"十二五"普通高等教育本科国家级规划教材

概率论与数理统计教程

第三版

茆诗松 程依明 濮晓龙 编著

高等教育出版社

"十二五"普通高等教育本科国家级规划教材

概率论与数理统计教程

第三版

茆诗松 程依明 濮晓龙 编著

高等教育出版社·北京

内容提要

本书为“十二五”普通高等教育本科国家级规划教材。全书共八章，前四章为概率论部分，主要叙述各种概率分布及其性质，后四章为数理统计部分，主要叙述各种参数估计与假设检验。本次修订适当补充了数字资源(以符号标识)。

本书的编写从实例出发，图文并茂，通俗易懂，注重讲清楚基本概念与统计思想，强调各种方法的应用，适合初次接触概率统计的读者阅读。全书插图100多幅，例题250多道，习题600余道。

本书可供高等学校数学类专业与统计学专业作为教材使用，亦可供其他专业类似课程参考，也适合自学使用。

图书在版编目(CIP)数据

概率论与数理统计教程/茆诗松，程依明，濮晓龙编著. --3版.--北京：高等教育出版社，2019.11(2020.4重印)

ISBN 978-7-04-051148-2

Ⅰ.①概… Ⅱ.①茆… ②程… ③濮… 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(2019)第009658号

策划编辑	李　蕊	责任编辑	李　蕊	特约编辑	高　旭	封面设计	王凌波
版式设计	马　云	插图绘制	于　博	责任校对	王　雨	责任印制	刘思涵

出版发行	高等教育出版社	网　　址	http://www.hep.edu.cn
社　　址	北京市西城区德外大街4号		http://www.hep.com.cn
邮政编码	100120	网上订购	http://www.hepmall.com.cn
印　　刷	肥城新华印刷有限公司		http://www.hepmall.com
开　　本	787mm×1092mm　1/16		http://www.hepmall.cn
印　　张	30.5	版　　次	2004年10月第1版
字　　数	670千字		2019年11月第3版
购书热线	010-58581118	印　　次	2020年4月第2次印刷
咨询电话	400-810-0598	定　　价	59.00元

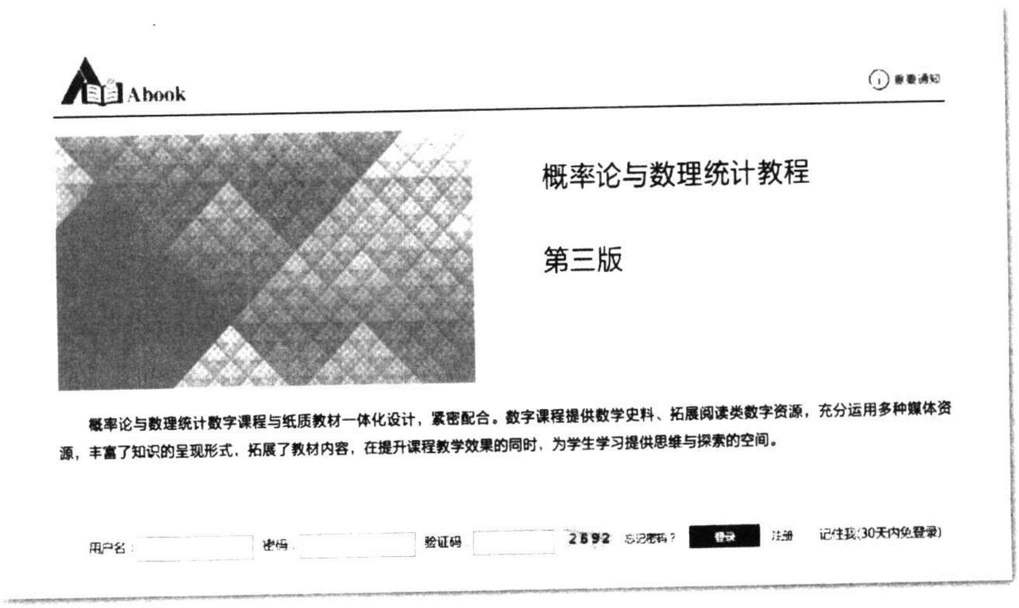
物 料 号　51148-00

概率论与数理统计教程

第三版

茆诗松 程依明 濮晓龙 编著

1. 计算机访问http://abook.hep.com.cn/1225539，或手机扫描二维码、下载并安装Abook应用。
2. 注册并登录，进入“我的课程”。
3. 输入封底数字课程账号（20位密码，刮开涂层可见），或通过Abook应用扫描封底数字课程账号二维码，完成课程绑定。
4. 单击“进入课程”按钮，开始本数字课程的学习。



课程绑定后一年为数字课程使用有效期。受硬件限制，部分内容无法在手机端显示，请按提示通过计算机访问学习。

如有使用问题，请发邮件至abook@hep.com.cn。

扫描二维码
下载Abook应用

概率论简史

数理统计简史

http://abook.hep.com.cn/1225539

目录

出的结果与其他两人不一样，那么由他付账；如果三个人掷出的结果是一样的，那么就重新掷，一直这样下去，直到确定了由谁来付账.求以下事件的概率：

（1）进行到了第 2 轮确定了由谁来付账；

（2）进行了 3 轮还没有确定付账人.

12. 从一个装有 m 个白球、n 个黑球的袋中有返回地摸球，直到摸到白球时停止.试求取出黑球数的期望.

13. 某种产品上的缺陷数 X 服从下列分布列：

$$P(X=k)=\frac{1}{2^{k+1}},\quad k=0,1,\cdots,$$

求此种产品上的平均缺陷数.

14. 设随机变量 X 的密度函数为

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

以 Y 表示对 X 的三次独立重复观察中事件 $\{X\leqslant 1/2\}$ 出现的次数，试求 $P(Y=2)$.

15. 某产品的不合格品率为 0.1，每次随机抽取 10 件进行检验，若发现其中不合格品数多于 1，就去调整设备.若检验员每天检验 4 次，试问每天平均要调整几次设备？

16. 一个系统由多个元件组成，各个元件是否正常工作是相互独立的，且各个元件正常工作的概率为 p.若在系统中至少有一半的元件正常工作，那么整个系统就有效.问 p 取何值时，5 个元件的系统比 3 个元件的系统更有可能有效？

17. 设随机变量 X 服从参数为 λ 的泊松分布，试证明

$$E(X^n)=\lambda E[(X+1)^{n-1}],$$

利用此结果计算 $E(X^3)$.

18. 令 $X(n,p)$ 表示服从二项分布 $b(n,p)$ 的随机变量，试证明：

$$P(X(n,p)\leqslant i)=1-P(X(n,1-p)\leqslant n-i-1).$$

19. 设随机变量 X 服从参数为 p 的几何分布，试证明：

$$E\left(\frac{1}{X}\right)=\frac{-p\ln p}{1-p}$$

20. 设随机变量 $X\sim b(n,p)$，试证明：

$$E\left(\frac{1}{X+1}\right)=\frac{1-(1-p)^{n+1}}{(n+1)p}.$$

§2.5 常用连续分布

在连续分布场合，密度函数与分布函数是可以相互导出的，含有相同信息，各有各的用处，但在图形上密度函数对各种连续分布的特性能得到直观显示.如正态与偏态、单峰与平顶都是依密度函数图形命名的，因而人们对密度函数更为注意.

2.5.1 正态分布

正态分布是概率论与数理统计中最重要的一个分布，高斯（Gauss，1777—1855）在研究误差理论时首先用正态分布来刻画误差的分布，所以正态分布又称为高斯分布.

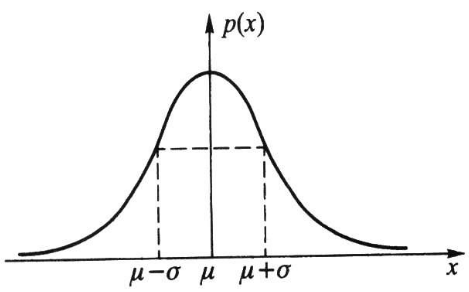
本书第四章的中心极限定理表明:一个随机变量如果是由大量微小的、独立的随机因素的叠加结果,那么这个变量一般都可以认为服从正态分布.因此很多随机变量可以用正态分布描述或近似描述,譬如测量误差、产品重量、人的身高、年降雨量等都可用正态分布描述.

一、正态分布的密度函数和分布函数

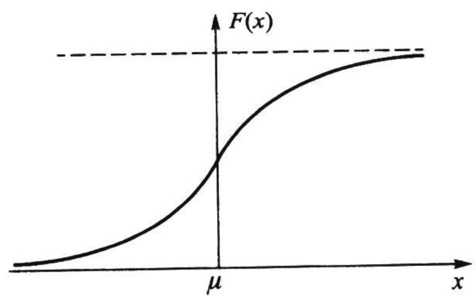
若随机变量 X 的密度函数为

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<\infty,\tag{2.5.1}$$

则称 X 服从**正态分布**,称 X 为**正态变量**,记作 $X\sim N(\mu,\sigma^2)$.其中参数 $-\infty<\mu<\infty$,$\sigma>0$.其密度函数 $p(x)$ 的图形如图 2.5.1(a)所示.$p(x)$ 是一条钟形曲线,中间高、两边低、左右关于 $x=\mu$ 对称,μ 是正态分布的中心,且在 $x=\mu$ 附近取值的可能性大,在两侧取值的可能性小.$\mu\pm\sigma$ 是该曲线的拐点.



(a) 密度函数 $p(x)$



(b) 分布函数 $F(x)$

图 2.5.1 正态分布

正态分布 $N(\mu,\sigma^2)$ 的分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt.\tag{2.5.2}$$

它是一条光滑上升的 S 形曲线,见图 2.5.1(b).

图 2.5.2 给出了在 μ 和 σ 变化时,相应正态密度曲线的变化情况.

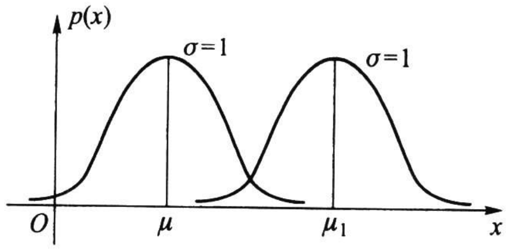
- 从图 2.5.2(a)中可以看出:如果固定 σ,改变 μ 的值,则图形沿 x 轴平移,而不改变其形状.也就是说正态密度函数的位置由参数 μ 所确定,因此亦称 μ 为**位置参数**.
- 从图 2.5.2(b)中可以看出:如果固定 μ,改变 σ 的值,则分布的位置不变,但 σ 愈小,曲线呈高而瘦,分布较为集中;σ 愈大,曲线呈矮而胖,分布较为分散.也就是说正态密度函数的尺度由参数 σ 所确定,因此称 σ 为**尺度参数**.

二、标准正态分布

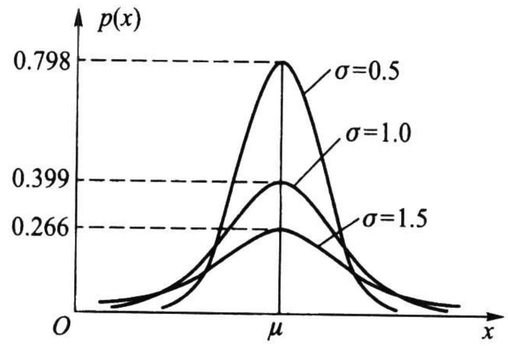
称 $\mu=0,\sigma=1$ 时的正态分布 $N(0,1)$ 为**标准正态分布**.

通常记标准正态变量为 U,记标准正态分布的密度函数为 $\varphi(u)$,分布函数为 $\Phi(u)$,即

$$\varphi(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}},\quad -\infty<u<\infty,$$



(a) σ固定，μ值改变



(b) μ固定，σ值改变

图 2.5.2 正态密度函数

$$\Phi(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t,\quad -\infty<u<\infty.$$

由于标准正态分布的分布函数不含任何未知参数,故其值 $\Phi(u)=P(U\leqslant u)$ 完全可以算出,附表 2 对 $u\geqslant 0$ 给出了 $\Phi(u)$ 的值.对于 $\Phi(u)$ 有

- $\Phi(-u)=1-\Phi(u)$.
- $P(U>u)=1-\Phi(u)$.
- $P(a<U<b)=\Phi(b)-\Phi(a)$.
- $P(|U|<c)=2\Phi(c)-1\ (c\geqslant 0)$.

这些等式都不难推得.

例 2.5.1 设 $U\sim N(0,1)$,利用附表 2,求下列事件的概率:

(1) $P(U<1.52)=\Phi(1.52)=0.935\ 7$.

(2) $P(U>1.52)=1-\Phi(1.52)=1-0.935\ 7=0.064\ 3$.

(3) $P(U<-1.52)=1-\Phi(1.52)=0.064\ 3$.

(4)
$$\begin{aligned}P(-0.75\leqslant U\leqslant 1.52)&=\Phi(1.52)-\Phi(-0.75)\\&=\Phi(1.52)-[1-\Phi(0.75)]\\&=0.935\ 7-1+0.773\ 4=0.709\ 1.\end{aligned}$$

(5) $P(|U|\leqslant 1.52)=2\Phi(1.52)-1=2\times 0.935\ 7-1=0.871\ 4$.

三、正态变量的标准化

正态分布有一个家族

$$\mathscr{P}=\{N(\mu,\sigma^2):-\infty<\mu<\infty,\sigma>0\},$$

标准正态分布 $N(0,1)$ 是其一个中心成员.以下定理说明:一般正态变量都可以通过一个线性变换(标准化)化成标准正态变量.因此与正态变量有关的一切事件的概率都可通过查标准正态分布函数表获得.可见标准正态分布 $N(0,1)$ 对一般正态分布 $N(\mu,$

σ^2)的计算起着关键的作用.

定理 2.5.1 若随机变量 $X \sim N(\mu,\sigma^2)$,则 $U=(X-\mu)/\sigma \sim N(0,1)$.

证明 记 X 与 U 的分布函数分别为 $F_X(x)$ 与 $F_U(u)$,则由分布函数的定义知

$$F_U(u)=P(U\leqslant u)=P\left(\frac{X-\mu}{\sigma}\leqslant u\right)=P(X\leqslant\mu+\sigma u)=F_X(\mu+\sigma u).$$

由于正态分布函数是严格单调增函数,且处处可导,因此若记 X 与 U 的密度函数分别为 $p_X(x)$ 与 $p_U(u)$,则有

$$p_U(u)=\frac{\mathrm{d}}{\mathrm{d}u}F_X(\mu+\sigma u)=p_X(\mu+\sigma u)\cdot\sigma=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-u^2/2},$$

由此得

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

由以上定理,我们马上可以得到一些在实际中有用的计算公式,若随机变量 $X\sim N(\mu,\sigma^2)$,则

$$P(X\leqslant c)=\Phi\left(\frac{c-\mu}{\sigma}\right). \tag{2.5.3}$$

$$P(a<X\leqslant b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right). \tag{2.5.4}$$

例 2.5.2 设随机变量 X 服从正态分布 $N(108,3^2)$,试求:

(1) $P(102<X<117)$;

(2) 常数 a,使得 $P(X<a)=0.95$.

解 利用公式(2.5.4)及查附表 2 得

(1)

$$\begin{aligned}P(102<X<117)&=\Phi\left(\frac{117-108}{3}\right)-\Phi\left(\frac{102-108}{3}\right)\\&=\Phi(3)-\Phi(-2)=\Phi(3)+\Phi(2)-1\\&=0.998\ 7+0.977\ 2-1=0.975\ 9.\end{aligned}$$

(2) 由

$$P(X<a)=\Phi\left(\frac{a-108}{3}\right)=0.95,\quad 或 \quad \Phi^{-1}(0.95)=\frac{a-108}{3},$$

其中 Φ^{-1} 为 Φ 的反函数.从附表 2 由里向外反查得

$$\Phi(1.64)=0.949\ 5,\qquad \Phi(1.65)=0.950\ 5,$$

再用线性内插法可得 $\Phi(1.645)=0.95$,即 $\Phi^{-1}(0.95)=1.645$,故

$$\frac{a-108}{3}=1.645,$$

从中解得 $a=112.935$.

从上例我们可以看出,有些场合下给定 $\Phi(x)$ 的值 p,可以从附表 2 中由里向外反查表来得到 x_p,使 $\Phi(x_p)=p$ 或 $\Phi^{-1}(p)=x_p$,这时 x_p 称为标准正态分布的 p 分位数.在上例中 1.645 就是标准正态分布的 0.95 分位数,更一般叙述见2.7.3节.分位数在统计

中被大量使用.

例 2.5.3 在考试中,如果考生的成绩 X 近似地服从正态分布,则通常认为这次考试(就合理地划分考生成绩的等级而言)是正常的.教师经常把分数超过 $\mu+\sigma$ 的评为 A 等,分数在 μ 到 $\mu+\sigma$ 之间的评为 B 等,分数在 $\mu-\sigma$ 到 μ 之间的评为 C 等,分数在 $\mu-2\sigma$ 到 $\mu-\sigma$ 之间的评为 D 等,分数在 $\mu-2\sigma$ 以下的评为 F 等.由此可计算得:

$$P(X\geqslant \mu+\sigma)=P\left(\frac{X-\mu}{\sigma}\geqslant 1\right)=1-\Phi(1)\approx 0.158\ 7,$$

$$P(\mu\leqslant X<\mu+\sigma)=P\left(0\leqslant\frac{X-\mu}{\sigma}<1\right)=\Phi(1)-\Phi(0)\approx 0.341\ 3,$$

$$P(\mu-\sigma\leqslant X<\mu)=P\left(-1\leqslant\frac{X-\mu}{\sigma}<0\right)=\Phi(0)-\Phi(-1)\approx 0.341\ 3,$$

$$P(\mu-2\sigma\leqslant X<\mu-\sigma)=P\left(-2\leqslant\frac{X-\mu}{\sigma}<-1\right)=\Phi(-1)-\Phi(-2)\approx 0.135\ 9,$$

$$P(X<\mu-2\sigma)=P\left(\frac{X-\mu}{\sigma}<-2\right)=\Phi(-2)\approx 0.022\ 8.$$

这说明:用这种方法划分成绩的等级,获得 A 等的约占 16%,B 等的约占 34%,C 等的约占 34%,D 等的约占 14%,F 等的约占 2%.

四、正态分布的数学期望与方差

设随机变量 $X\sim N(\mu,\sigma^2)$,由于 $U=(X-\mu)/\sigma\sim N(0,1)$,所以 U 的数学期望为

$$E(U)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}u\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u,$$

注意到上述积分的被积函数为一个奇函数,所以其积分值等于 0,即 $E(U)=0$.又因为 $X=\mu+\sigma U$,所以由数学期望的性质得

$$E(X)=\mu+\sigma\times 0=\mu.$$

也就是说,正态分布 $N(\mu,\sigma^2)$ 中的 μ 为数学期望.

又因为

$$\mathrm{Var}(U)=E(U^2)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}u^2\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}u\mathrm{d}\left(-\mathrm{e}^{-\frac{u^2}{2}}\right)$$

$$=\frac{1}{\sqrt{2\pi}}\left(-u\mathrm{e}^{-\frac{u^2}{2}}\Big|_{-\infty}^{\infty}+\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u=\frac{1}{\sqrt{2\pi}}\sqrt{2\pi}=1,$$

且 $X=\mu+\sigma U$,所以由方差的性质得

$$\mathrm{Var}(X)=\mathrm{Var}(\mu+\sigma U)=\sigma^2.$$

这说明,正态分布 $N(\mu,\sigma^2)$ 中另一个参数 σ^2 就是 X 的方差.

在求正态分布的数学期望和方差过程中,用到了一种变换:令 $U=(X-\mu)/\sigma$,则

$E(U)=0, \mathrm{Var}(U)=1$.这个变换在很多场合也可使用,也就是对任意非正态随机变量 X,如果 X 的数学期望 $E(X)$ 和方差 $\mathrm{Var}(X)$ 存在,则称

$$X^{*}=\frac{X-E(X)}{\sqrt{\mathrm{Var}(X)}}$$

为 X 的**标准化随机变量**,且可得

$$E(X^{*})=0, \qquad \mathrm{Var}(X^{*})=1.$$

五、正态分布的 3σ 原则

设随机变量 $X \sim N(\mu,\sigma^2)$,则

$$P(\mu-k\sigma<X<\mu+k\sigma)=P\left(\left|\frac{X-\mu}{\sigma}\right|<k\right)=\Phi(k)-\Phi(-k)=2\Phi(k)-1$$

当 $k=1,2,3$ 时,有

$$\begin{aligned} &P(\mu-\sigma<X<\mu+\sigma)=2\Phi(1)-1=0.682\ 6, \\ &P(\mu-2\sigma<X<\mu+2\sigma)=2\Phi(2)-1=0.954\ 5, \\ &P(\mu-3\sigma<X<\mu+3\sigma)=2\Phi(3)-1=0.997\ 3. \end{aligned} \tag{2.5.5}$$

这是正态分布的重要性质.假如某随机变量取值的概率近似满足(2.5.5),则可认为这个随机变量近似服从正态分布;假如(2.5.5)三式中有一个偏差较大,则可以认为这个随机变量不服从正态分布.这就是正态分布的 3σ 原则,这个原则在 X 的观察值较多(成百上千个)时,常用于判断 X 的分布是否近似服从正态分布.

在生产中某产品的质量要求常规定其上、下控制限,若上、下控制限能覆盖区间 $(\mu-3\sigma,\mu+3\sigma)$,则称该生产过程受控制,并称其比值

$$C_p=\frac{\text{上控制限}-\text{下控制限}}{6\sigma}$$

为过程能力指数.当 $C_p<1$ 时,认为生产过程不足;当 $C_p\geqslant 1.33$ 时,认为生产过程正常;当 C_p 为其他值时,常认为生产过程不稳定,需要改进.

2.5.2　均匀分布

一、均匀分布的密度函数和分布函数

前面曾以例子形式说明过均匀分布,这里给出一般的叙述:若随机变量 X 的密度函数(见图 2.5.3(a))为

$$p(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b, \\ 0, & \text{其他}.\end{cases} \tag{2.5.6}$$

则称 X 服从区间 (a,b) 上的**均匀分布**,记作 $X\sim U(a,b)$,其分布函数(见图2.5.3(b))为