=\sqrt{\pi}.$$

因此，若 n 为自然数，则

$$\Gamma(n+1)=n!,\quad \Gamma(n+1/2)=(n-1/2)(n-3/2)\cdots(1/2)\sqrt{\pi}.$$

注 2.3.6　不难从 Γ 分布的定义和注 2.3.5 看出，$\Gamma(1,\beta)$ 分布就是参数为 β 的指数分布. 由定义 5.4.1 可知当 n 为正整数时，$\Gamma(n/2,1/2)$ 分布就是在统计中常用的 n 个自由度的 χ^2（卡方）分布 χ_n^2.

n个独立标准正态随机变量的和

Beta 分布

在贝叶斯统计中常常用到 β 分布. 如果随机变量 X 的密度函数为

$$p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{(0,1)}(x),$$

其中 $\alpha>0$ 和 $\beta>0$ 都是参数,则称 X 服从 β 分布(或贝塔分布).

Weibull 分布

在寿命分布中常常用到威布尔分布. 如果随机变量 X 的密度函数为

$$p(x)=\frac{\alpha}{\eta}(x/\eta)^{\alpha-1}\exp\{-(x/\eta)^{\alpha}\}I_{(0,+\infty)}(x),$$

其中 $\alpha>0$ 和 $\eta>0$ 都是参数,则称 X 服从**威布尔分布**.