

第二章 随机变量及概率分布

2.1 一维随机变量

2.1.1 随机变量的概念

顾名思义,随机变量就是“其值随机会而定”的变量,正如随机事件是“其发生与否随机会而定”的事件.机会表现为试验结果,一个随机试验有许多可能的结果,到底出现哪一个要看机会,即有一定的概率.最简单的例子莫如掷骰子,掷出的点数 X 是一个随机变量,它可以取 $1, \dots, 6$ 等 6 个值.到底是哪一个,要等掷了骰子以后才知道.因此又可以说,随机变量就是试验结果的函数.从这一点看,它与通常的函数概念又没有什么不同.把握这个概念的关键之点在于试验前后之分:在试验前,我们不能预知它将取何值,这要凭机会,“随机”的意思就在这里,一旦试验后,取值就确定了.比如你在 3 月 31 日买了一张奖券,到 6 月 30 开奖.当你买下这张奖券的后我就对你说:你中奖的金额 X 是一个随机变量,其值要到 6 月 30 日“抽奖试验”做过以后才能知道.

明白了这一点就不难举出一大堆随机变量的例子.比如,你在某厂大批产品中随机地抽出 100 个,其中所含废品数 X ;一月内某交通路口的事故数 X ;用天平秤量某物体的重量的误差 X ;随意在市场上买来一架电视机,其使用寿命 X 等等,都是随机变量.

随机变量的反面是所谓“确定性变量”,即其取值遵循某种严格的规律的变量.例如你以每小时 a 公里的匀速从某处向东行,则经 t 小时后,你距该处 at 公里.这一点我不待你做完这个试验(即走了 t 小时后)就能准确预知.在这种理想的条件下,你与该处的距离 X 并非随机变量.然而,你的速度必然会受到许多因素,包括随机性因素的影响,而成为不能预知的,这使你在 t 时间内行走

的距离 X 成为随机变量. 从绝对的意义讲, 许多通常视为确定性变量的量, 本质上都有随机性, 只是由于随机性干扰不大, 以至在所要求的精度之内, 不妨把它作为确定性变量来处理.

再考虑一个打靶的试验. 在靶面上取定一个直角坐标系 Oxy , 则命中的位置由其坐标 (X, Y) 来刻画, X, Y 都是随机变量, 而 (X, Y) 则称为一个二维随机向量或二维随机变量, 多维随机向量 (X_1, \dots, X_n) 的意义据此推广. 前面几个例子中的 X 都是一维随机变量, 通常就简称随机变量.

关于随机变量(及向量)的研究, 是概率论的中心内容. 这是因为, 对于一个随机试验, 我们所关心的往往是与所研究的特定问题有关的某个或某些量, 而这些量就是随机变量. 当然, 有时我们所关心的是某个或某些特定的随机事件. 例如, 在特定一群人中, 年收入十万元以上的高收入者, 及年收入在 8000 元以下的低收入者, 各自的比率如何, 这看上去像是两个孤立的事件. 可是, 若我们引进一个随机变量的 X :

$X =$ 随机抽出一个人其年收入

则 X 是我们关心的随机变量. 上述两个事件可分别表为 $\{X > 10000\}$ 和 $\{X < 3000\}$. 这就看出: 随机事件这个概念实际上是包容在随机变量这个更广的概念之内. 也可以说: 随机事件是从静态的观点来研究随机现象, 而随机变量则是一种动态的观点, 一如数学分析中的常量与变量的区分那样. 变量概念是高等数学有别于初等数学的基础概念. 同样, 概率论能从计算一些孤立事件的概念发展为一个更高的理论体系, 其基础概念是随机变量.

随机变量按其可能取的值的全体的性质, 区分为两大类.

一类叫离散型随机变量. 其特征是只能取有限个值, 或虽则在理论上讲能取无限个值, 但这些值可以毫无遗漏地一个接一个排列出来. 前者的例子如掷骰子的点数 X (6 个可能值), 从大批产品中抽出 100 个其中的废品数 X (101 个可能值); 后者的例子如一月内某交通路口的车祸数, 它理论上讲可以取 $0, 1, 2, \dots$ 等任一非负整数为值. 从实用的观点说, 这变量也只能取有限个值. 例如, 可

肯定它不会超过 10^{10} . 但由于不像前两例那样有一个明确的界线, 不如把它视为能取无穷个值, 理论上倒反简便些.

另一类叫连续型随机变量. 这种变量的全部可能取值不仅是无穷多的, 并且还不能无遗漏地逐一排列, 而是充满一个区间. 例如秤量一物体重量的误差, 由于我们难于明确指出误差的可能范围, 不妨就把它取为 $(-\infty, \infty)$ 更方便. 又如电视机的寿命, 其范围可取为 $(0, \infty)$, 也是一种抽象.

说到底, “连续型变量”这个概念只是一个数学上的抽象. 任何量都有一定单位, 都只能在该单位下量到一定的精度, 故必然为离散的. 但是当单位极小时, 其可能值在一范围内会很密集, 不如视为连续量在数学上更易处理. 其次, 关于连续型随机变量这个概念还需补充其一个重要方面, 这留到本节 2.1.3 段再谈.

2.1.2 离散型随机变量的分布及重要例子

研究一个随机变量, 不只是要看它能取哪些值, 更重要的是它取各种值的概率如何. 例如从一大批产品中随机抽出 100 个其中所含废品数 X . 当废品率小时, X 取 $0, 1, \dots$ 等小值的概率大. 反之, 若废品率很高, 则 X 取大值的概率就上升.

定义 1.1 设 X 为离散型随机变量, 其全部可能值为 $\{a_1, a_2, \dots\}$. 则

$$p_i = P(X = a_i), i = 1, 2, \dots \quad (1.1)$$

称为 X 的概率函数.

显然有

$$p_i \geq 0, p_1 + p_2 + \dots = 1 \quad (1.2)$$

后一式是根据加法定理, 因为事件 $\{X = a_1, \text{ 或 } a_2, \dots\}$ 为必然事件, 而又可表为一些互斥事件 $\{X = a_1\}, \{X = a_2\}, \dots$ 之和.

因此, 概率函数(1.1)给出了: 全部概率 1 是如何在其可能值之间分配的. 或者说, 它指出了概率 1 在其可能值集 $\{a_1, a_2, \dots\}$ 上的分布情况. 有鉴于此, 常把(1.1)称为随机变量 X 的“概率分布”. 它可以列表的形式给出:

可 能 值	a_1	a_2	\cdots	a_i	\cdots	(1.3)
概 率	p_1	p_2	\cdots	p_i	\cdots	

有时也把(1.3)称为 X 的分布表. 它也可以形象地用图 2.1 表出. 图中横轴上标出可能值之坐标 a_i , 而在 a_i 处的竖线之长则表示事件 $\{X = a_i\}$ 的概率.

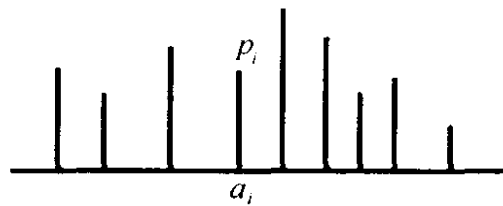


图 2.1

例如, 掷两粒均匀骰子, 以 X 记出现点数之和, 则 X 取 2, 3, \cdots , 12 等共 11 个可能值. 要确定其概率分布, 只好对上述每个 i 去计算 $P(X = i)$, 例如 $i = 6$. 投两个骰子可出现 36 种不同的但等可能的组合, 其中有利于事件 $\{X = 6\}$ 的组合有 5 种, 即 (1, 5), (5, 1), (2, 4), (4, 2), (3, 3). 故 $p_6 = P(X = 6) = 5/36$. 类似地算出其他 p_i . X 的分布表为

可 能 值	2	3	4	5	6	7	8	9	10	11	12	(1.4)
概 率	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	

对离散型变量, 用概率函数去表达其概率分布是最方便的. 也可以用下面定义的分函数:

定义 1.2 设 X 为一随机变量, 则函数

$$P(X \leq x) = F(x), \quad -\infty < x < \infty \quad (1.5)$$

称为 X 的分布函数. 注意这里并未限定 X 为离散型的: 它对于任何随机变量都有定义. 对离散型随机变量而言, 概率函数与分布函数在下述意义上是等价的, 即知道其一即可决定另一个. 事实上, 若知道概率函数(1.1), 则

$$F(x) = P(X \leq x) = \sum_{\{i: a_i \leq x\}} p_i$$

这个和号的意思,是指求和只对满足条件 $a_i \leq x$ 的那些 i 去进行. 如对上例而言,由分布表(1.4)算出

$$F(-1) = 0, F(2.5) = 1/36,$$

$$F(5) = (1 + 2 + 3 + 4)/36 = 5/18$$

等等. 反过来,由分布函数也易决定分布表. 仍以此例来说,如知道了 X 的分布函数 $F(x)$, 则为算 $p_i = P(X = i), i = 2, 3, \dots, 11$, 只须注意

$$\{X \leq i\} = \{X \leq i - 1\} + \{X = i\}$$

且右边两事件互斥. 于是

$$\begin{aligned} F(i) = P(X \leq i) &= P(X \leq i - 1) + P(X = i) \\ &= F(i - 1) + P(X = i) \end{aligned}$$

因而 $p_i = P(X = i) = F(i) - F(i - 1)$.

对任何随机变量 X , 其分布函数 $F(x)$ 具有下面的一般性质:

1° $F(x)$ 是单调非降的: 当 $(x_1 < x_2)$ 时, 有 $F(x_1) \leq F(x_2)$.

这是因为当 $x_1 < x_2$ 时, 事件 $\{X \leq x_1\}$ 蕴含事件 $\{X \leq x_2\}$, 因而前者的概率不能超过后者的概率.

2° 当 $x \rightarrow \infty$ 时, $F(x) \rightarrow 1$; 当 $x \rightarrow -\infty$ 时, $F(x) \rightarrow 0$.

这是因为, 当 $x \rightarrow \infty$ 时, $\{X \leq x\}$ 愈来愈接近于必然事件, 故其概率, 即 $F(x)$, 应趋于必然事件的概率, 即 1. 类似地得出后一论断.

下面来讨论几个在应用上常见的离散型随机变量的例子.

例 1.1 设某事件 A 在一次试验中发生的概率为 p . 现把这试验独立地重复 n 次, 以 X 记 A 在这 n 次试验中发生的次数. 则 X 可取 $0, 1, \dots, n$ 等值. 为确定其概率分布, 考虑事件 $\{X = i\}$. 要这个事件发生, 必须在这 n 次试验的原始记录

$$AA\bar{A}\bar{A}\cdots\bar{A}\bar{A}\bar{A}$$

中, 有 i 个 A , $n - i$ 个 \bar{A} . 每个 A 有概率 p 而每个 \bar{A} 有概率 $1 - p$. 又“ n 次试验独立”表示在每次中 A 出现与否与其他次试验的结

果独立. 因此概率乘法定理给出: 每个这样的原始结果序列发生的概率, 为 $p^i(1-p)^{n-i}$. 又因为在 n 个位置中 A 可以占据任何 i 个位置, 故一共有 $\binom{n}{i}$ 种可能. 由此得出

$$p_i = b(i; n, p) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, \dots, n \quad (1.6)$$

X 所遵从的概率分布 (1.6) 称为二项分布, 并常记为 $B(n, p)$. 以后, 当随机变量 X 服从某种分布 F 时, 我们用 $X \sim F$ 来表达这一点. 例如, X 服从二项分布就记为 $X \sim B(n, p)$.

二项分布是最重要的离散型概率分布之一. 上面已指出: 变量 X 服从这个分布有两个重要条件: 一是各次试验的条件是稳定的, 这保证了事件 A 的概率 p 在各次试验中保持不变; 二是各次试验的独立性. 现实生活中有许多现象程度不同地符合这些条件, 而不一定分厘不差. 例如, 某厂每天生产 n 个产品, 若原材料质量、机器设备、工人操作水平等在一段时期内大体保持稳定, 且每件产品之合格与否与其他产品合格与否并无显著关联, 则每日的废品数 X 大体上服从二项分布. 又如一大批产品 N 个, 其废品率为 p . 从其中逐一抽取产品检验其是否废品, 共抽 n 个. 若每次抽出检验后又放回且保证了每次抽取时, 每个产品有同等的 $1/N$ 的机会被抽出, 则这 n 个产品中所含废品数 X 就相当理想地遵从二项分布 $B(n, p)$ 了. 反之, 如果每抽出一个检验后即不放回去, 则下一次抽取时, 废品率已起了变化, 这时 X 就不再服从二项分布了. 但是, 若 N 远大于 n , 则即使不放回, 对废品率影响也极小. 这时, X 仍可近似地作为二项分布来处理.

例 1.2 波哇松分布. 若随机变量 X 的可能取值为 $0, 1, 2, \dots$, 且概率分布为

$$P(X = i) = e^{-\lambda} \lambda^i / i! \quad (1.7)$$

则称 X 服从波哇松分布, 常记为 $X \sim P(\lambda)$. 此处 $\lambda > 0$ 是某一常数. (1.7) 右边对 $i = 0, 1, \dots$ 求和的结果为 1, 可以从熟知的公式 e^λ

$$= \sum_{i=0}^{\infty} \lambda^i / i! \text{ 得出.}$$

这个分布也是最重要的离散型分布之一,它多是出现在当 X 表示在一定的时间或空间内出现的事件个数这种场合.前面提到的在一定时间内某交通路口所发生的事故个数,是一个典型的例子.这分布产生的机制也可以通过这个例子来解释.为方便计,设所观察的这段时间为 $[0, 1)$. 取一个很大的自然数 n , 把时间段 $[0, 1)$ 分为等长的 n 段:

$$l_1 = \left[0, \frac{1}{n}\right), l_2 = \left[\frac{1}{n}, \frac{2}{n}\right), \dots, l_i = \left[\frac{i-1}{n}, \frac{i}{n}\right), \dots, \\ l_n = \left[\frac{n-1}{n}, 1\right)$$

作几个假定:

1° 在每段 l_i 内,恰发生一个事故的概率,近似地与这段时间之长 $\frac{1}{n}$ 成正比,即可取为 λ/n . 又假定在 n 很大因而 $1/n$ 很小时,在 l_i 这么短暂的一段时间内,要发生两次或更多的事故是不可能的.因此,在 l_i 时段内不发生事故的概率为 $1 - 1/n$.

2° l_1, \dots, l_n 各段是否发生事故是独立的.

把在 $[0, 1)$ 时段内发生的事故数 X 视作在 n 个小时段 l_1, \dots, l_n 内有事故的时段数,则按上述 1°, 2° 两条假定, X 应服从二项分布 $B(n, \lambda/n)$. 于是

$$P(X = i) = \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \quad (1.8)$$

严格讲, (1.8) 只是近似成立而非严格等式. 因为在假定 1° 中, 在每时段内发生一次事故的概率只是近似地为 λ/n . 当 $n \rightarrow \infty$ 取极限时, 就得到确切的答案. 注意当 $n \rightarrow \infty$ 时

$$\binom{n}{i} / n^i \rightarrow 1/i!, \quad \left(1 - \frac{\lambda}{n}\right)^n \rightarrow e^{-\lambda}$$

得知 (1.8) 式右边以 $e^{-\lambda} \lambda^i / i!$ 为极限. 由此得出 (1.7).

从上述推导看出: 波哇松分布可作为二项分布的极限而得到.

一般地说,若 $X \sim B(n, p)$, 其中 n 很大, p 很小而 $np = \lambda$ 不太大时, 则 X 的分布接近于波哇松分布 $P(\lambda)$. 这个事实在所述条件下可将较难计算的二项分布转化为波哇松分布去计算, 看一个例子.

例 1.3 现在需要 100 个符合规格的元件. 从市场上买的该元件有废品率 0.01. 故如只买 100 个, 则它们全都符合规格的机会恐怕不大, 为此, 我们买 $100 + a$ 个. a 这样取, 以使“在这 $100 + a$ 个元件中至少有 100 个符合规格”这事件 A 的概率不小于 0.95. 问 a 至少要多大?

在此, 我们自然假定各元件是否合格是独立的. 以 X 记在这 $100 + a$ 个元件中所含废品数, 则 X 有二项分布 $B(100 + a, 0.01)$. 事件 A 即事件 $\{X \leq a\}$, 于是 A 的概率为

$$P(A) = \sum_{i=0}^a P(X = i) = \sum_{i=0}^a \binom{100+a}{i} (0.01)^i (0.99)^{100+a-i} \quad (1.9)$$

为确定最小的 a 使 $P(A) \geq 0.95$, 我们得从 $a = 0$ 开始, 对 $a = 0, 1, 2, \dots$ 依次计算 (1.9) 式右边之值, 直到算出 ≥ 0.95 的结果为止. 这很麻烦.

由于 $100 + a$ 这个数较大而 0.01 很小, $(100 + a)(0.01) = 1 + a(0.01)$ 大小中, 可近似地用波哇松分布计算. 由于平均在 100 个中只有一个废品, a 谅必相当小. 故可以用 1 近似地取代 $1 + a(0.01)$. 由此, X 近似地服从波哇松分布 $P(1)$, 因而

$$P(X \leq a) \approx \sum_{i=0}^a e^{-1}/i!$$

计算出当 $a = 0, 1, 2, 3$ 时, 上式右边分别为 0.368, 0.736, 0.920 和 0.981. 故取 $a = 3$ 已够了.

除了二项和波哇松这两个最重要的离散型分布外, 还有几个离散型分布, 其重要性略次一些, 但也很常用. 其中有超几何分布与负二项分布.

例 1.4 考虑第一章的例 2.1, 以 X 记从 N 个产品中随机抽出 n 个里面含废品数. 按该例的计算, X 的分布为 (第一章 (2.7))

式):

$$P(X = m) = \binom{M}{m} \binom{N - M}{n - m} / \binom{N}{n} \quad (1.10)$$

至于 m 的取值范围, 必须 $0 \leq m \leq M$ 及 $n - m \leq N - M$. 例如, $N = 500, n = 50, M = 25$, 则 m 的范围为 $0 \leq m \leq 25$. (1.10) 称为超几何分布, 是因为其形式与“超几何函数”的级数展式的系数有关.

这个分布在涉及抽样的问题中常用, 特别当 N 不大时. 因为通常在抽样时, 多是像在本例中这样“无放回的”, 即已抽出的个体不再有放回去以供再次抽出的机会, 这就与把 n 个同时抽出的效果一样. 如果一个一个地抽而抽出过的仍放回, 则如在例 1.1 中已指出的, 结果是二项分布. 在例 1.1 中也曾指出: 若 n/N 很小, 则放回与不放回差别不大. 由此可见, 在这种情况下超几何分布应与二项分布很接近. 确切地说, 若 X 服从超几何分布 (1.10), 则当 n 固定, $M/N = p$ 固定, $N \rightarrow \infty$ 时, X 近似地服从二项分布 $B(n, p)$.

例 1.5 为了检查某厂产品的废品率 p 大小, 有两个试验方案可采取: 一是从该厂产品中抽出若干个, 检查其中的废品数 X , 这一方案导致二项分布, 已于前述. 另一个方案是先指定一个自然数 r . 一个一个地从该厂产品中抽样检查, 直到发现第 r 个废品为止. 以 X 记到当时为止已检出的合格品个数. 显然, 若废品率 p 小, 则 X 倾向于取较大之值, 反之当 p 大时, 则 X 倾向于取小值. 故 X 可用于考究 p 的目的.

为计算 X 的分布, 假定各次抽取的结果(是废品或否)是独立的, 且每次抽得废品的概率, 保持固定为 p . 考察 $\{X = i\}$ 这个事件, 为使这个事件发生, 需要以下两个事件同时发生: ①在前 $i + r - 1$ 次抽取中, 恰有 $r - 1$ 个废品. ②第 $i + r$ 次抽出废品. 按所作假定, 这两事件的概率分别为 $b(r - 1; i + r - 1, p)$ 和 p . 再由独立性, 即得

$$P(X = i) = b(r - 1; i + r - 1, p)p$$

$$= \binom{i+r-1}{r-1} p^r (1-p)^i, \quad (1.11)$$

$$(i = 0, 1, 2, \dots)$$

这个分布称为负二项分布. 这名称的来由, 一则由于“负指数二项展开式”

$$(1-x)^{-r} = \sum_{i=0}^{\infty} \binom{-r}{i} (-x)^i = \sum_{i=0}^{\infty} \binom{i+r-1}{i} x^i$$

$$= \sum_{i=0}^{\infty} \binom{i+r-1}{r-1} x^i$$

中令 $x=1-p$ 并两边乘以 p^r , 得

$$1 = p^r [1 - (1-p)]^{-r} = \sum_{i=0}^{\infty} \binom{i+r-1}{r-1} p^r (1-p)^i$$

(这验证了分布(1.11)确满足(1.2)). 另一则由于例中所描述的试验方式, 它与二项分布比是“反其道而行之”: 二项分布是定下总抽样个数 n 而把废品个数 X 作为变量; 负二项分布则相反, 它定下废品个数 r 而把总抽样次数减去 r 作为变量.

一个重要的特例是 $r=1$. 这时, 注意到 $\binom{i}{0} = 1$ 之约定, (1.11) 成为

$$P(X=i) = p(1-p)^i, \quad i=0, 1, 2, \dots \quad (1.12)$$

概率 $p, p(1-p), p(1-p)^2, \dots$ 呈公比作为 $1-p$ 的几何级数, 故分布(1.12)又常称为几何分布.

2.1.3 连续型随机变量的分布及重要例子

连续型随机变量的意义已在 2.1.1 段中解释过. 对这种变量的概率分布, 不能用像离散型变量那种方法去描述. 原因在于, 这种变量的取值充满一个区间, 无法一一排出. 若指定一个值 a , 则变量 X 恰好是 a 一丝不差, 事实上不可能. 如在秤量误差的例中, 如果你认定天平上的读数(刻度)是“无限精细”, 则“误差正好为 $\pi-3$ ”虽原则上不能排除, 但可能性也极微以至只能取为 0. 如在

靶面上指定一个几何意义下的点(即只有位置而无任何向度),则“射击时正好命中该点”的概率,也只能取为 0.

刻画连续型随机变量的概率分布的一个方法,是使用(1.5)式所定义的概率分布函数.但是,在理论和实用上更方便因而更常用的方法,是使用所谓“概率密度函数”或简称密度函数.

定义 1.3 设连续性随机变量 X 有概率分布函数 $F(x)$,则 $F(x)$ 的导数 $f(x) = F'(x)$,称为 X 的概率密度函数.

“密度函数”这名词的来由可解释如下.取定一个点 x ,则按分布函数的定义,事件 $\{x < X \leq x + h\}$ 的概率($h > 0$ 为常数),应为 $F(x + h) - F(x)$.所以,比值 $[F(x + h) - F(x)]/h$ 可以解释为在 x 点附近 h 这么长的区间 $(x, x + h)$ 内,单位长所占有的概率.令 $h \rightarrow 0$,则这个比的极限,即 $F'(x) = f(x)$,也就是在 x 点处(无穷小区段内)单位长的概率,或者说,它反映了概率在 x 点处的“密集程度”.你可以设想一条极细的无穷长的金属杆,总质量为 1,概率密度相当于杆上各点的质量密度.

连续型随机变量 X 的密度函数 $f(x)$ 都具有以下三条基本性质:

- 1° $f(x) \geq 0$
- 2° $\int_{-\infty}^{\infty} f(x) dx = 1$
- 3° 对任何常数 $a < b$ 有*

$$P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x) dx \quad (1.13)$$

1°显然.2°是说“全部概率为 1”.3°是微积分的基本定理(定积分与导数的关系)的直接应用.实际上,2°是 3°当 $a = -\infty$ 和 $b = \infty$ 的特例.

图 2.2(a),(b)分别表示某一连续型变量 X 的分布函数 F 和

* 由于连续型变量取一个点的概率为 0,故区间的端点是否包括在内无影响,也就是说, $\{a \leq X \leq b\}$, $\{a < X < b\}$, $\{a < X \leq b\}$ 和 $\{a \leq X < b\}$ 这四个事件都有同样的概率(1.13).

密度函数 f . 从密度函数的图上可以明显看出该分布的一些特点. 例如概率最大的集中区在 μ 点附近, 而在这点的两边呈对称性的衰减. 图中斜线标出部分的面积表示变量 X 落在 a, b 之间的概率. 这些特点从分布函数的图上就不那么容易看出来.

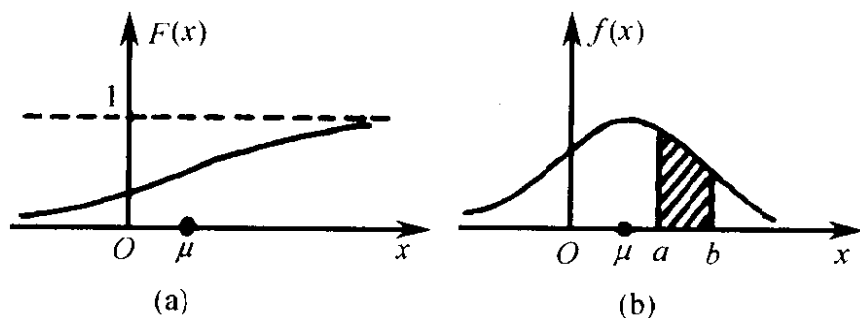


图 2.2

下面举一些重要的连续型分布的例子.

例 1.6 正态分布.

如果一个随机变量具有概率密度函数

$$f(x) = (\sqrt{2\pi}\sigma)^{-1}e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty \quad (1.14)$$

则称 X 为正态随机变量并记为 $X \sim N(\mu, \sigma^2)$. 这里 N 为“Normal”一词的首字母. μ 和 σ^2 都是常数, μ 可以取任何实数值而 $0 < \sigma^2 < \infty$. 它们称为这个分布的“参数”, 其概率意义将在第三章说明.

需要证明 $f(x)$ 确可以作为一个概率密度. 为此须验证 $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x)dx = 1$. 前者显然. 为证后者, 作变数代换 $t = (x - \mu)/\sigma$, 转化为证明

$$I = \int_{-\infty}^{\infty} e^{-t^2/2} dt = \sqrt{2\pi} \quad (1.15)$$

为证此式, 考虑

$$I^2 = \int_{-\infty}^{\infty} e^{-t^2/2} dt \int_{-\infty}^{\infty} e^{-u^2/2} du = \iint_{-\infty}^{\infty} e^{-(t^2+u^2)/2} dt du$$

转化成极坐标 $t = r\cos\theta$, $u = r\sin\theta$, 上式转化为

$$I^2 = \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} e^{-r^2/2} r dr = 2\pi$$

即(1.15).

函数(1.14)的图形约如图 2.2b. 它关于 μ 点对称, 而后往两个方向衰减, 属于“两头低, 中间高”这种正常状况下一般事物下所处的状态. 例如一群人的身高或体重, 特大和特小的居少而中间状态的居多. 举凡人的收入, 大批制造的同一产品的某一指标等, 都在不同程度上符合这一分布. 这不但说明了“正态”这名字的来由, 也说明了这种分布的重要性. 正态分布还有理论上的解释, 这一点留待下一章 3.4 节再谈.

当 $\mu = 1, \sigma^2 = 1$ 时, (1.14) 成为

$$f(x) = e^{-x^2/2} / \sqrt{2\pi} \quad (1.16)$$

它是正态分布 $N(0, 1)$ 的密度函数. $N(0, 1)$ 称为“标准正态分布”. 在概率论著作中, 其密度函数和分布函数常分别记为 $\varphi(x)$ 和 $\Phi(x)$, 并造有很仔细的表. 本书也附有一个简单的 $\Phi(x)$ 的表. 标准正态分布之所以重要, 一个原因在于: 任意的正态分布 $N(\mu, \sigma^2)$ 的计算很容易转化为标准正态分布 $N(0, 1)$. 事实上, 容易证明:

$$\text{若 } X \sim N(\mu, \sigma^2), \text{ 则 } Y = (X - \mu) / \sigma \sim N(0, 1) \quad (1.17)$$

事实上

$$\begin{aligned} P(Y \leq x) &= P((X - \mu) / \sigma \leq x) = P(X \leq \mu + \sigma x) \\ &= (\sqrt{2\pi}\sigma)^{-1} \int_{-\infty}^{\mu + \sigma x} e^{-(t - \mu)^2 / 2\sigma^2} dt \\ &= (\sqrt{2\pi})^{-1} \int_{-\infty}^x e^{-u^2/2} du \end{aligned}$$

其导数, 即 Y 的密度函数, 正是 $(\sqrt{2\pi})^{-1} e^{-x^2/2}$. 这证明了 (1.17).

例如, $X \sim N(1.5, 2^2)$, 要计算 $P(-1 \leq X \leq 2)$, 则因 $(X - 1.5) / 2 \sim N(0, 1)$, 故

$$P(-1 \leq X \leq 2) = P\left(\frac{-1 - 1.5}{2} \leq \frac{X - 1.5}{2} \leq \frac{2 - 1.5}{2}\right)$$

$$\begin{aligned}
&= P(-1.25 \leq (X - 1.5)/2 \leq 0.25) \\
&= \Phi(0.25) - \Phi(-1.25) \quad (1.18)
\end{aligned}$$

然后查标准正态分布 Φ 的表. 表上只有 $\Phi(x)$ 当 $x \geq 0$ 之值. 对 $x < 0$, 可利用公式

$$\Phi(x) = 1 - \Phi(-x) \quad (1.19)$$

而转化为 $x > 0$ 的情况. (1.19) 的证明很简单:

$$\begin{aligned}
\Phi(x) &= (\sqrt{2\pi})^{-1} \int_{-\infty}^x e^{-t^2/2} dt = (\sqrt{2\pi})^{-1} \int_{-x}^{\infty} e^{-t^2/2} dt \\
&= (\sqrt{2\pi})^{-1} \int_{-\infty}^{\infty} e^{-t^2/2} dt - (\sqrt{2\pi})^{-1} \int_{-\infty}^{-x} e^{-t^2/2} dt \\
&= 1 - \Phi(-x)
\end{aligned}$$

用(1.19), 由(1.18)式得

$$P(-1 \leq X \leq 2) = \Phi(0.25) + \Phi(1.25) - 1$$

查 $\Phi(x)$ 的表, 得 $\Phi(0.25) = 0.5987$, $\Phi(1.25) = 0.8944$, 于是得到 $P(-1 \leq X \leq 2) = 0.4931$.

例 1.7 指数分布.

若随机变量 X 有概率密度函数

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{当 } x > 0 \\ 0, & \text{当 } x \leq 0 \end{cases} \quad (1.20)$$

则称 X 服从指数分布. 其中 $\lambda > 0$ 为参数*, 其意义将在后面阐明.

由于 $f(x) = 0$ 当 $x \leq 0$, 表示随机变量取负值的概率为 0: X 只取正值. $f(x)$ 在 $x = 0$ 处之值 $\lambda > 0$, 故密度函数 $f(x)$ 在 $x = 0$ 处不连续. 图 2.3 中描出了这函数当 $\lambda = 1$ (虚线) 和 $\lambda = 2$ (实线) 时的图形.

变量 X 的分布函数易求得为

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & \text{当 } x \leq 0 \\ 1 - e^{-\lambda x}, & \text{当 } x > 0 \end{cases} \quad (1.21)$$

* 因为 $\lambda > 0, x > 0$, (1.2) 式 $e^{-\lambda x}$ 之指数 $-\lambda x$ 总取负值. 由于这个原因, 也有把 (1.20) 称为负指数分布的.

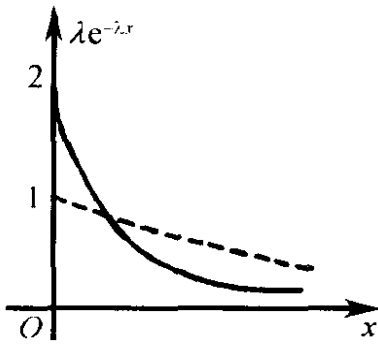


图 2.3

指数分布最常见的一个场合是寿命分布. 设想一种大批生产的电子元件, 其寿命 X 是随机变量. 以 $F(x)$ 记 X 的分布函数. 我们来证明: 在一定的条件下, $F(x)$ 就是(1.21).

我们要作的假定, 从技术上说就是“无老化”. 就是说: “元件在时刻 x 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 x 无关”. 失效率

就是单位长度时间内失效的概率. 用条件概率的形式, 上述假定可表为

$$P(x \leq X \leq x+h | X > x) / h = \lambda, h \rightarrow 0$$

此式解释如下: 元件在时刻 x 时尚正常工作, 表示其寿命大于 x , 即 $X > x$. 在 x 处, 长为 h 的时间段内失效, 即 $x \leq X \leq x+h$. 把这个条件概率除以时间段之长 h , 即得在 x 时刻的平均失效率. 再令 $h \rightarrow 0$, 得瞬时失效率, 按假定, 它应为常数 λ .

按条件概率定义, 注意到 $P(X > x) = 1 - F(x)$, 又

$$\{X > x\} \{x \leq X \leq x+h\} = \{x < X \leq x+h\}$$

有

$$\begin{aligned} & P(x \leq X \leq x+h | X > x) / h \\ &= P(x < X \leq x+h) / (h(1 - F(x))) \\ &= [(F(x+h) - F(x)) / h] / (1 - F(x)) \\ &\rightarrow F'(x) / (1 - F(x)) = \lambda \end{aligned}$$

这个微分方程的通解为 $F(x) = 1 - Ce^{-\lambda x}$ (当 $x > 0$. $x \leq 0$ 时 $F(x)$ 为 0). 常数 C 可用初始条件 $F(0) = 0$ (因为 $F(0) = P(X \leq 0)$, 而寿命 ≤ 0 的概率为 0) 定出为 1, 这样就得到(1.21)式.

从这个推导也可以窥见参数 λ 的意义. λ 为失效率. 失效率愈高, 平均寿命就愈小. 下一章(见第三章例 1.3)将证明: λ^{-1} 就是平均寿命.

由本例可见: 指数分布描述了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.又如人的寿命,一般,在 50 岁或 60 岁以前,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,人的寿命分布在这个阶段也应接近指数分布.

例 1.8 威布尔分布.

若考虑老化,则应取失效率随时间而上升,不能为常数,而应取为一个 x 的增函数,例如 λx^m ,对某个常数 $\lambda > 0, m > 0$.在这个条件下,按上例的推理,将得出:寿命分布 $F(x)$ 满足微分方程 $F'(x)/[1-F(x)] = \lambda x^m$,此与初始条件 $F(0) = 0$ 结合,得出

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

取 $\alpha = m + 1 (\alpha > 1)$,并把 $\lambda / (m + 1)$ 记为 λ ,得出

$$F(x) = 1 - e^{-\lambda x^\alpha}, x > 0 \quad (1.22)$$

而 $F(x) = 0$ 当 $x \leq 0$. 此分布之密度函数为

$$f(x) = \begin{cases} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (1.23)$$

(1.22)和(1.23)分别称为威布尔分布函数和威布尔密度函数.它与指数分布一样,在可靠性统计分析中占重要的地位.实际上指数分布是威布尔分布当 $\alpha = 1$ 时的特例.

例 1.9 均匀分布.

设随机变量 X 有概率密度函数

$$f(x) = \begin{cases} 1/(b-a), & \text{当 } a \leq x \leq b \\ 0, & \text{其他 } x \end{cases} \quad (1.24)$$

则称 X 服从区间 $[a, b]$ 上的均匀分布,并常记为 $X \sim R(a, b)$. 这里 a, b 都是常数, $-\infty < a < b < \infty$. 均匀分布这个名称的来由很明显:因为密度函数 f 在区间 $[a, b]$ 上为常数,故在这区间上,概率在各处的密集程度一样.或者说,概率均匀地分布在这区间上.均匀分布 $R(a, b)$ 的分布函数是

$$F(x) = \begin{cases} 0, & x \leq a \\ (x-a)/(b-a), & a < x < b \\ 1, & x \geq b \end{cases} \quad (1.25)$$

f 和 F 的图形分别如图 2.4a, b 所示.

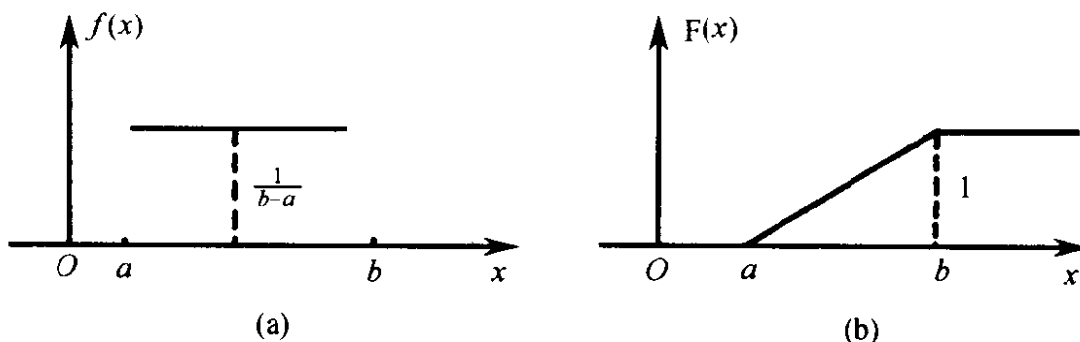


图 2.4

在计算时因“四舍五入”而产生的误差,若以被舍入的那一位的前一位为单位,则可认为这个舍入误差服从均匀分布 $R(-1/2, 1/2)$. 均匀分布的一个好处是:借助于它容易实现对分布的模拟. 首先,若以某种方法产生“随机数”(即像 $0, 1, \dots, 9$ 这十个数字出现的概率都是 $1/10$ 的那种数字,它可以用“摸球”等方式来实现. 实用上用计算机程序可在短时间内产生大量随机数——严格地说,计算机中产生的并非完全随机,但很接近,故有时称为伪随机数),则如取 n 足够大,而独立地产生 n 个随机数字 a_1, \dots, a_n 时,则 $X = 0 \cdot a_1 a_2 \dots a_n$ 就很接近于 $[0, 1]$ 均匀分布 $R(0, 1)$. 对一般分布函数 $F(x)$,若 $F(x)$ 处处连续且严格上升,则其反函数 G 存在,这时易见,若 $X \sim R(0, 1)$,则 $G(X) \sim F$. 事实上, $\{G(X) \leq x\}$ 这个事件,就是 $\{F(G(X)) \leq F(x)\}$ 即 $\{X \leq F(x)\}$,因而(注意到 $R(0, 1)$ 的分布函数为 $F(x) = x$ 当 $0 < x < 1$)

$$P(G(X) \leq x) = P(X \leq F(x)) = F(x)$$

这证明了 $G(X) \sim F$. 这样,用上述模拟方法产生 X 的模拟值后,代入 G 中即得分布 F 的模拟值. 这个方法在模拟研究中常用,而显示了均匀分布的重要性. 均匀分布还有其他重要的理论性质,不能在此细论了.

还有几个在统计应用上很重要的连续型分布,这留待本章 2.4 节去讨论.