

北大社普通高等教育“十三五”数字化建设规划教材

GAILÜLUN YU SHULI TONGJI 概率论与数理统计

韩旭里 谢永钦 / 主编



北京大学出版社
PEKING UNIVERSITY PRESS

北大社普通高等教育“十三五”数字化建设规划教材

概率论与数理统计

主 编 韩旭里 谢永钦



北京大学出版社
PEKING UNIVERSITY PRESS

目 录

第一章 概率论的基本概念	001
第一节 样本空间、随机事件	002
第二节 概率、古典概型	006
第三节 条件概率、全概率公式	015
第四节 独立性	021
小 结	026
习题一	027
第二章 随机变量	032
第一节 随机变量及其分布函数	033
第二节 离散型随机变量及其分布	034
第三节 连续型随机变量及其分布	042
第四节 随机变量函数的分布	051
小 结	054
习题二	056
第三章 随机向量	061
第一节 二维随机向量及其分布	062
第二节 边缘分布	067
第三节 条件分布	070
第四节 随机变量的独立性	073
第五节 两个随机变量函数的分布	074
小 结	081
习题三	083
第四章 随机变量的数字特征	088
第一节 数学期望	089
第二节 方差	098
第三节 协方差与相关系数	103
第四节 矩、协方差矩阵	108
小 结	110
习题四	112
第五章 大数定律与中心极限定理	117
第一节 大数定律	118

第二节 中心极限定理	122
小 结	125
习题五	126
第六章 数理统计的基本概念	128
第一节 随机样本	129
第二节 抽样分布	132
小 结	136
习题六	137
第七章 参数估计	139
第一节 点估计	140
第二节 估计量的评价标准	146
第三节 区间估计	148
小 结	152
习题七	154
第八章 假设检验	158
第一节 概述	159
第二节 单个正态总体的假设检验	163
第三节 两个正态总体的假设检验	170
第四节 总体分布函数的假设检验	175
小 结	179
习题八	180
第九章 方差分析	182
第一节 单因素试验的方差分析	183
第二节 双因素试验的方差分析	188
第三节 正交试验设计及其方差分析	196
小 结	203
习题九	203
第十章 回归分析	206
第一节 回归分析的概述	207
第二节 参数估计	209
第三节 假设检验	213
第四节 预测与控制	216
第五节 非线性回归的线性化处理	218
小 结	220
习题十	221
附表	224
习题参考答案	246
参考文献	255

对表 2-1 所示的一般的分布律, 其分布函数 $F(x)$ 表示一条阶梯状右连续曲线, 在 $X = x_k (k = 1, 2, \dots)$ 处有跳跃, 跳跃的高度恰为 $p_k = P\{X = x_k\}$, 从左至右, 由水平直线 $F(x) = 0$, 分别按阶高 $p_1, p_2, \dots$ 升至水平直线 $F(x) = 1$.

以上是已知分布律求分布函数. 反过来, 若已知离散型随机变量 X 的分布函数 $F(x)$, 则 X 的分布律也可由分布函数所确定:

$$p_k = P\{X = x_k\} = F(x_k) - F(x_k - 0).$$

第三节 连续型随机变量及其分布

上一节我们研究了离散型随机变量, 这类随机变量的特点是它的可能取值及其相对应的概率能被逐个地列出. 这一节我们将要研究的连续型随机变量就不具有这样的性质了. 连续型随机变量的特点是它的可能取值连续地充满某个区间甚至整个数轴. 例如, 测量一个工件长度, 因为在理论上说这个长度的值 X 可以取区间 $(0, +\infty)$ 上的任何一个值. 此外, 连续型随机变量取某特定值的概率总是零 (关于这点将在以后还会说明). 例如, 抽检一个工件其长度 X 丝毫不差刚好是其固定值 (如 1.824 cm) 的事件 $\{X = 1.824\}$ 几乎是不可能的, 应认为 $P\{X = 1.824\} = 0$. 因此讨论连续型随机变量在某点的概率是毫无意义的. 于是, 对于连续型随机变量就不能用对离散型随机变量那样的方法进行研究了. 为了方便论述, 我们先来看一个例子.

例 2.8 || 一个半径为 2 m 的圆盘靶, 设击中靶上任一同心圆盘上的点的概率与该圆盘的面积成正比, 并设射击都能中靶, 以 X 表示弹着点与圆心的距离, 试求随机变量 X 的分布函数.

解 (1) 若 $x < 0$, 因为事件 $\{X \leq x\}$ 是不可能事件, 所以

$$F(x) = P\{X \leq x\} = 0.$$

(2) 若 $0 \leq x \leq 2$, 由题意 $P\{0 \leq X \leq x\} = kx^2$, k 是常数. 为了确定 k 的值, 取 $x = 2$, 有 $P\{0 \leq X \leq 2\} = 2^2 k$, 因事件 $\{0 \leq X \leq 2\}$ 是必然事件, 故 $P\{0 \leq X \leq 2\} = 1$, 即 $2^2 k = 1$, 所以 $k = \frac{1}{4}$, 则

$$P\{0 \leq X \leq x\} = \frac{1}{4}x^2.$$

于是

$$F(x) = P\{X \leq x\} = P\{X < 0\} + P\{0 \leq X \leq x\} = \frac{1}{4}x^2.$$

(3) 若 $x \geq 2$, 由于 $\{X \leq 2\}$ 是必然事件, 因此

$$F(x) = P\{X \leq x\} = 1.$$

综上所述

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{4}x^2, & 0 \leq x < 2, \\ 1, & x \geq 2, \end{cases}$$

它的图形是一条连续曲线, 如图 2-2 所示.

另外, 容易看到本例中 X 的分布函数 $F(x)$ 还可写成如下形式:

$$F(x) = \int_{-\infty}^x f(t) dt,$$

其中
$$f(t) = \begin{cases} \frac{1}{2}t, & 0 < t < 2, \\ 0, & \text{其他.} \end{cases}$$

这就是说 $F(x)$ 恰好是非负函数 $f(t)$ 在区间 $(-\infty, x]$ 上的积分, 我们称这种随机变量 X 为连续型随机变量. 一般地有如下定义.

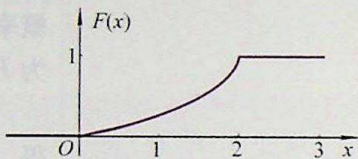


图 2-2

定义 2.3 若对随机变量 X 的分布函数 $F(x)$, 存在非负函数 $f(x)$, 使对于任意实数 x , 有

$$F(x) = \int_{-\infty}^x f(t) dt, \quad (2-8)$$

则称 X 为连续型随机变量, 其中 $f(x)$ 称为 X 的概率密度函数, 简称概率密度或密度函数(density function).

由(2-8)式可知连续型随机变量 X 的分布函数 $F(x)$ 是连续函数. 由分布函数的性质 $F(-\infty) = 0$, $F(+\infty) = 1$ 及 $F(x)$ 单调不减, 知 $F(x)$ 是一条位于直线 $y = 0$ 与 $y = 1$ 之间的单调不减的连续(但不一定光滑)曲线.

由定义 2.3 可知, $f(x)$ 具有以下性质:

- 1° $f(x) \geq 0$;
- 2° $\int_{-\infty}^{+\infty} f(x) dx = 1$;
- 3° $P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1)$
 $= \int_{x_1}^{x_2} f(x) dx \quad (x_1 \leq x_2);$

4° 若 $f(x)$ 在 x 点处连续, 则有 $F'(x) = f(x)$.

由性质 2° 可知, 介于曲线 $y = f(x)$ 与 $y = 0$ 之间的面积为 1; 由性质 3° 可知, X 落在区间 $(x_1, x_2]$ 内的概率 $P\{x_1 < X \leq x_2\}$ 等于区间 $(x_1, x_2]$ 上曲线 $y = f(x)$ 之下曲边梯形的面积; 由性质 4° 可知, 在 $f(x)$ 的连续点 x 处有



拓展知识

$$f(x) = \lim_{\Delta x \rightarrow 0^+} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0^+} \frac{P\{x < X \leq x + \Delta x\}}{\Delta x}.$$

这种形式恰与物理学中线密度定义相类似,这也正是为什么称 $f(x)$ 为概率密度的原因. 同样我们也要指出,反过来,任一满足以上 1°, 2° 两个性质的函数 $f(x)$, 一定可以作为某个连续型随机变量的密度函数.

前面我们曾指出对连续型随机变量 X 而言, 取任一特定值 a 的概率为零, 即 $P\{X = a\} = 0$. 事实上, 令 $\Delta x > 0$, 设 X 的分布函数为 $F(x)$, 则由

$$\{X = a\} \subset \{a - \Delta x < X \leq a\},$$

得

$$0 \leq P\{X = a\} \leq P\{a - \Delta x < X \leq a\} = F(a) - F(a - \Delta x).$$

由于 $F(x)$ 连续, 因此

$$\lim_{\Delta x \rightarrow 0} F(a - \Delta x) = F(a),$$

当 $\Delta x \rightarrow 0$ 时, 由夹逼定理, 得

$$P\{X = a\} = 0,$$

由此很容易推导出

$$\begin{aligned} P\{a \leq X < b\} &= P\{a < X \leq b\} = P\{a \leq X \leq b\} \\ &= P\{a < X < b\}, \end{aligned}$$

即在计算连续型随机变量落在某区间上的概率时, 可不必区分该区间端点的情况. 此外还要说明的是, 事件 $\{X = a\}$ “几乎不可能发生”, 但并不保证绝不会发生, 它是“零概率事件”而不是不可能事件.

例 2.9

设连续型随机变量 X 的分布函数为

$$F(x) = \begin{cases} 0, & x < 0, \\ Ax^2, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

试求:

- (1) 系数 A ;
- (2) X 落在区间 $(0.3, 0.7)$ 内的概率;
- (3) X 的密度函数.

解 (1) 由于 X 为连续型随机变量, 故 $F(x)$ 是连续函数, 因此有

$$1 = F(1) = \lim_{x \rightarrow 1^-} F(x) = \lim_{x \rightarrow 1^-} Ax^2 = A,$$

即 $A = 1$, 于是有

$$F(x) = \begin{cases} 0, & x < 0, \\ x^2, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

$$(2) P\{0.3 < X < 0.7\} = F(0.7) - F(0.3) = (0.7)^2 - (0.3)^2 = 0.4.$$

(3) X 的密度函数为

$$f(x) = F'(x) = \begin{cases} 2x, & 0 \leq x < 1, \\ 0, & \text{其他.} \end{cases}$$

由定义 2.3 知, 改变密度函数 $f(x)$ 在个别点的函数值, 不影响分布函数 $F(x)$ 的取值, 因此, 并不在乎改变密度函数在个别点上的值(比如在 $x = 0$ 或 $x = 1$ 上 $f(x)$ 的值).

例 2.10 || 设随机变量 X 具有密度函数

$$f(x) = \begin{cases} kx, & 0 \leq x < 3, \\ 2 - \frac{x}{2}, & 3 \leq x \leq 4, \\ 0, & \text{其他.} \end{cases}$$

试求:

- (1) 系数 k ;
- (2) X 的分布函数 $F(x)$;
- (3) $P\left\{1 < X \leq \frac{7}{2}\right\}$.

解 (1) 由 $\int_{-\infty}^{+\infty} f(x) dx = 1$, 得

$$\int_0^3 kx dx + \int_3^4 \left(2 - \frac{x}{2}\right) dx = 1,$$

解得 $k = \frac{1}{6}$, 故 X 的密度函数为

$$f(x) = \begin{cases} \frac{x}{6}, & 0 \leq x < 3, \\ 2 - \frac{x}{2}, & 3 \leq x \leq 4, \\ 0, & \text{其他.} \end{cases}$$

(2) 当 $x < 0$ 时, $F(x) = P\{X \leq x\} = \int_{-\infty}^x f(t) dt = 0$;

$$\begin{aligned} \text{当 } 0 \leq x < 3 \text{ 时, } F(x) &= P\{X \leq x\} = \int_{-\infty}^x f(t) dt \\ &= \int_{-\infty}^0 f(t) dt + \int_0^x f(t) dt \\ &= \int_0^x \frac{t}{6} dt = \frac{x^2}{12}; \end{aligned}$$

$$\begin{aligned} \text{当 } 3 \leq x < 4 \text{ 时, } F(x) &= P\{X \leq x\} = \int_{-\infty}^x f(t) dt \\ &= \int_{-\infty}^0 f(t) dt + \int_0^3 f(t) dt + \int_3^x f(t) dt \end{aligned}$$

$$= \int_0^3 \frac{t}{6} dt + \int_3^x \left(2 - \frac{t}{2}\right) dt$$

$$= -\frac{x^2}{4} + 2x - 3;$$

当 $x \geq 4$ 时, $F(x) = P\{X \leq x\} = \int_{-\infty}^x f(t) dt$

$$= \int_{-\infty}^0 f(t) dt + \int_0^3 f(t) dt + \int_3^4 f(t) dt + \int_4^x f(t) dt$$

$$= \int_0^3 \frac{t}{6} dt + \int_3^4 \left(2 - \frac{t}{2}\right) dt = 1.$$

故

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{x^2}{12}, & 0 \leq x < 3, \\ -\frac{x^2}{4} + 2x - 3, & 3 \leq x < 4, \\ 1, & x \geq 4. \end{cases}$$

$$(3) P\left\{1 < X \leq \frac{7}{2}\right\} = F\left(\frac{7}{2}\right) - F(1) = \frac{41}{48}.$$

下面介绍 3 种常见的连续型随机变量.

1. 均匀分布

若连续型随机变量 X 具有概率密度

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0, & \text{其他}, \end{cases} \quad (2-9)$$

则称 X 在区间 (a, b) 上服从均匀分布(uniform distribution), 记为 $X \sim U(a, b)$.

易知 $f(x) \geq 0$, 且

$$\int_{-\infty}^{+\infty} f(x) dx = \int_a^b \frac{1}{b-a} dx = 1.$$

由(2-9)式可得

$$1^\circ \quad P\{X \geq b\} = \int_b^{+\infty} 0 dx = 0, P\{X \leq a\} = \int_{-\infty}^a 0 dx = 0,$$

即

$$P\{a < X < b\} = 1 - P\{X \geq b\} - P\{X \leq a\} = 1;$$

2° 若 $a \leq c < d \leq b$, 则

$$P\{c < X < d\} = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}.$$

因此,在区间 (a,b) 上服从均匀分布的随机变量 X 的物理意义是: X 以概率1在区间 (a,b) 内取值,而以概率0在区间 (a,b) 以外取值,并且 X 值落入 (a,b) 中任一子区间 (c,d) 中的概率与子区间的长度成正比,而与子区间的位置无关.

由(2-8)式易得 X 的分布函数为

$$F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x < b, \\ 1, & x \geq b. \end{cases} \quad (2-10)$$

密度函数 $f(x)$ 和分布函数 $F(x)$ 的图形分别如图2-3和图2-4所示.

在数值计算中,由于四舍五入,小数点后第一位小数所引起的误差 X 一般可以看做是一个在 $[-0.5, 0.5]$ 上服从均匀分布的随机变量;又如在 (a,b) 中随机掷质点,则该质点的坐标 X 一般也可看做是一个在 (a,b) 上服从均匀分布的随机变量.

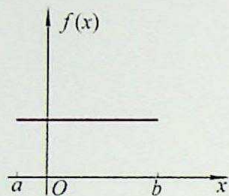


图 2-3

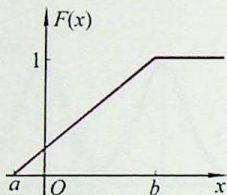


图 2-4

例 2.11 || 某公共汽车站从上午7:00开始,每15 min来一辆车,如某乘客到达此站的时间是7:00—7:30之间的服从均匀分布的随机变量,试求乘客等车少于5 min的概率.

解 设乘客于7:00过 X min到达车站,由于 X 在 $[0, 30]$ 上服从均匀分布,即有

$$f(x) = \begin{cases} \frac{1}{30}, & 0 \leq x \leq 30, \\ 0, & \text{其他.} \end{cases}$$

显然,只有乘客在7:10—7:15之间或7:25—7:30之间到达车站时,他(或她)等车的时间才少于5 min,因此所求概率为

$$P\{10 < X \leq 15\} + P\{25 < X \leq 30\} = \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{3}.$$

2. 指数分布

若随机变量 X 的密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases} \quad (2-11)$$

其中 $\lambda > 0$ 为常数,则称 X 服从参数为 λ 的指数分布(exponential distribution),记作 $X \sim E(\lambda)$.

显然 $f(x) \geq 0$,且 $\int_{-\infty}^{+\infty} f(x) dx = \int_0^{+\infty} \lambda e^{-\lambda x} dx = 1$.

容易得到 X 的分布函数为

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

指数分布最常见的一个场合是寿命分布. 因为指数分布具有“无记忆性”, 即对于任意 $s, t > 0$, 有

$$P\{X > s+t | X > s\} = P\{X > t\}. \quad (2-12)$$

如果用 X 表示某一元件的寿命, 那么上式表明, 在已知元件已使用了 s h 的条件下, 它还能再使用至少 t h 的概率, 与从开始使用时算起它至少能使用 t h 的概率相等. 这就是说该元件对它已使用过 s h 没有记忆. 当然, 指数分布描述的是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.

(2-12) 式是容易证明的. 事实上,

$$\begin{aligned} P\{X > s+t | X > s\} &= \frac{P\{X > s, X > s+t\}}{P\{X > s\}} = \frac{P\{X > s+t\}}{P\{X > s\}} \\ &= \frac{1 - F(s+t)}{1 - F(s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t} = P\{X > t\}. \end{aligned}$$

3. 正态分布

若连续型随机变量 X 的密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty, \quad (2-13)$$

其中 $\mu, \sigma (\sigma > 0)$ 为常数, 则称 X 服从参数为 μ, σ 的正态分布 (normal distribution), 记为 $X \sim N(\mu, \sigma^2)$.

显然 $f(x) \geq 0$, 下面来证明 $\int_{-\infty}^{+\infty} f(x) dx = 1$. 令 $\frac{x-\mu}{\sigma} = t$, 得到

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt,$$

记 $I = \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$, 则有

$$I^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{(t^2+s^2)}{2}} dt ds.$$

作极坐标变换: $s = r \cos \theta, t = r \sin \theta$, 得到

$$I^2 = \int_0^{2\pi} \int_0^{+\infty} r e^{-\frac{r^2}{2}} dr d\theta = 2\pi,$$

而 $I > 0$, 故有 $I = \sqrt{2\pi}$, 即有

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi},$$

于是

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1.$$

正态分布是概率论和数理统计中最重要的分布之一. 在实际问题中大量的随机变量服从或近似服从正态分布. 只要某一个随机变量受到许多相互独立随机因素的影响, 而每个个别因素的影响都不能起决定性作用, 那么就可以断定该随机变量服从或近似服从正态分布. 例如, 因为人的身高、体重受到种族、饮食习惯、地域、运动等因素影响, 但这些因素又不能对身高、体重起决定性作用, 所以我们可以认为身高、体重服从或近似服从正态分布.

参数 μ, σ 的意义将在第四章中说明. $f(x)$ 的图形如图 2-5 所示, 它具有如下性质:

- 1° 曲线关于 $x = \mu$ 对称;
- 2° 曲线在 $x = \mu$ 处取到最大值, x 离 μ 越远, $f(x)$ 值越小, 这表明对于同样长度的区间, 当区间离 μ 越远, X 落在这个区间上的概率越小;

3° 曲线在 $\mu \pm \sigma$ 处有拐点;

4° 曲线以 x 轴为渐近线;

5° 若固定 μ , 当 σ 越小时图形越尖陡 (见图 2-6), 因而 X 落在 μ 附近的概率越大; 若固定 σ , μ 值改变, 则图形沿 x 轴平移, 而不改变其形状. 故称 σ 为精度参数, μ 为位置参数.

由 (2-13) 式得 X 的分布函数为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt. \quad (2-14)$$

特别地, 当 $\mu = 0, \sigma = 1$ 时, 称 X 服从标准正态分布 $N(0, 1)$, 其密度函数和分布函数分别用 $\varphi(x), \Phi(x)$ 表示, 即有

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad (2-15)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt. \quad (2-16)$$

易知, $\Phi(-x) = 1 - \Phi(x)$.

人们已事先编制了 $\Phi(x)$ 的函数值表 (见本书附表 2).

一般地, 若 $X \sim N(\mu, \sigma^2)$, 则有 $\frac{X-\mu}{\sigma} \sim N(0, 1)$.

事实上, $Z = \frac{X-\mu}{\sigma}$ 的分布函数为

$$\begin{aligned} P\{Z \leq x\} &= P\left\{\frac{X-\mu}{\sigma} \leq x\right\} = P\{X \leq \mu + \sigma x\} \\ &= \int_{-\infty}^{\mu+\sigma x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \end{aligned}$$

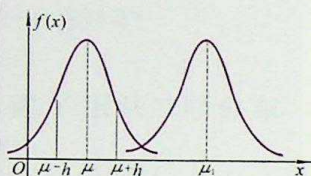


图 2-5

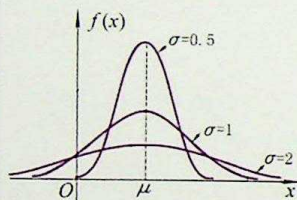


图 2-6

令 $\frac{t-\mu}{\sigma} = s$, 得

$$P\{Z \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{s^2}{2}} ds = \Phi(x),$$

由此知 $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

因此,若 $X \sim N(\mu, \sigma^2)$, 则可利用标准正态分布函数 $\Phi(x)$, 通过查表求得 X 落在任一区间 $(x_1, x_2]$ 内的概率, 即

$$\begin{aligned} P\{x_1 < X \leq x_2\} &= P\left\{\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{x_2-\mu}{\sigma}\right\} \\ &= P\left\{\frac{X-\mu}{\sigma} \leq \frac{x_2-\mu}{\sigma}\right\} - P\left\{\frac{X-\mu}{\sigma} \leq \frac{x_1-\mu}{\sigma}\right\} \\ &= \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right). \end{aligned}$$

例如, 设 $X \sim N(1.5, 4)$, 可得

$$\begin{aligned} P\{-1 \leq X \leq 2\} &= P\left\{\frac{-1-1.5}{2} \leq \frac{X-1.5}{2} \leq \frac{2-1.5}{2}\right\} \\ &= \Phi(0.25) - \Phi(-1.25) \\ &= \Phi(0.25) - [1 - \Phi(1.25)] \\ &= 0.5987 - 1 + 0.8944 = 0.4931. \end{aligned}$$

设 $X \sim N(\mu, \sigma^2)$, 由 $\Phi(x)$ 的函数表可得

$$P\{\mu - \sigma < X < \mu + \sigma\} = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.6826,$$

$$P\{\mu - 2\sigma < X < \mu + 2\sigma\} = \Phi(2) - \Phi(-2) = 0.9544,$$

$$P\{\mu - 3\sigma < X < \mu + 3\sigma\} = \Phi(3) - \Phi(-3) = 0.9974.$$

我们看到, 尽管正态变量的取值范围是 $(-\infty, +\infty)$, 但它的值落在 $(\mu - 3\sigma, \mu + 3\sigma)$ 内几乎是肯定的事, 因此在实际问题中, 基本上可以认为有 $|X - \mu| < 3\sigma$. 这就是人们所说的“ 3σ 原则”.

例 2.12 公共汽车车门的高度是按成年男子与车门顶碰头的机会在 1% 以下来设计的. 设男子身高 X 服从 $\mu = 170(\text{cm})$, $\sigma = 6(\text{cm})$ 的正态分布, 即 $X \sim N(170, 6^2)$, 问车门高度应如何确定?

解 设车门高度为 $h(\text{cm})$, 按设计要求 $P\{X \geq h\} \leq 0.01$ 或 $P\{X < h\} \geq 0.99$, 因为 $X \sim N(170, 6^2)$, 所以

$$P\{X < h\} = P\left\{\frac{X-170}{6} < \frac{h-170}{6}\right\} = \Phi\left(\frac{h-170}{6}\right) \geq 0.99,$$

查表得

$$\Phi(2.33) = 0.9901 > 0.99.$$

故取 $\frac{h-170}{6} = 2.33$, 则 $h \approx 184$, 即设计车门高度为 184 cm 时, 可使成年男子与车门碰头的机会不超过 1%.

例 2.13 || 测量到某一目标的距离时发生的随机误差 X (单位: m) 具有密度函数

$$f(x) = \frac{1}{40\sqrt{2\pi}} e^{-\frac{(x-20)^2}{3200}}.$$

试求在 3 次测量中至少有一次误差的绝对值不超过 30 m 的概率.

解 X 的密度函数为

$$f(x) = \frac{1}{40\sqrt{2\pi}} e^{-\frac{(x-20)^2}{3200}} = \frac{1}{\sqrt{2\pi} \times 40} e^{-\frac{(x-20)^2}{2 \times 40^2}}.$$

即 $X \sim N(20, 40^2)$, 故一次测量中随机误差的绝对值不超过 30 m 的概率为

$$\begin{aligned} P\{|X| \leq 30\} &= P\{-30 \leq X \leq 30\} = \Phi\left(\frac{30-20}{40}\right) - \Phi\left(\frac{-30-20}{40}\right) \\ &= \Phi(0.25) - \Phi(-1.25) = 0.5987 - (1 - 0.8944) \\ &= 0.4931. \end{aligned}$$

设 Y 为 3 次测量中误差的绝对值不超过 30 m 的次数, 则 Y 服从二项分布 $b(3, 0.4931)$, 故

$$P\{Y \geq 1\} = 1 - P\{Y = 0\} = 1 - (0.5069)^3 \approx 0.8698.$$

为了便于今后应用, 对于标准正态变量, 我们引入了 α 分位点的定义.

设 $X \sim N(0, 1)$, 若 z_α 满足条件:

$$P\{X > z_\alpha\} = \alpha, \quad 0 < \alpha < 1, \quad (2-17)$$

则称点 z_α 为标准正态分布的上 α 分位点. 例如, 由查表可得 $z_{0.05} = 1.645$, $z_{0.001} = 3.1$. 故 1.645 与 3.1 分别是标准正态分布的上 0.05 分位点与上 0.001 分位点.

第四节 随机变量函数的分布

我们常常遇到一些随机变量, 它们的分布往往难于直接得到 (如测量轴承滚珠体积值 Y 等), 但是与它们有函数关系的另一些随机变量, 其分布却是容易知道的 (如滚珠直径测量值 X). 因此, 要研究随机变量之间的函数关系, 从而通过这种关系由已知的随机变量的分布求出与其有函数关系的另一个随机变量的分布.

例 2.14 || 设随机变量 X 具有表 2-6 所示的分布律, 试求 X^2 的分布律.

表 2-6

X	-1	0	1	1.5	3
p_i	0.2	0.1	0.3	0.3	0.1

图书在版编目(CIP)数据

概率论与数理统计/韩旭里, 谢永钦主编. —北京: 北京大学出版社, 2018.7

ISBN 978-7-301-29547-2

I. ①概… II. ①韩… ②谢… III. ①概率论—高等学校—教材 ②数理统计—高等学校—教材
IV. ①O21

中国版本图书馆 CIP 数据核字(2018)第 098740 号

- 书 名 概率论与数理统计
GAILÜLUN YU SHULI TONGJI
- 著作责任者 韩旭里 谢永钦 主编
- 责任编辑 潘丽娜
- 标准书号 ISBN 978-7-301-29547-2
- 出版发行 北京大学出版社
- 地 址 北京市海淀区成府路 205 号 100871
- 网 址 <http://www.pup.cn>
- 电子信箱 zpup@pup.cn
- 新浪微博 @北京大学出版社
- 电 话 邮购部 62752015 发行部 62750672 编辑部 62752021
- 印刷者 长沙超峰印刷有限公司
- 经 销 者 新华书店
- 787 毫米×1092 毫米 16 开本 16.75 印张 412 千字
2018 年 7 月第 1 版 2023 年 6 月第 9 次印刷
- 定 价 49.50 元

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有, 侵权必究

举报电话: 010-62752024 电子邮箱: fd@pup.pku.edu.cn

图书如有印装质量问题, 请与出版部联系, 电话: 010-62756370

GAILÜLUN YU SHULI TONGJI 概率论与数理统计

配套云资源的使用说明



打开微信，扫描关注
公众号



刮开涂层，点击公众号里
的激活课程，扫描激活



“北京大学出版社”
微信公众号

ISBN 978-7-301-29547-2



9 787301 295472

定价：49.50元