《初等概率论》第 5 讲

邓 婉 璐

清华大学
统计学研究中心

October 8, 2018

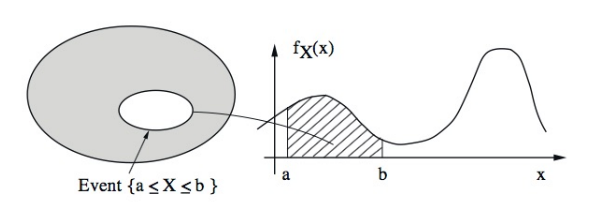
一、连续型随机变量

定义：连续型随机变量、概率密度

设随机变量 X，如果存在非负函数 $f(x)$ 使得对任意的 $a<b$，

$$\mathbb{P}(a<X\leq b)=\int_a^b f(x)\,dx,$$

则称 X 是连续型随机变量 (continuous r.v.)，称 $f(x)$ 是 X 的概率密度函数，简称概率密度或者密度 (density).



一、连续型随机变量

设 $f(x)$ 是 X 的概率密度，则

1. $\int_{-\infty}^{\infty} f(x)\,dx = 1.$

证明. 事件 $A_n = \{X \in (-n, n]\}$ 单调增，利用概率的连续性得

$$\int_{-\infty}^{\infty} f(x)\,dx = \lim_{n\to\infty} \int_{-n}^{n} f(x)\,dx = \lim_{n\to\infty} \mathbb{P}(A_n)$$
$$= \mathbb{P}\Big(\cup_{n=1}^{\infty} A_n\Big) = \mathbb{P}(X \in (-\infty, \infty)) = 1.$$

2. $\mathbb{P}(X = a) = 0$，于是 $\mathbb{P}(a < X \le b) = \mathbb{P}(a \le X \le b)$.

证明. 利用概率的连续性得

$$\mathbb{P}(X = a) \le \mathbb{P}(X \in (a-\varepsilon, a]) = \int_{a-\varepsilon}^{a} f(x)\,dx \to 0, \quad \text{as } \varepsilon \downarrow 0.$$

3. 对任意的 Borel 集 A，有 $\mathbb{P}(X \in A) = \int_A f(x)\,dx$.

一、连续型随机变量

♣ 以后可以证明，在 $\mathbb{R}$ 上的积分值等于 1 的非负函数一定是某个随机变量的概率密度函数.

♣ 虽然 X 是连续型随机变量，但是并不意味着 $X = X(\omega)$ 是连续函数，因为 X 的定义域 — 样本空间 Ω 没有任何拓扑结构，根本谈不上任何连续性，连续型随机变量的意思是它的分布函数绝对连续.

♣ 密度函数在其定义域内未必有界.

例 1.1

考虑函数

$$f(x) = \frac{1}{2\sqrt{x}}, \quad x \in (0, 1].$$

则 $f(x)$ 是密度函数，但 $f(x)$ 在 $(0, 1]$ 上无界，因为 $\lim_{x \downarrow 0} f(x) = +\infty$.

一、连续型随机变量

♣ 常见的连续型随机变量

1. 均匀分布 $\mathcal{U}(a,b)$

对 $a<b$，如果 X 的概率密度是

$$f(x)=\begin{cases}\frac{1}{b-a}, & x\in(a,b),\\ 0, & x\notin(a,b),\end{cases}$$

称 X 服从区间 (a,b) 上的均匀分布，记作 $X\sim\mathcal{U}(a,b)$.
$\mathcal{U}$ 是 Uniform 的首字母.

显然区间 (a,b) 也可以写成 $(a,b]$, $[a,b)$ 或 $[a,b]$. 密度函数 $f(x)$ 还可以写成示性函数的形式

$$f(x)=\frac{1}{b-a}I_{(a,b)}(x) \text{ 或 } f(x)=\frac{1}{b-a}I_{x\in(a,b)}.$$

一、连续型随机变量

对任意的 Borel 集 A, 如果 A 的测度 $m(A)=\int_A dx<\infty$, 可以类似地定义 A 上的均匀分布：如果 X 的密度是

$$f(x)=\begin{cases}\frac{1}{m(A)}, & x\in A,\\ 0, & x\in A^c,\end{cases}$$

称 X 服从 Borel 集 A 上的均匀分布，记作 $X\sim\mathcal{U}(A)$. 对任意的 Borel 集 B,

$$\mathbb{P}(B)=\frac{m(A\cap B)}{m(A)}.$$

一、连续型随机变量

2. 指数分布 $\mathcal{E}(\lambda)$

对正常数 λ，如果 X 的密度是

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

称 X 服从参数 λ 的指数分布，记作 $X \sim \mathcal{E}(\lambda)$. $\mathcal{E}$ 是 Exponential 的首字母.

密度函数 $f(x)$ 还可以写成示性函数的形式

$$f(x) = \lambda e^{-\lambda x} I_{[0,\infty)} \text{ 或 } f(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

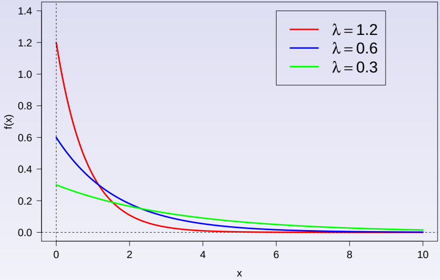
♣ 应用

- 到发生某个事件为止所用的时间
- 一台仪器的使用寿命
- ···

与几何分布十分相似，稍后会看到两者的比较.

一、连续型随机变量

♣ X 超过某个值的概率，随着这个值的增加而按指数递减，即 $P(X \geq a) = e^{-\lambda a}, \forall a \geq 0.$



图：指数分布密度函数图形.

♣ 当随机变量 X 使得 $\mathbb{P}(X < 0) = 0$, 称 X 是非负随机变量.

一、连续型随机变量

定理 1.1

设 X 是连续型非负随机变量，则 X 服从指数分布的充要条件是对任意的 $s,t\geq 0$，有

$$\mathbb{P}(X>s+t|X>s)=\mathbb{P}(X>t). \tag{1}$$

性质 (1) 称为无后效性或者无记忆性，是指数分布的特征.

证明. ($\Longrightarrow$). 设 $X\sim\mathcal{E}(\lambda)$，则

$$\mathbb{P}(X>x)=\int_x^{\infty}\lambda e^{-\lambda y}dy=e^{-\lambda x}$$

和条件概率公式得

$$\begin{aligned}\mathbb{P}(X>s+t|X>s)&=\frac{\mathbb{P}(X>s+t,X>s)}{\mathbb{P}(X>s)}=\frac{\mathbb{P}(X>s+t)}{\mathbb{P}(X>s)}\\&=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}=\mathbb{P}(X>t).\end{aligned}$$

一、连续型随机变量

$(\Longleftarrow)$. 设 X 具有性质 (1), 令

$$G(x) = \mathbb{P}(X > x).$$

由

$$\mathbb{P}(X > s + t | X > s) = \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} = \mathbb{P}(X > t).$$

得 $G(s + t) = G(s)G(t)$，即 $\log G(s + t) = \log G(s) + \log G(t)$. 由于 $G(x)$ 单调非增函数，所以是可测的. 由高等数学的知识知

$$\log G(x) = \{\log G(1)\}x, \quad \text{i.e.,} \quad G(x) = \mathrm{e}^{-\lambda x},$$

其中，$\lambda = -\log \mathbb{P}(X > 1) > 0$. 最后，对任意的 $0 \le a < b$,

$$\mathbb{P}(a < X \le b) = \mathbb{P}(X \le b) - \mathbb{P}(X \le a) = \mathrm{e}^{-\lambda a} - \mathrm{e}^{-\lambda b}$$
$$= \int_a^b \lambda \mathrm{e}^{-\lambda y} dy,$$

知 $X \sim \mathcal{E}(\lambda)$.

一、连续型随机变量

♣ 从技术上说，“无记忆性”就是说：元件在时刻 s 尚能正常工作的条件下，其失效率总保持为某个常数 $\lambda > 0$，与 s 无关. 失效率就是单位长度时间内失效的概率.

指数分布描述了元件无老化时的寿命分布，但“无老化”是不可能的，因而只是一种近似. 对一些寿命长的元件，在初期阶段老化现象很小. 在这一阶段，指数分布比较确切地描述了其寿命分布情况. 如人的寿命，一般在 50 岁或 60 岁之前，由于生理上老化而死亡的因素是次要的，若排除那些意外情况，人的寿命分布在此阶段应该接近指数分布. 若考虑老化，则应取失效率随时间而上升，不能为常数，而应取为 x 的增函数，比如 λx^m，对某个常数 $\lambda > 0, m > 0$.

3 正态分布 $\mathcal{N}(\mu, \sigma^2)$

如果 X 的密度是

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

称 X 服从参数 (μ, σ^2) 的正态分布，记作 $X \sim \mathcal{N}(\mu, \sigma^2)$. $\mathcal{N}$ 是 Normal 的首字母. 特别地，当 $X \sim \mathcal{N}(0, 1)$ 时，称 X 服从标准正态分布. 标准正态分布的密度函数:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right), \quad x \in R,$$

一、连续型随机变量

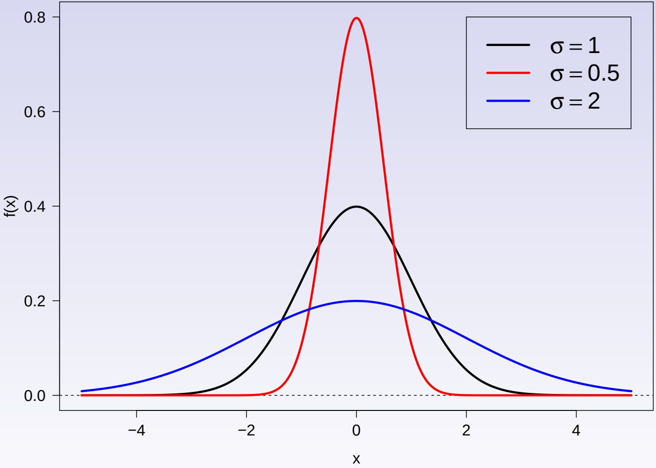
♣ 背景与应用

正态分布最早由 Gauss 在研究测量误差时得到，所以正态分布又称为 Gauss 分布. 在 Brown motion 的研究中，人们也得到了正态分布. 正态分布在概率论和数理统计中有着特殊的地位. 事实表明，产品的许多质量指标，生物和动物的许多生理指标等都服从或近似服从正态分布. 大量相互独立且具有相同分布的随机变量 (Independent and identically distributed random variables, i.i.d.) 的累积也近似服从正态分布.

正态分布的密度函数呈钟型，又称为钟型分布. 具有以下性质：

1. $f(x)$ 关于 $x = \mu$ 对称；
2. $f(\mu) = (\sigma\sqrt{2\pi})^{-1}$ 是最大值；
3. $f(x)$ 在 $x = \mu \pm \sigma$ 处有拐点.

一、连续型随机变量



图：正态分布密度函数图形.

一、连续型随机变量

♣ $\varphi(x)$ 确实是个密度函数.

$$\begin{aligned}
&\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-x^2/2}\,dx\right)\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-y^2/2}\,dy\right)\\
&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{-(x^2+y^2)/2}\,dxdy\\
&=\frac{1}{2\pi}\int_{0}^{2\pi}d\theta\int_{0}^{\infty}r\mathrm{e}^{-r^2/2}\,dr\\
&=\frac{1}{2}\int_{0}^{\infty}\mathrm{e}^{-r^2/2}\,dr^2\\
&=1.
\end{aligned}$$

♣ (2) 中的函数 $f(x)$ 是密度函数.

$$\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-x^2/2}\,dx=1.$$

4. Beta 分布 $\mathrm{B}(\alpha,\beta)$

设 α,β 是正常数，如果 X 的密度是

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

称 X 服从参数 (α,β) 的 Beta 分布，记作 $X\sim\mathrm{B}(\alpha,\beta)$, 其中

$$\mathrm{B}(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

♣ 当 $\alpha=\beta=1$ 时，为均匀分布 $\mathcal{U}(0,1)$;
当 $\alpha=2,\beta=1$ 时，密度函数呈直线型;
当 $\alpha=\beta=\frac{1}{2}$ 时，密度函数呈上 U 型;
当 $\alpha=\beta=2$ 时，密度函数呈下 U 型.

♣ 常在贝叶斯统计方法中作为参数的先验分布. 特别地，它是二项分布的共轭分布. $(X|p\sim Binomial(n,p)$. 参数 p 的先验 $\mathrm{B}(\alpha,\beta)$, 观测 X 后得到后验 $\mathrm{B}(\alpha+X,\beta+n-X)$; 直观解释; 学习条件分布后将可以证明)

一、连续型随机变量

5 Weibull 分布

如果 X 的密度是

$$f(x;\lambda,\alpha) = \lambda\alpha x^{\alpha-1}\exp\left(-\lambda x^{\alpha}\right), \quad x \geq 0,$$

称 X 服从 Weibull 分布.

(a) 当 $\alpha = 1$ 时，Weibull 分布为指数分布；

(b) 当 $\alpha = 2$ 时，Weibull 分布为 Rayleigh 分布，其密度为：

$$f(x;\sigma) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right), \quad x \geq 0,$$

♣ 指数分布和 Weibull 分布在可靠性分析中占重要地位.

一、连续型随机变量

6. Gamma 分布 $\Gamma(\alpha,\lambda)$

设 α,λ 是正常数，如果 X 的密度是

$$f(x)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x},\quad x\geq 0,$$

称 X 服从参数 (α,λ) 的 Gamma 分布，记作 $X\sim\Gamma(\alpha,\lambda)$, 其中

$$\Gamma(\alpha)=\int_0^{\infty}x^{\alpha-1}\mathrm{e}^{-x}dx.$$

称为 Gamma 函数.

♣ Gamma 函数的性质:
(a) $\Gamma(n)=(n-1)!$; (b) $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$; (c) $\Gamma(1/2)=\sqrt{\pi}$.

♣ α 被称为形状参数 (shape parameter);
λ 被称为 the rate parameter;

一、连续型随机变量

♣ 当 $\alpha=1$ 时，为指数分布.

♣ 当 α 为正整数时，称为 Erlang 分布. (用在“排队论”中).

♣ 英国著名统计学家 Karl Pearson 在研究物理、生物及经济中的随机变量时，发现很多连续型随机变量的分布不是正态分布. 这些随机变量的特点是只取非负值，于是他致力于这类随机变量的研究，参阅 Wikipedia 上的 “Pearson distribution”。在气象学中，干旱地区的年、季或月降水量被认为服从 Γ 分布.

7 χ_ν^2 分布

设 $\nu > 0$，如果 X 的密度是

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu}{2}-1} \mathrm{e}^{-\frac{x}{2}}, \quad x \geq 0,$$

称 X 服从参数 ν 的 χ_ν^2 分布，记作 $X \sim \chi_\nu^2$.

8. Student's t_ν 分布

设 $\nu > 0$，如果 X 的密度是

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in R,$$

称 X 服从参数 ν 的 Student 分布，记作 $X \sim t_\nu$.

♣ 对称性. 若 $T \sim t_\nu$，则 $-T \sim t_\nu$.

♣ 比正态分布重尾.

♣ 当 ν 很大时，t_ν 分布看起来和标准正态分布很像 (以后将用大数定律证明).

❾ Cauchy 分布

如果 X 的密度是

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

称 X 服从 Cauchy 分布。显然记作 $X \sim t_1$.

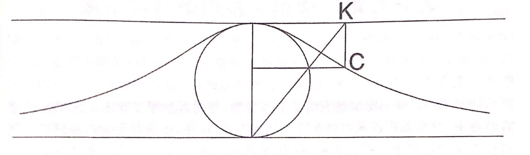


图: 当 K 从 $-\infty$ 变到 ∞ 时，点 C 的轨迹描绘出了 Maria Gaetana Agnesi(1718/5/16- 1799/1/9, 意大利女数学家兼哲学家) 箕舌线.

10 $F(m,n)$ 分布【Fisher Snedecor distribution】

设 m, n 是正整数，如果 X 的密度是

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2} x^{\frac{m}{2}-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)(mx+n)^{\frac{m+n}{2}}}, \quad x \geq 0,$$

称 X 服从参数为 (m,n) 的 F 分布，记作 $X \sim F(m,n)$.

二、概率分布函数

♣ 可否将连续变量、离散变量统一表达?

定义：分布函数 (distribution function)

对 r.v. X，称 x 的函数

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

为 X 的**概率分布函数**或**累积分布函数** (cumulative distribution function, CDF)，简称为**分布函数**.

二、概率分布函数

♣ 如果 X 是离散型随机变量，有概率分布

$$p_k = \mathbb{P}(X = x_k), \quad k = 1, 2, \ldots,$$

或

X	x_1	x_2	x_3	$\cdots$
$\mathbb{P}$	p_1	p_2	p_3	$\cdots$

则 X 的分布函数为

$$F(x) = \mathbb{P}(X \le x) = \mathbb{P}\Big(\bigcup_{j: x_j \le x} \{X = x_j\}\Big) = \sum_{j: x_j \le x} p_j.$$

二、概率分布函数

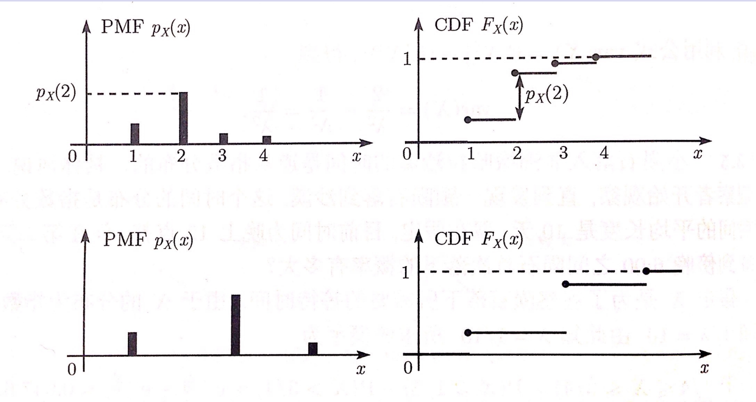


图: 某些离散随机变量的 CDF.

二、概率分布函数

注意到 $F(x)$ 是单调不减的阶梯函数，它在每个 x_j 处有跳跃 p_j. 此时，也称 $F(x)$ 是分布列 $\{p_j\}$ 的分布函数. 进一步，如果 $\{x_k\}$ 是单增的，则

$$p_k = \mathbb{P}(X \le x_k) - \mathbb{P}(X \le x_{k-1}) = F(x_k) - F(x_{k-1}).$$

♣ 令 $H(x)$ 是 Heaviside 函数：$H(x) = I_{[0,\infty)}(x)$. 则

$$F(x) = \sum_{k=1}^{\infty} p_k H(x - x_k). \tag{3}$$

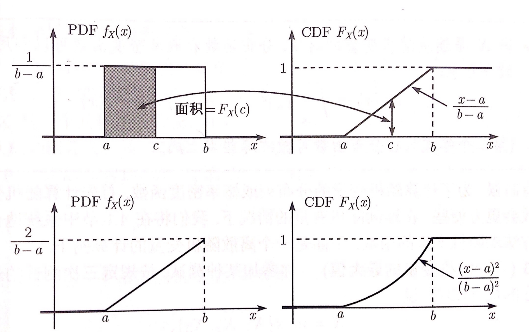
♣ 如果分布函数 $F(x)$ 可以表示为 (3) 这种形式，称 $F(x)$ 是离散的【$F(x)$ is called *discrete.*】.

二、概率分布函数

♣ 如果 X 是连续型随机变量，有概率密度 $f(x)$，则

$$F(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f(t)\,dt, \quad x \in R$$

是连续函数，并且在 $f(x)$ 的连续点 x 有 $f(x) = F'(x)$. 称 $F(x)$ 是 $f(x)$ 的分布函数.



图：某些连续型随机变量的 CDF.

二、概率分布函数

例：标准正态分布的分布函数：

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,dt.$$

利用 $\varphi(t)$ 的对称性，得

$$\Phi(x) + \Phi(-x) = 1, \quad \text{或} \quad \Phi(-x) = 1 - \Phi(x), \quad x \in R.$$

对一般的正态分布，$X \sim \mathcal{N}(\mu, \sigma^2)$，有

$$\begin{aligned}\mathbb{P}(X \le a) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(a-\mu)/\sigma} \mathrm{e}^{-y^2/2}\,dy \\ &= \Phi\left(\frac{a-\mu}{\sigma}\right).\end{aligned}$$

二、概率分布函数

当 $X \sim \mathcal{N}(\mu, \sigma^2)$ 时，

$$\mathbb{P}(|X-\mu| \le \sigma) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 68.27\%;$$
$$\mathbb{P}(|X-\mu| \le 2\sigma) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 95.45\%;$$
$$\mathbb{P}(|X-\mu| \le 3\sigma) = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 = 99.73\%;$$
$$\mathbb{P}(|X-\mu| \le 6\sigma) = 2\Phi(6) - 1 = 99.999999802682\%.$$

在产品质量或服务质量管理中，所谓的 3σ 或 6σ 管理原则就是要求产品质量的合格率、产品性能的可靠性或服务系统的满意度等达到 $\mathbb{P}(|X-\mu| \le 3\sigma)$ 或 $\mathbb{P}(|X-\mu| \le 6\sigma)$.

二、概率分布函数

♣ 分布函数对一切随机变量都适用。可以用来探讨离散和连续随机变量之间的关系。

例如，考虑几何随机变量与指数随机变量间的关系.

$X_1 \sim G(p)$, 即 $P(X_1 = k) = p(1-p)^{k-1}$. 则 X_1 的 CDF 为

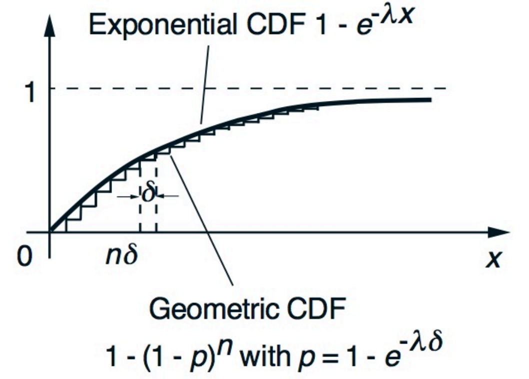
$$F_1(x) = \mathbb{P}(X_1 \le x) = \sum_{k=1}^{n} p(1-p)^{k-1}$$
$$= p\frac{1-(1-p)^n}{1-(1-p)} = 1-(1-p)^n, \quad n = 1, 2, \ldots$$

$X_2 \sim \mathcal{E}(\lambda)$, 则 X_2 的 CDF 为

$$F_2(x) = \mathbb{P}(X_2 \le x) = \int_{-\infty}^{x} \lambda e^{-\lambda t} dt$$
$$= -e^{-\lambda t}\big|_0^x = 1 - e^{-\lambda x}, \quad x > 0.$$

二、概率分布函数

比较两分布函数. 令 $\delta = -\frac{ln(1-p)}{\lambda}$, i.e. $e^{-\lambda\delta} = 1 - p$.
则对于 $n = 1, 2, \ldots$, 有 $F_1(n) = F_2(n\delta)$. 相当于，快速抛掷硬币，每 δ 秒抛一次，每次正面的概率为 p. 则 X_1 表示第一次得到正面所需抛掷次数，$X_1\delta$ 表示这个时刻. 可从下图的分布函数中看到 $X_1\delta \approx X_2$.



二、概率分布函数

♣ 分布函数对一切随机变量都适用。也可以用来研究其他类型的随机变量。例：抛均匀硬币，正面直接得钱 0.5 元，反面转一个幸运转盘，可得钱为 0 到 1（元）的均匀分布。

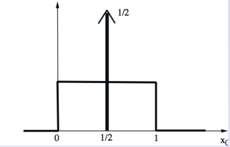


图: PMF 和 PDF 的示意图.

其分布函数：

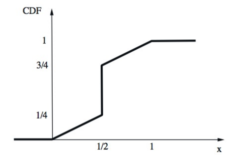


图: 某混合型随机变量的 CDF.

二、概率分布函数

♣ 分布函数 $F(x)$ 的性质

1. $F(x)$ 单调非降；
2. $F(-\infty) = \lim\limits_{x\to-\infty} F(x) = 0$, $F(\infty) = \lim\limits_{x\to\infty} F(x) = 1$.
3. $F(x)$ 是右连续的.

证明. (1). 对任意的 $x < y$, 有 $F(y) - F(x) = \mathbb{P}(x < X \leq y) \geq 0$, 即 $F(x) \leq F(y)$. 故结论 (1) 成立.

(2). 由概率的连续性，可得

$$F(\infty) = \lim_{n\to\infty} F(n) = \lim_{n\to\infty} \mathbb{P}(X \leq n)$$

$$= \mathbb{P}\left(\bigcup_{n=1}^{\infty}\{X \leq n\}\right) = \mathbb{P}(X < \infty) = 1.$$

同理可证 $F(-\infty) = 0$.

二、概率分布函数

(3). 由 $F(x)$ 的单调性和概率的连续性，可得

$$\begin{aligned}\lim_{n\to\infty} F(x+n^{-1}) &= \lim_{n\to\infty} \mathbb{P}(X\le x+n^{-1}) \\ &= \mathbb{P}\Big(\bigcap_{n=1}^{\infty}\{X\le x+n^{-1}\}\Big) \\ &= \mathbb{P}(X\le x) = F(x).\end{aligned}$$

例 2.1

一个使用了 t 小时的热敏电阻在 $\triangle t$ 内失效的概率是 $\lambda\triangle t + o(\triangle t)$. 设该热敏电阻的使用寿命是连续型随机变量，求该热敏电阻的使用寿命的分布.

解. 用 X 表示该热敏电阻的使用寿命，要求的是 $F(x) = \mathbb{P}(X \leq x)$. 由题意得

$$\mathbb{P}(t < X \leq t + \triangle t | X > t) = \frac{\mathbb{P}(t < X \leq t + \triangle t)}{\mathbb{P}(X > t)}$$
$$= \lambda\triangle t + o(\triangle t),$$

即
$$\frac{F(t+\triangle t) - F(t)}{\bar{F}(t)} = \lambda\triangle t + o(\triangle t), \quad t \geq 0,$$

其中 $\bar{F}(x) = 1 - F(x)$ 被称为 X 的生存函数. 因此

$$\frac{F'_+(t)}{\bar{F}(t)} = \lambda, \quad t \geq 0.$$

二、概率分布函数

完全类似地，对 $t>0$，当 $s=t-\triangle t>0$ 时，有

$$\frac{F'_-(t)}{\overline{F}(t)}=\lambda,\quad t\geq 0.$$

故

$$\frac{F'(t)}{\overline{F}(t)}=\lambda,\quad t\geq 0.$$

即 $d\log\overline{F}(t)=-\lambda dt$, 积分后可得 $\overline{F}(t)=c\mathrm{e}^{-\lambda t}$, 其中 c 是一常数. 注意到 $\overline{F}(0)=c=1$, 所以 $F(t)=1-\mathrm{e}^{-\lambda t}$, 故 $X\sim\mathcal{E}(\lambda)$.

二、概率分布函数

例 2.2

如果对 $\forall x \in R$ 都有 $\mathbb{P}(X=x)=0$，则 $F(x)=\mathbb{P}(X \le x)$ 是连续的.

证明. 由分布函数的右连续性和

$$
\begin{aligned}
0 = \mathbb{P}(X=x) &= \mathbb{P}\Big(\bigcap_{m=1}^{\infty}\{x-\tfrac{1}{m} < X \le x\}\Big) \\
&= \lim_{m\to\infty}\{F(x)-F(x-m^{-1})\} \\
&= F(x)-F(x-),
\end{aligned}
$$

即 $F(x)$ 是左连续的. 又因为 $F(x)$ 是右连续的，因此 $F(x)$ 是连续的.

二、概率分布函数

定理 2.1

设 X 的分布函数 $F(x)$ 连续，数集 A 中任何两点之间的距离大于某个正数 δ. 如果在 A^c 上，导数 $F'(x)$ 存在且连续，则

$$f(x) = \begin{cases} F'(x), & \text{当 } x \notin A, \\ 0, & \text{当 } x \in A. \end{cases}$$

是 X 的密度函数.

证明. 对任何 $a < b$，不妨设 (a, b) 中只有集合 A 中的 $a_1, a_2, \ldots, a_k$，并且 $a_0 = a < a_1 < a_2 < \cdots < a_k < b = a_{k+1}$. 此时，

$$P(a < X \le b) = \sum_{j=0}^{k} \mathbb{P}(a_j < X \le a_{j+1}) = \sum_{j=0}^{k} [F(a_{j+1}) - F(a_j)]$$

二、概率分布函数

$$= \sum_{j=0}^{k} \int_{a_j}^{a_{j+1}} f(t)\,dt = \int_a^b f(t)\,dt.$$

故 $f(x)$ 是 X 的概率密度.

注：在该定理中，F 连续的条件是至关重要的. 尽管 F 连续不能保证 X 是连续型随机变量，但是 F 不连续能保证 X 不是连续型随机变量. 当 F 是二项分布的随机变量的分布函数时，除去有限个点外，$F'(x) = 0$，所以 X 的密度不存在.

二、概率分布函数

(选修) 请注意：$F(x)$ 连续，并不能保证 X 是连续型随机变量. 下面是一个反例.

♣ A function F is called ***singular*** *if and only if* it is continuous, not identically zero, F' exists a.e., and $F'(t) = 0$ a.e.

二、概率分布函数

♣ Uniform distribution on the Cantor set

The Cantor set C is defined by removing $(1/3, 2/3)$ from $[0, 1]$ and then removing the middle third of each interval that remains. We define an associated d.f. by setting

$F(x) = 0$ for $x \le 0$,

$F(x) = 1$ for $x \ge 1$,

$F(x) = 1/2$ for $x \in [1/3, 2/3]$,

$F(x) = 1/4 = 1/2^2$ for $x \in [1/9, 2/9] = [1/3^2, 2/3^2]$,

$F(x) = 3/4 = 1 - 1/2^2$ for $x \in [7/9, 8/9]$
$= [(3^2 - 2)/3^2, (3^2 - 1)/3^2]$,

....

二、概率分布函数

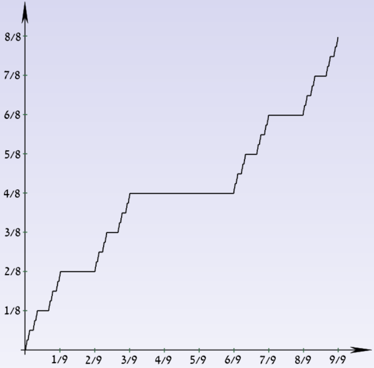


图: Cantor 分布函数

二、概率分布函数

It can be shown that the resulting function F is defined for all $x \in R$. Further it is nondecreasing, continuous, and $F(-\infty) = 0$ and $F(\infty) = 1$, i.e., it is a d.f. (Check carefully about the continuity of F.) Also, it is easy to see that $F'(x) = 0$ for all $x \in R$ except perhaps for those on the Cantor set (i.e., those points $m/3^n$, $m, n \in \mathbb{N}$). By definition, F is clearly a singular d.f., which is called “Lebesgue’s singular function”.

小结

知识点

- 离散型随机变量、分布列
- 连续型随机变量、密度函数
- 分布函数（统一性）、分布函数与分布列/密度函数的关系

技巧

- 利用情景建立不同分布之间的关系
- 利用分布间的关系理解分布的含义
- 分布函数与分布列/密度函数之间的转换运算

作业

作业：第三章：5，8，11

$$Thank\quad you!$$