概率论 · 复习概要

黄正华*

2012 年 12 月 26 日

目录

*Email: huangzh@whu.edu.cn

0 考点提要

0.1 重要知识点

(1) 六种重要的分布.

(2) 六种重要分布的期望与方差.

(3) $D(aX \pm bY) = a^2D(X) + b^2D(y)$, 其中 X 与 Y 相互独立.

(4) 密度函数 $f(x)$ 的概念与求法.

(5) 分布函数 $F(x)$ 的概念与求法.

(6) 边缘密度函数、联合密度函数的概念与求法.

(7) 全概率公式和 Bayes 公式.

(8) 中心极限定理.

(9) 正态分布的标准化: 若 $X \sim N(\mu, \sigma^2)$, 则 $\dfrac{X-\mu}{\sigma} \sim N(0, 1)$.

(10) 独立性与相关性.

0.2 常见考点

(1) 全概率公式和 Bayes 公式.

(2) 求分布函数 $F(x)$.

(3) 求 $Y = g(X)$ 的密度函数.

(4) 求边缘密度函数、联合密度函数.

(5) 正态分布的性质.

(6) 期望与方差的性质.

(7) 期望与方差的计算.

(8) 独立性、相关性的判断.

(9) 中心极限定理.

1 随机事件与概率

主要问题: 事件间的关系与运算, 条件概率, 乘法公式, 全概率公式, 贝叶斯公式, 事件的独立性, 伯努利概型.

常见考点: 全概率公式和 Bayes 公式; 加法公式与乘法公式.

1.1 加法公式与乘法公式

和事件的概率 $P(A \cup B)$ 在不同场合下的求法:

- 一般形式:

$$P(A \cup B) = P(A) + P(B) - P(AB),$$
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

- 若 A, B 互不相容: $P(A\cup B)=P(A)+P(B)$.
- 若 A, B 相互独立:

$$\begin{aligned}P(A\cup B)&=1-P(\overline{A\cup B})=1-P(\overline{A}\,\overline{B})\\&=1-P(\overline{A})P(\overline{B}).\end{aligned}$$

积事件的概率 $P(AB)$ 的求法:

- 一般形式:

$$P(AB)=P(A)P(B|A)=P(B)P(A|B).$$

- 若 A, B 相互独立:

$$P(AB)=P(A)P(B).$$

乘法公式来自于条件概率公式:

$$P(B|A)=\frac{P(AB)}{P(A)}.$$

1.2 全概率公式和 Bayes 公式

全概率公式

$$\begin{aligned}P(A)&=P(AB_1)+P(AB_2)+\cdots+P(AB_n)\\&=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+\cdots+P(B_n)P(A|B_n).\end{aligned}$$

Bayes 公式

$$P\left(B_i|A\right)=\frac{P\left(B_i\right)P\left(A|B_i\right)}{\sum\limits_{j=1}^{n}P\left(B_j\right)P\left(A|B_j\right)},\quad i=1,2,\cdots,n. \tag{1}$$

Bayes 公式本质上是条件概率公式:

$$P\left(B_i|A\right)=\frac{P(AB_i)}{P(A)},$$

只是其分子、分母进一步分别使用了乘法公式和全概率公式.

- 全概率公式表达了“综合考虑引起结果 A 的各种原因 B_i, 计算导致结果 A 出现的可能性的大小”; 如果一个事件的发生有多个“诱因”, 就要用到全概率公式.
- Bayes 公式则反映了“当结果 A 出现时, 它是由原因 B_i 引起的可能性的大小”. Bayes 公式常用来追究责任, 或者“执果索因”. 也就是计算各个“诱因”对事件发生的“贡献”.

典型例题:

例 1 发报台分别以概率 0.6 和 0.4 发出信号 0 和 1. 由于通讯系统受到干扰, 当发出信号 0 时, 收报台未必收到信号 0, 而是分别以 0.8 和 0.2 的概率收到 0 和 1; 同样, 发出 1 时分别以 0.9 和 0.1 的概率收到 1 和 0. 如果收报台收到 0, 问它没收错的概率? (答案: 12/13.)

1.3 其他

- 抽签公平性. (见教材 P.13 例 1.3.3)
- $A \cup B = (A - B) \cup B = (A - AB) \cup B$ 的灵活应用.
- 建议要看的题型
  - 事件的关系与运算: P.35 习题 6, 7.
  - 全概率公式和 Bayes 公式: P.36 习题 28.
  - 伯努利概型: P.38 习题 8.

2 随机变量及其概率分布

主要问题: 密度函数 $f(x)$, 分布函数 $F(x)$ 的定义与性质; 六种重要分布; 随机变量的函数的分布.

常见考点: $f(x)$, $F(x)$ 的性质; 求密度函数 $f(x)$; 求分布函数 $F(x)$; 求 $Y = g(X)$ 的密度函数.

2.1 已知密度函数 $f(x)$, 求分布函数 $F(x)$

密度函数 $f(x)$ 一般是分段函数. 由 $f(x)$ 求 $F(x)$, 本质上是分段函数求积分的问题, 是大家的薄弱环节. 要引起重视!

典型例题:

例 2 设随机变量 X 具有概率密度

$$f(x) = \begin{cases} kx, & 0 \leqslant x < 3, \\ 2 - \dfrac{x}{2}, & 3 \leqslant x \leqslant 4, \\ 0, & \text{其它}. \end{cases}$$

① 确定常数 k; ② 求 X 的分布函数 $F(x)$; ③ 求 $P\{1 < X \leqslant 3.5\}$.

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

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

解得 $k = \dfrac{1}{6}$.

更为详细的解释是: 由积分的区域可加性, 得

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x &= \int_{-\infty}^{0} f(x)\,\mathrm{d}x + \int_0^3 f(x)\,\mathrm{d}x + \int_3^4 f(x)\,\mathrm{d}x + \int_4^{+\infty} f(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{0} 0\,\mathrm{d}x + \int_0^3 kx\,\mathrm{d}x + \int_3^4 (2 - \frac{x}{2})\,\mathrm{d}x + \int_4^{+\infty} 0\,\mathrm{d}x \\ &= \int_0^3 kx\,\mathrm{d}x + \int_3^4 (2 - \frac{x}{2})\,\mathrm{d}x. \end{aligned}$$

从而 X 的概率密度函数为

$$f(x)=\begin{cases}\dfrac{x}{6}, & 0\leqslant x<3,\\ 2-\dfrac{x}{2}, & 3\leqslant x\leqslant 4,\\ 0, & \text{其它}.\end{cases}$$

② 对 $F(x)=\int_{-\infty}^{x} f(t)\,\mathrm{d}t$,

- $x<0$ 时, $F(x)=\int_{-\infty}^{x} f(t)\,\mathrm{d}t=\int_{-\infty}^{x} 0\,\mathrm{d}t=0$.
- $0\leqslant x<3$ 时, $F(x)=\int_{-\infty}^{0} 0\,\mathrm{d}t+\int_{0}^{x}\frac{t}{6}\,\mathrm{d}t=\frac{x^2}{12}$.
- $3\leqslant x<4$ 时,

$$\begin{aligned}F(x)&=\int_{-\infty}^{0} 0\,\mathrm{d}t+\int_{0}^{3}\frac{t}{6}\,\mathrm{d}t+\int_{3}^{x}(2-\frac{t}{2})\,\mathrm{d}t\\ &=-3+2x-\frac{x^2}{4}.\end{aligned}$$

- $x\geqslant 4$ 时, $F(x)=1$.

即

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

☞ 做完一定要验算: $F'(x)=f(x)$.

③ $P\{1<x\leqslant 3.5\}=F(3.5)-F(1)=41/48$. ■

2.2 随机变量的函数

已知连续型随机变量 X 的概率密度 $f_X(x)$, 求随机变量 $Y=g(X)$ 的概率密度 $f_Y(y)$, 两种方法:

- 分布函数微分法;
- 积分转化法.

典型例题:

例 3 设随机变量 X 具有概率密度

$$f_X(x)=\begin{cases}\dfrac{x}{8}, & 0<x<4,\\ 0, & \text{其他}.\end{cases}$$

求 $Y=2X+8$ 的概率密度 $f_Y(y)$.

解 先求 Y 的分布函数. (请自己注明下述各个步骤的理由.)

$$\begin{aligned}F_Y(y) &= P\{Y \leqslant y\}\\ &= P\{2X+8 \leqslant y\}\\ &= P\left\{X \leqslant \frac{y-8}{2}\right\}\\ &= \int_{-\infty}^{\frac{y-8}{2}} f_X(x)\,\mathrm{d}x.\end{aligned}$$

注意到积分上限函数求导法则 $\left(\int_{-\infty}^{\varphi(x)} f(x)\,\mathrm{d}x\right)' = f\big(\varphi(x)\big)\varphi'(x)$, 上式两端关于 y 求导, 得

$$\begin{aligned}f_Y(y) &= f_X\left(\frac{y-8}{2}\right)\cdot\left(\frac{y-8}{2}\right)'_y\\ &= \frac{1}{2}f_X\left(\frac{y-8}{2}\right)\\ &= \begin{cases}\dfrac{1}{2}\cdot\dfrac{\frac{y-8}{2}}{8}, & 0<\dfrac{y-8}{2}<4,\\ 0, & \text{其他}.\end{cases}\\ &= \begin{cases}\dfrac{y-8}{32}, & 8<y<16,\\ 0, & \text{其他}.\end{cases}\end{aligned}$$

■

上述方法体现为下面的一般结论. 称为**单调函数公式法**:

已知随机变量 X 具有概率密度

$$f_X(x) = \begin{cases} f(x), & a<x<b,\\ 0, & \text{其他}.\end{cases}$$

对随机变量 $Y=g(X)$, 要求 $f_Y(y)$. 则对函数关系 $y=g(x)$, 给出反函数 $x=h(y)$, 有

$$f_Y(y) = \begin{cases} f\big(h(y)\big)\cdot|h'(y)|, & a<h(y)<b,\\ 0, & \text{其他}.\end{cases}$$

其中函数 $y=g(x)$ 处处可导且单调.

积分转化法参看教材 P.77.

2.3 正态分布

正态分布的全部细节都要非常清楚!

- 正态分布的标准化: 若 $X\sim N(\mu,\sigma^2)$, 则 $\dfrac{X-\mu}{\sigma}\sim N(0,1)$.

对一般的随机变量也可以 "标准化", 即使它不一定服从正态分布. 事实上, X 标准化变量为

$$X^* = \frac{X - E(X)}{\sqrt{D(X)}},$$

则

$$E(X^*) = 0, \qquad D(X^*) = 1.$$

- 正态分布的再生性: 设 X, Y 相互独立, $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, 则

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2), \tag{2}$$
$$aX \pm bY \sim N(a\mu_1 \pm b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2). \tag{3}$$

- $\Phi(-x) = 1 - \Phi(x)$.
- $z_{1-\alpha} = -z_\alpha$.

例 4 设随机变量 X, Y 相互独立, $X \sim N(-1, 2)$, $Y \sim N(1, 3)$, 则 $X + 2Y \sim$____________, $X - 2Y \sim$____________.

解 $X + 2Y \sim N(1, 14)$, $X - 2Y \sim N(-3, 14)$. ■

3 多维随机变量及其概率分布

主要问题: 边缘概率密度, 联合概率密度; 随机变量的独立性; 随机变量函数的分布

常见考点: 求边缘概率密度 $f_X(x)$; 独立性的判断; 卷积公式.

3.1 边缘概率密度的求法

计算公式:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)\,\mathrm{d}y,$$
$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)\,\mathrm{d}x.$$

重要题型: P.95 例 3.3.2.

3.2 独立性的判断

独立性的判断, 即看下列式子是否成立:

$$\text{联合} = \text{边缘} \times \text{边缘}.$$

重要题型: P.105 例 3.4.4.

3.3 随机变量函数的分布

例 5 设二维随机变量 (X, Y) 的概率密度为

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

求: (I) (X, Y) 的边缘概率密度 $f_X(x)$, $f_Y(y)$;

(II) $Z = 2X - Y$ 的概率密度 $f_Z(z)$.

解 (I) 注意到 $f(x, y)$ 在 X–型区域 $\begin{cases} 0 < y < 2x, \\ 0 < x < 1 \end{cases}$ 上有非零表达式, 该区域可以转化为 Y–型区域 $\begin{cases} \frac{y}{2} < x < 1, \\ 0 < y < 2. \end{cases}$ 则

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)\,\mathrm{d}y = \begin{cases} \int_0^{2x} \mathrm{d}y, & 0 < x < 1, \\ 0, & \text{其他}. \end{cases} = \begin{cases} 2x, & 0 < x < 1, \\ 0, & \text{其他}. \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)\,\mathrm{d}x = \begin{cases} \int_{\frac{y}{2}}^{1} \mathrm{d}x, & 0 < y < 2, \\ 0, & \text{其他}. \end{cases} = \begin{cases} 1 - \dfrac{y}{2}, & 0 < y < 2, \\ 0, & \text{其他}. \end{cases}$$

(II) 用积分转化法. 此时 $g(x, y) = 2x - y$. 对任何有界连续函数 $h(z)$,

$$\begin{aligned} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h\big[g(x, y)\big] f(x, y)\,\mathrm{d}x\,\mathrm{d}y &= \int_0^1 \Big(\int_0^{2x} h(2x - y)\cdot 1\,\mathrm{d}y\Big)\,\mathrm{d}x \\ &= \int_0^1 \Big(\int_{2x}^{0} h(z)(-1)\,\mathrm{d}z\Big)\,\mathrm{d}x \qquad (\text{换元 } z = 2x - y) \\ &= \int_0^1 \Big(\int_0^{2x} h(z)\,\mathrm{d}z\Big)\,\mathrm{d}x \\ &= \int_0^2 \Big(h(z)\int_{\frac{z}{2}}^{1} \mathrm{d}x\Big)\,\mathrm{d}z \qquad (\text{交换积分次序}) \\ &= \int_0^2 h(z)\big(1 - \frac{z}{2}\big)\,\mathrm{d}z, \end{aligned}$$

得 Z 的概率密度为

$$f_Z(z) = \begin{cases} 1 - \dfrac{z}{2}, & 0 < z < 2, \\ 0, & \text{其他}. \end{cases}$$

■

4 随机变量的数字特征

主要问题: 期望的性质、计算; 方差的性质、计算; 协方差与相关系数的计算.

常见考点: 求 $E(x)$, $D(x)$; 计算协方差与相关系数.

4.1 期望的性质与计算

计算公式

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x,$$
$$E\big(g(X)\big) = \int_{-\infty}^{+\infty} g(x) f(x)\,\mathrm{d}x.$$

4.2 方差的性质与计算

方差的计算:

$$\begin{aligned} D(X) &= E\big[\big(X - E(X)\big)^2\big] && \text{(定义式)} \\ &= E(X^2) - \big(E(X)\big)^2. && \text{(计算公式)} \end{aligned}$$

记 $E(X) = \mu$, 由方差定义式 $D(X) = E\big[(X-\mu)^2\big]$, 可见方差其实是一个期望, 是随机变量函数 $(X-\mu)^2$ 的期望. 由随机变量函数期望的求法, 故有

$$D(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,\mathrm{d}x.$$

方差的性质:

(1) $D(C) = 0$, $D(X+C) = D(X)$.

(2) $D(aX) = a^2 D(X)$, $D(-X) = D(X)$.

(3) $D(X+Y) = D(X) + D(Y) + 2\,\mathrm{Cov}(X, Y)$.
X 与 Y 不相关 $\iff D(X+Y) = D(X) + D(Y)$.
X 与 Y 相互独立 $\Longrightarrow D(X+Y) = D(X) + D(Y)$.

(4) $D(aX \pm bY) = a^2 D(X) + b^2 D(Y)$, 其中 X 与 Y 相互独立.

例 6 设随机变量 X, Y 相互独立, 且都服从均值为 0, 方差为 $\dfrac{1}{2}$ 的正态分布. 求随机变量 $|X-Y|$ 的方差.

解 令 $Z = X - Y$. 由题设知, $Z \sim N(0, 1)$. 对

$$\begin{aligned} D(|X-Y|) = D\big(|Z|\big) &= E\big(|Z|^2\big) - \big[E\big(|Z|\big)\big]^2 \\ &= E(Z^2) - \big[E\big(|Z|\big)\big]^2. \end{aligned}$$

由 $E(Z^2) = D(Z) + \big[E(Z)\big]^2 = 1 + 0 = 1$, 且

$$\begin{aligned} E\big(|Z|\big) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |z| \mathrm{e}^{-z^2/2}\,\mathrm{d}z = \frac{2}{\sqrt{2\pi}} \int_{0}^{+\infty} |z| \mathrm{e}^{-z^2/2}\,\mathrm{d}z = \frac{2}{\sqrt{2\pi}} \int_{0}^{+\infty} z \mathrm{e}^{-z^2/2}\,\mathrm{d}z \\ &= -\frac{2}{\sqrt{2\pi}} \int_{0}^{+\infty} \mathrm{d}\big(\mathrm{e}^{-z^2/2}\big) = \frac{2}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}\Big|_0^{+\infty} \\ &= \sqrt{\frac{2}{\pi}}. \end{aligned}$$

故 $D\big(|X-Y|\big) = E(Z^2) - \big[E\big(|Z|\big)\big]^2 = 1 - \dfrac{2}{\pi}$. ■

4.3 协方差与相关系数

协方差的计算:

$$\begin{aligned}\operatorname{Cov}(X, Y) &= E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big] && \text{(定义式)} \\ &= E(XY) - E(X)E(Y). && \text{(计算公式)}\end{aligned}$$

相关系数的计算:

$$\begin{aligned}\rho_{XY} &= \frac{\operatorname{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \\ &= \frac{E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big]}{\sqrt{D(X)}\sqrt{D(Y)}}.\end{aligned}$$

随机变量的相关系数 = 随机变量 "标准化" 后的协方差. 事实上, X, Y 标准化为

$$X^* = \frac{X - E(X)}{\sqrt{D(X)}}, \qquad Y^* = \frac{Y - E(Y)}{\sqrt{D(Y)}},$$

则

$$\rho_{XY} = \frac{\operatorname{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \operatorname{Cov}(X^*, Y^*).$$

相关系数的性质

- $|\rho| \leqslant 1$. 其中
  (1) $|\rho| = 1 \Longleftrightarrow X$ 与 Y 之间存在线性关系;
  (2) $\rho = 0 \Longleftrightarrow X$ 与 Y 之间不存在线性关系, 或称 X 与 Y **不相关**.
  强调: 不相关是 "不线性相关" 的简称!
- 以下命题是等价的:
  (1) X 与 Y 不相关.
  (2) $\rho_{XY} = 0$.
  (3) $\operatorname{Cov}(X, Y) = 0$.
  (4) $E(XY) = E(X)E(Y)$.
  (5) $D(X + Y) = D(X) + D(Y)$.
- X 与 Y 独立 $\Longrightarrow$ X 与 Y 不相关. 反之不一定成立.

4.4 切比雪夫不等式

$$P\big\{|X - E(X)| \geqslant \varepsilon\big\} \leqslant \frac{D(X)}{\varepsilon^2},$$

或等价地

$$P\big\{|X - E(X)| < \varepsilon\big\} \geqslant 1 - \frac{D(X)}{\varepsilon^2}.$$

5 概率极限定理

主要问题: 大数定律, 中心极限定理.

常见考点: 使用中心极限定理解题.

中心极限定理即言: 大量独立同分布的随机变量之和, 近似服从正态分布.

中心极限定理:

(1) 设随机变量 $X_1, X_2, \cdots, X_n$ 独立同分布, $E(X_k)=\mu$, $D(X_k)=\sigma^2$, $k=1, 2\cdots, n$. 从而,

$$E\big(\sum_{k=1}^{n} X_k\big)=n\mu, \qquad D\big(\sum_{k=1}^{n} X_k\big)=n\sigma^2.$$

则近似地有

$$\sum_{k=1}^{n} X_k \sim N(n\mu, n\sigma^2), \tag{4}$$

进一步 "标准化" 得

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \sim N(0, 1).$$

等价地,

$$\frac{\frac{1}{n}\sum_{k=1}^{n} X_k - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

记 $\overline{X}=\frac{1}{n}\sum_{k=1}^{n} X_k$, 则

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

等价地,

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}). \tag{5}$$

也可以直接由 $E(X_k)=\mu$, $D(X_k)=\sigma^2$, 得 $E(\overline{X})=\mu$, $D(\overline{X})=\frac{\sigma^2}{n}$.

在实际应用中, 公式 (4) 求解和的概率问题; 公式 (5) 求解平均值的概率问题. 这两个公式很重要!

(2) 设 n_A 为 n 重伯努利试验中事件 A 出现的次数, 且 A 在每次实验中发生的概率为 p. 则 n_A 服从二项分布 $B(n, p)$, 从而

$$E(n_A)=np, \qquad D(n_A)=np(1-p).$$

当 n 很大时, n_A 的 "标准化" 变量 $\frac{n_A-E(n_A)}{\sqrt{D(n_A)}}$ 近似服从正态分布, 即

$$\frac{n_A-np}{\sqrt{np(1-p)}} \sim N(0, 1).$$

例 7 设 $X_1, X_2, \cdots$ 为独立同分布随机变量序列, 已知 $E(X_n^k)=a_k$ $(k=1, 2, 3, 4;\ n=1, 2, \cdots)$, 且 $a_4-a_2^2>0$. 试问: 当 n 充分大时, 随机变量 $Y_n=\frac{1}{n}\sum_{i=1}^{n} X_k^2$ 近似服从什么分布? 指出其分布参数.

解 $X_1, X_2, \cdots$ 独立同分布, 则 $X_1^2, X_2^2, \cdots$ 也独立同分布. 由 $E(X_n^k) = a_k$ $(k=1, 2, 3, 4)$, 得

$$E(X_n^2) = a_2,$$
$$D(X_n^2) = E(X_n^4) - \big(E(X_n^2)\big)^2 = a_4 - a_2^2.$$

其中 $n = 1, 2, \cdots$.

又 $a_4 - a_2^2 > 0$, 根据中心极限定理, 则当 n 充分大时, 随机变量 $Y_n = \dfrac{1}{n}\sum\limits_{k=1}^{n} X_k^2$ 近似服从参数为 $\big(a_2, \dfrac{a_4 - a_2^2}{n}\big)$ 的正态分布. ■

例 8 一生产线生产的产品成箱包装, 每箱的重量是随机的. 假设每箱平均重 50 千克, 标准差为 5 千克. 若用最大载重量为 5 吨的汽车承运, 试利用中心极限定理说明每辆车最多可以装多少箱, 才能保障不超载的概率大于 0.977.

解 设所求箱数为 n, 每箱的重量记为 $X_1, X_2, \cdots, X_n$. 由题设可把 $X_1, X_2, \cdots, X_n$ 视为独立同分布随机变量. 又

$$E(X_i) = 50, \qquad D(X_i) = 5^2, \qquad (i = 1, 2, \cdots, n)$$

根据中心极限定理, 有 $\sum\limits_{i=1}^{n} X_i$ 近似服从正态分布 $(n\cdot 50, n\cdot 5^2)$.

问题即求 n 使

$$P\Big\{\sum_{i=1}^{n} X_i \leqslant 5000\Big\} > 0.977.$$

其中

$$P\Big\{\sum_{i=1}^{n} X_i \leqslant 5000\Big\} = P\bigg\{\frac{\sum\limits_{i=1}^{n} X_i - 50n}{5\sqrt{n}} \leqslant \frac{5000 - 50n}{5\sqrt{n}}\bigg\}$$
$$\approx \Phi\Big(\frac{1000 - 10n}{\sqrt{n}}\Big),$$

故

$$\Phi\Big(\frac{1000 - 10n}{\sqrt{n}}\Big) > 0.977 = \Phi(2),$$

即

$$\frac{1000 - 10n}{\sqrt{n}} > 2,$$

从而 $n < 98.0199$, 即最多可以装 98 箱. ■

说明: 教材原题是 0.997. 这里改为 0.977.

另外, P.198 习题 7 的第 (2) 小题, 将 “概率小于 0.90”, 改为 “概率不小于 0.90”.