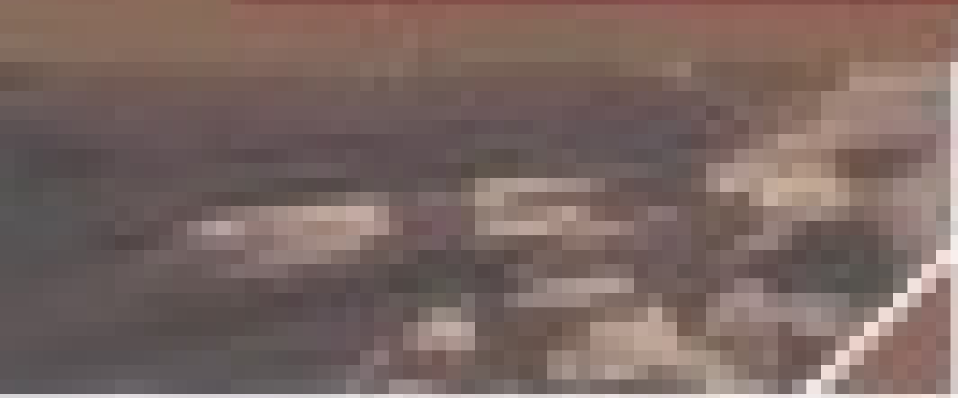
高职高专资源勘查专业规划教材

地学数据常用数理统计

姜启明　鲁挑建◎编著

哈尔滨工程大学出版社
Harbin Engineering University Press

地学数据常用数理统计



高职高专资源勘查专业规划教材

地学数据常用数理统计

姜启明　鲁挑建　编著

哈尔滨工程大学出版社

内容简介

本书系统地叙述了地学数据处理常用的统计和分析方法。全书共分9章，并有附录。内容包括数据处理的基本理论、数据的类型和分布形式、数据的各类变换方法、物化探场的基本理论和求解方法、物化探数据内部信息的挖掘等。

本书插图较多，叙述通俗易懂，理论联系实际。书中编入了很多例题，每章后都附有习题，每个习题都对应于相应的例题，便于对照例题进行练习；书末还附有习题答案，便于自学之用。本书注重基本理论的应用，多处强调了由数据得出的地质结论。

本书可作为高等院校的物探专业教材，也可供从事矿产勘查的地质工作者参考。

图书在版编目(CIP)数据

地学数据常用数理统计/姜启明，鲁挑建编著. —哈尔滨：哈尔滨工程大学出版社，2010. 1
ISBN 978-7-81133-595-8

Ⅰ. ①地… Ⅱ. ①姜…②鲁… Ⅲ. ①地球科学-数据处理-高等学校：技术学校-教材 Ⅳ. ①P-37

中国版本图书馆 CIP 数据核字(2010)第007431号

出版发行	哈尔滨工程大学出版社
社　　址	哈尔滨市南岗区东大直街124号
邮政编码	150001
发行电话	0451-82519328
传　　真	0451-82519699
经　　销	新华书店
印　　刷	黑龙江省地质测绘印制中心印刷厂
开　　本	787mm×1 092mm　1/16
印　　张	12.75
字　　数	312千字
版　　次	2010年1月第1版
印　　次	2010年1月第1次印刷
定　　价	25.00元

http://press.hrbeu.edu.cn
E-mail:heupress@hrbeu.edu.cn

近几年，随着我国经济建设的发展，对矿产资源的需求量猛增，地质行业又迎来了难得的发展机遇。各类院校相继开设资源勘查类专业，各种与资源勘查相关的职业教育也得到了迅猛的发展。由于历史原因，地学高职高专教育起步较晚，使其发展受到很大限制，其中教材奇缺直接影响到教学质量。本书就是在这种形势下应运而生的。

随着新的找矿技术、方法和仪器的应用，地表能够找到的矿体几乎都已找到，而通过地表工作直接找到矿体的难度变得越来越大。这就需要对现有的地学测量数据（这些数据包含大量找矿信息）进行统计分析，从中找出有用信息。筛选有用信息必须运用一定的数学方法，故本书前半部分（第1～4章）主要是地学数据的统计整理；后半部分（第5～9章）主要是通过数学方法，从地学数据中筛选出有用信息，称之为数学分析。

本书的地学数据主要是指物化探数据，此外还包括岩（矿）石测试数据、矿产开采工艺数据等与地学密切相关的数据。地学数据来源于复杂多变的地质总体（土壤、岩石），而地质总体经过漫长而复杂的地质作用，其中大量的有用信息被掩盖。对于资源勘查类专业的学生来说，将来走上工作岗位以后，很可能要接触到这类数据，面对大量纷繁复杂的实测数据，既要认清数据的复杂性，又不能掉入数据的海洋而不知所措，这就需要对数据进行仔细的统计分析，从大量的数据中筛选出有用信息，找出规律性的东西，深入挖掘找矿信息，使现有信息得到充分利用，进一步指导找矿，这就是本书编写的目的。

本书可供高职高专资源勘查类专业教学使用，亦可作为从事资源勘查工作人员的参考书或工具书。

本书是在作者近20年野外实践经验的基础上编写而成的，编写原则是：学以致用，遵循必须、够用为度的原则。为了适应具备高职高专文化程度的人员的需要，本书对大量的数学原理和不常用的数据处理方法进行了删减（比如多元统计分析方法等），突出公式的应用性，并编写了大量的例题，每章后附有习题，书末还附有习题答案。通过例题的学习可以加深对公式的理解，通过作业可以训练学生的实际演算能力，逐步培养学生的数学应用能力和动手操作能力，使学生逐步养成认真细致、一丝不苟的工作作风。

书中的例题和习题大部分都是作者近年来的野外实测原始资料（多数资料尚未发表），把这些资料编入本书，可以使本书的例题更贴近于野外工作实际，也可以增加学生的学习兴趣。书中对于地学类专业经常遇到或可能遇到的数据处理方法作了必要汇总，所使用的数学方法也都是最常用的、较为简单的。例题和习题从生产实践的需要出发，讲解数据统计、分析方法，所

以实用性很强。

本书对《工程数学》中的《概率论与数理统计》《线性代数》等知识要求不高,没有《工程数学》知识的人也能看懂。对牵涉到《高等数学》和《工程数学》方面的问题作了适当的补充,如第1章内容和第8章前半部分。教师在授课时可作适当删节,删节后不会影响以后知识的学习。

本书预计教授75学时左右。由于学时数量有限,故本书在编写中对公式应用的步骤进行了详细的阐述(如第8章后半部分),这样就可以保证数学底子较薄的学生通过课后自学达到掌握知识的目的。

本书的主要参考资料是张锦由、黎春华编写的《铀矿物化探数据处理方法》,在编写过程中得到张锦由教授的悉心指导,同时也参考或引用了一些国内学者的论著,在此一并致谢。

核地质二〇三研究所研究员级高级工程师张云宜老师审阅了全部书稿,并提出了很多宝贵的修改意见,在此表示诚挚的谢意。

本书第7~9章由鲁挑建编写,其余章节由姜启明编写。

由于编写时间仓促,书中难免出现错漏之处,恳请读者批评指正。

编者

2009年10月

CONTENTS 目录

CONTENTS

第 1 章　数理统计的数学基础

1.1　Γ 函数及其应用

1.1.1　Γ 函数

在学习数理统计时，首先要学习一个数理统计中最常用的函数——Γ 函数。

Γ 函数的定义式为

$$\Gamma(s)=\int_0^{+\infty}e^{-x}x^{s-1}dx,s>0 \tag{1-1}$$

这是一个广义的收敛积分。它有三条重要性质。

1. 递推公式 $\Gamma(s+1)=s\Gamma(s)$，当 $s>0$ 时成立

证　因为

$$\Gamma(s+1)=\int_0^{+\infty}e^{-x}x^s dx=\lim_{b\to+\infty}\lim_{\varepsilon\to+0}\int_\varepsilon^b e^{-x}x^s dx$$

应用分部积分法，设 $u=x^s$，$dv=e^{-x}dx$，则 $du=sx^{s-1}dx$，$v=-e^{-x}$，所以

$$\int_\varepsilon^b e^{-x}x^s dx=[-e^{-x}x^s]_\varepsilon^b+s\int_\varepsilon^b e^{-x}x^{s-1}dx$$

而

$$\lim_{b\to+\infty}\lim_{\varepsilon\to+0}[e^{-x}x^s]_\varepsilon^b=\lim_{b\to+\infty}\lim_{\varepsilon\to+0}[e^{-b}b^s-e^{-\varepsilon}\varepsilon^s]=\lim_{b\to+\infty}[e^{-b}b^s]$$

把上式右端写成分式，并多次使用洛必达法则，直至分子变为 $s!\ b^0$，而分母不变，故上式右端极限为 0，所以

$$\Gamma(s+1)=\lim_{b\to+\infty}\lim_{\varepsilon\to+0}s\int_\varepsilon^b e^{-x}x^{s-1}dx=s\int_0^{+\infty}e^{-x}x^{s-1}dx=s\Gamma(s)$$

显然

$$\Gamma(1)=\int_0^{+\infty}e^{-x}dx=[-e^{-x}]_0^{+\infty}=-[0-1]=1$$

反复运用递推公式，便有

$$\begin{aligned}\Gamma(2)&=1\Gamma(1)&&=1\\\Gamma(3)&=2\Gamma(2)&&=2!\\\Gamma(4)&=3\Gamma(3)&&=3!\\\Gamma(5)&=4\Gamma(4)&&=4!\\\vdots&\quad\ \vdots&&\quad\vdots\end{aligned}$$

一般地，对于任何正整数 n，有

$$\Gamma(n+1)=n! \tag{1-2}$$

故 Γ 函数可以看成是阶乘的推广。

2. 当 $s\to+0$ 时，$\Gamma(s)\to+\infty$

证　因为

$$\Gamma(s)=\frac{\Gamma(s+1)}{s},\Gamma(1)=1$$

所以当 $s\to+0$ 时,$\Gamma(s)\to+\infty$。

3. 当 $s=1/2$ 时

$$\Gamma\left(\frac{1}{2}\right)=2\int_0^{+\infty}\mathrm{e}^{-u^2}\mathrm{d}u=\sqrt{\pi} \tag{1-3}$$

这一性质证明时用到的数学知识较多,在此我们不作证明。

综合以上三条性质,结合高等数学知识,可作出 Γ 函数的图像,如图 1-1 所示。

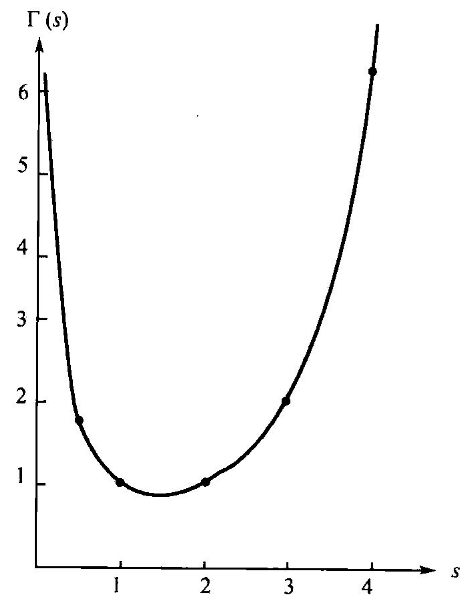


图 1-1　Γ 函数的图像

利用 Γ 函数这三条性质,可以证明一些数学公式。

例 1-1　利用 Γ 函数的性质证明

$$\Gamma\left(\frac{2k+1}{2}\right)=\frac{1\cdot3\cdot5\cdot\cdots\cdot(2k-1)\sqrt{\pi}}{2^k}$$

式中,k 为自然数。

证　这个公式可以利用中学学过的数学归纳法来证明。

(1) 当 $k=1$ 时,等式左端 $=\Gamma\left(\frac{3}{2}\right)=\Gamma\left(\frac{1}{2}+1\right)=\frac{1}{2}\Gamma\left(\frac{1}{2}\right)=\frac{1}{2}\sqrt{\pi}=$ 等式右端,等式成立。

(2) 当 $k=2$ 时,等式左端 $=\Gamma\left(\frac{5}{2}\right)=\Gamma\left(\frac{3}{2}+1\right)=\frac{3}{2}\Gamma\left(\frac{3}{2}\right)=\frac{3}{2}\cdot\frac{1}{2}\sqrt{\pi}=\frac{1\cdot3}{2^2}\sqrt{\pi}=$ 等式右端,等式成立。

(3) 设 $k=m$ 时等式成立,即

$$\Gamma\left(\frac{2m+1}{2}\right)=\frac{1\cdot3\cdot5\cdot\cdots\cdot(2m-1)\sqrt{\pi}}{2^m}$$

当 $k=m+1$ 时

$$\begin{aligned}\Gamma\left[\frac{2(m+1)+1}{2}\right]&=\Gamma\left(\frac{2m+1}{2}+1\right)=\frac{2m+1}{2}\Gamma\left(\frac{2m+1}{2}\right)\\&=\frac{2m+1}{2}\cdot\frac{1\cdot3\cdot5\cdot\cdots\cdot(2m-1)\sqrt{\pi}}{2^m}\\&=\frac{1\cdot3\cdot5\cdot\cdots\cdot(2m-1)(2m+1)\sqrt{\pi}}{2^{m+1}}\end{aligned}$$

即 k 为自然数时,$\Gamma\left(\frac{2k+1}{2}\right)=\frac{1\cdot3\cdot5\cdot\cdots\cdot(2k-1)\sqrt{\pi}}{2^k}$,公式得证。

1.1.2　Γ 函数的应用

这里只讲 Γ 函数在证明正态分布的平均数和均方差的数学期望中的应用,Γ 函数在抽

样分布中的应用将在下一节中讲述。

设 X 服从正态分布，则概率密度为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \sigma > 0, -\infty < x < +\infty \tag{1-4}$$

1. X 的数学期望

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

令 $\frac{x-\mu}{\sigma} = u$，则 $dx = \sigma du$，得

$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma u + \mu) e^{\frac{-u^2}{2}} du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma u e^{\frac{-u^2}{2}} du + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mu e^{\frac{-u^2}{2}} du \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} -e^{\frac{-u^2}{2}} d\left(\frac{-u^2}{2}\right) + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{-u^2}{2}} du \\ &= \frac{\sigma}{\sqrt{2\pi}} \left[-e^{\frac{-u^2}{2}} \right]_{-\infty}^{+\infty} + \frac{2\mu}{\sqrt{2\pi}} \int_{0}^{+\infty} e^{\frac{-u^2}{2}} du \end{aligned}$$

上式第一项为0，只需求出第二项的值。为此，再设 $t^2 = \frac{u^2}{2}$，则 $u = \sqrt{2}t$，$du = \sqrt{2}dt$，得

$$\begin{aligned} E(X) &= \frac{\sqrt{2}\mu}{\sqrt{\pi}} \int_{0}^{+\infty} e^{-t^2} \sqrt{2} dt \\ &= \frac{2\mu}{\sqrt{\pi}} \int_{0}^{+\infty} e^{-t^2} dt = \frac{\mu}{\sqrt{\pi}} \cdot \Gamma\left(\frac{1}{2}\right) = \frac{\mu}{\sqrt{\pi}} \cdot \sqrt{\pi} = \mu \end{aligned}$$

可见，正态概率密度中的 μ 就是 X 的数学期望。

注意变换 $\frac{x-\mu}{\sigma} = u$，这种变换称为标准化变换；有的书中把 u 写作 z，但其意义是相同的；还有的书中把这种变换直接称为 u 变换。标准化变换在数理统计中非常重要，其应用将在以后章节中详细讲述。

2. X 方差的数学期望

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

令 $\frac{x-\mu}{\sigma} = u$，则 $dx = \sigma du$，得

$$D(X) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^2 e^{\frac{-u^2}{2}} du = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} -u e^{\frac{-u^2}{2}} d\left(\frac{-u^2}{2}\right)$$

对上式采用分部积分法，得

$$D(X) = \frac{\sigma^2}{\sqrt{2\pi}} \left[-u e^{\frac{-u^2}{2}} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{\frac{-u^2}{2}} du \right]$$

对上式方括号内的第一项采用洛必达法则求极限，可证得其为0，则

$$D(X) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{-u^2}{2}} du$$

再次采用换元积分法，设 $t^2=\frac{u^2}{2}$，则 $u=\sqrt{2}t, \mathrm{d}u=\sqrt{2}\mathrm{d}t$，得

$$D(X) = \frac{2\sigma^2}{\sqrt{2\pi}}\int_0^{+\infty} e^{-t^2}\sqrt{2}\mathrm{d}t = \frac{2\sigma^2}{\sqrt{\pi}}\int_0^{+\infty} e^{-t^2}\mathrm{d}t = \frac{\sigma^2}{\sqrt{\pi}}\cdot\sqrt{\pi} = \sigma^2$$

可见，正态概率密度中的 σ 就是 X 方差的数学期望。

这就是说正态随机变量的概率密度中的两个参数 μ 和 σ 分别就是该随机变量的数学期望和均方差。因而，正态随机变量的分布完全可由它的数学期望和方差确定。

对任意正态分布的自变量 X 进行标准化变换后可得标准正态分布，此时 X 的数学期望为 $\mu=0$，X 方差的数学期望为 $\sigma^2=1$，故通常把标准正态分布写作 $N(0,1)$，亦称 0－1 分布。每个经标准化变换后的 X(即 X')对应的概率值已列入附录一，使用时可直接查表。

下面还要指出正态随机变量的三个重要数据

$$P\{\mu-\sigma < X < \mu+\sigma\} = \Phi(1)-\Phi(-1) = 0.6826 \tag{1-5}$$

$$P\{\mu-2\sigma < X < \mu+2\sigma\} = \Phi(2)-\Phi(-2) = 0.9544 \tag{1-6}$$

$$P\{\mu-3\sigma < X < \mu+3\sigma\} = \Phi(3)-\Phi(-3) = 0.9974 \tag{1-7}$$

上述三个公式中的 $\Phi(1)$，$\Phi(-1)$，$\Phi(2)$ 等，分别表示经标准化变换后的 $u=1$，$u=-1$，$u=2$ 时的概率值。

如图 1－2 所示，图中画出了 $\mu=0,\sigma=1$ 时的正态概率密度曲线，即标准正态分布概率的情况。图中(1)与概率密度曲线围成的面积，表示在 －0.674 5～0.674 5 的概率为50%，即(1)的面积是概率密度曲线与横轴围成的面积的一半；图中(2)与概率密度曲线围成的面积占 68.26%；图中(3)与概率密度曲线围成的面积占 95.44%；图中(4)与概率密度曲线围成的面积占 99.74%。由此可见，对于正态随机变量来说，它落在 $\mu-3\sigma$～$\mu+3\sigma$ 内的概率几乎是肯定的事件，而落在此区间以外的概率约为 0.26%，几乎成为不可能事件。这一规则被称为“3σ 规则”，在工程上被广泛使用，其理论基础就在于此。

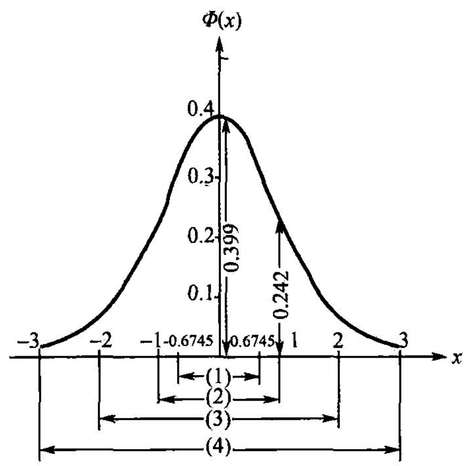


图 1－2　标准正态分布概率密度曲线

1.1.3　多元统计量的联合分布

本节先讨论两个随机变量的函数分布，然后进行推广。

1. $Z=X+Y$ 的分布

设 X 和 Y 是两个相互独立的随机变量，且都具有 $N(0,1)$ 分布，则 X 和 Y 的概率密度函数分别为

$$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}},\ -\infty < x < +\infty \tag{1-8}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}}e^{\frac{-y^2}{2}},\ -\infty < y < +\infty \tag{1-9}$$

令 $z=x+y$，则 $y=z-x$；变量 $Z=X+Y$ 的概率密度函数为

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) \mathrm{d}x = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{e}^{\frac{-x^2}{2}} \cdot \mathrm{e}^{\frac{-(z-x)^2}{2}} \mathrm{d}x = \frac{1}{2\pi}\mathrm{e}^{\frac{-z^2}{4}}\int_{-\infty}^{+\infty} \mathrm{e}^{-\left(x-\frac{z}{2}\right)^2} \mathrm{d}x$$

再令 $t = x - \frac{z}{2}$，得

$$f_Z(z) = \frac{1}{2\pi}\mathrm{e}^{\frac{-z^2}{4}}\int_{-\infty}^{+\infty} \mathrm{e}^{-t^2}\mathrm{d}t = \frac{1}{2\pi}\mathrm{e}^{\frac{-z^2}{4}}\sqrt{\pi} = \frac{1}{2\sqrt{\pi}}\mathrm{e}^{\frac{-z^2}{4}}$$

即 Z 具有 $N(0,2)$ 分布。

一般地，设 X,Y 相互独立，且 $X \sim N(a_1,\sigma_1^2)$，$Y \sim N(a_2,\sigma_2^2)$，则 $Z = X + Y$ 仍然具有正态分布，且 $Z \sim N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$。

这个结论还能推广到 n 个随机变量的和的情况，即若 $X_k \sim (a_k, \sigma_k^2)(k = 1,2,\cdots,n)$，且它们相互独立，则它们的和 $Z = X_1 + X_2 + \cdots + X_n$ 仍然服从正态分布，且 $Z \sim N(a_1 + a_2 + \cdots + a_k, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_k^2)$。还可以证明，有限个正态随机变量的线性组合仍然服从正态分布，这些结论在以后的 χ^2 检验中将有很大用处。

2. $Z = \sqrt{X^2 + Y^2}$ 的分布

我们只讨论下面的特殊情况。

设 X,Y 是两个相互独立的随机变量，且都具有 $N(0,\sigma^2)$ 分布，即

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2\sigma^2}}, -\infty < x < +\infty \tag{1-10}$$

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{\frac{-y^2}{2\sigma^2}}, -\infty < y < +\infty \tag{1-11}$$

现在来求 $Z = \sqrt{X^2 + Y^2}$ 的分布函数和概率密度函数。

先求 Z 的分布函数。由于 X,Y 相互独立，故二维随机变量 (X,Y) 的概率密度函数为

$$f(x,y) = f_X(x)f_Y(y) = \frac{1}{2\pi\sigma^2}\mathrm{e}^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{1-12}$$

因为 Z 是非负的，故 $Z<0$ 时其分布函数 $F_Z(z) = 0$；当 $Z>0$ 时

$$F_Z(z) = P\{Z \leqslant z\} = \iint\limits_{\sqrt{x^2+y^2}<z} \frac{1}{2\pi\sigma^2}\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}\mathrm{d}x\mathrm{d}y$$

区域 $\sqrt{x^2+y^2} \leqslant z$ 是以原点为中心，z 为半径的圆域。采用极坐标来计算上述二重积分，做变换 $x = r\cos\theta, y = r\sin\theta$，得

$$F_Z(z) = \frac{1}{2\pi\sigma^2}\int_0^{2\pi}\int_0^z \mathrm{e}^{-\frac{r^2}{2\sigma^2}} r\mathrm{d}r\mathrm{d}\theta = \frac{1}{2\pi\sigma^2}\int_0^{2\pi}\left[\int_0^z(-\sigma^2)\mathrm{e}^{-\frac{r^2}{2\sigma^2}}\mathrm{d}\left(-\frac{r^2}{2\sigma^2}\right)\right]\mathrm{d}\theta$$

$$= \frac{1}{2\pi\sigma^2}\int_0^{2\pi}(-\sigma^2)\left[\mathrm{e}^{-\frac{r^2}{2\sigma^2}}\right]_0^z\mathrm{d}\theta = \frac{1}{2\pi}\int_0^{2\pi} -\left(\mathrm{e}^{\frac{-z^2}{2\sigma^2}} - 1\right)\mathrm{d}\theta = 1 - \mathrm{e}^{-\frac{z^2}{2\sigma^2}}$$

于是 Z 的分布函数为

$$F_Z(z) = \begin{cases} 1 - \mathrm{e}^{-\frac{z^2}{2\sigma^2}}, z \geqslant 0 \\ 0, \text{其他} \end{cases}$$

Z 的概率密度函数就是对 Z 的分布函数求一阶导数，即

$$f_Z(z) = \begin{cases} \dfrac{z}{\sigma^2}\mathrm{e}^{-\frac{z^2}{2\sigma^2}}, z \geqslant 0 \\ 0, \text{其他} \end{cases}$$

$f_Z(z)$的图像如图1-3所示，图中分别画出了$\sigma=\sigma_1$，$\sigma=\sigma_2$时的曲线，我们称Z服从参数为$\sigma(\sigma>0)$的瑞利(Rayleigh)分布。

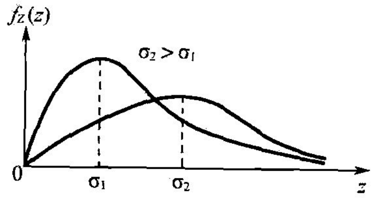


图1-3 瑞利分布概率密度曲线

1.2 抽样分布

统计量的分布又称抽样分布。

1.2.1 正态分布的统计量

设$X\sim N(\mu,\sigma^2)$，$X_1,X_2,\cdots,X_n$是它的一个样本，则$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$也服从正态分布，并且它的数学期望和方差分别是

$$E(\overline{X})=\mu$$

$$D(\overline{X})=\frac{\sigma^2}{n}$$

亦即

$$\overline{X}\sim N(\mu,\frac{\sigma^2}{n}) \tag{1-13}$$

注意：$E(\overline{X})$与总体平均数μ相等，但$D(\overline{X})$只等于总体方差σ^2的n分之一。也就是说，n越大，$\overline{X}$越向总体平均数μ集中。

1.2.2 χ^2分布

设$X\sim N(0,1)$，$X_1,X_2,\cdots,X_n$是它的一个样本，它们的平方和记作χ^2，即

$$\chi^2=X_1^2+X_2^2+\cdots+X_n^2 \tag{1-14}$$

我们称χ^2为服从参数为n的χ^2分布，记作$\chi^2\sim\chi^2(n)$，其中n称为自由度。

$\chi^2(n)$分布的概率密度函数为

$$f(y)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}y^{\frac{n}{2}-1}e^{\frac{-y}{2}},y\geqslant 0\\0,y<0\end{cases} \tag{1-15}$$

这个公式我们不作证明，但有几种常见情况需说明一下。

当$n=1$时，$f(y)=\dfrac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)}y^{\frac{-1}{2}}e^{-\frac{y}{2}}=\dfrac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}e^{-\frac{y}{2}}=\dfrac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}}$，从高等数学中函数的单调性可知，随着$y$的增大，$f(y)$单调减小，且很快衰减；甚至比单纯的指数规律衰减得更快。

当$n=2$时，$f(y)=\dfrac{1}{2\Gamma(1)}y^0e^{-\frac{y}{2}}=\dfrac{1}{2}e^{-\frac{y}{2}}$，可见，随着$y$的增大，$f(y)$呈指数规律衰减，这与指数分布的概率密度公式很相似。

当$n=4$时，$f(y)=\dfrac{1}{4}ye^{-\frac{y}{2}}$，经高等数学讨论可见，随着$y$的增大，$f(y)$先增大到某个数值后，又开始呈指数规律衰减。函数图像与后面讲的泊松分布$\lambda=2\sim3$时的图像很相似。

当 n 很大时，$f(y)$ 的图像与正态分布的概率密度函数图像亦很相似。

$f(y)$ 的图像如图 1－4 所示。

对于给定的正数 α，$0<\alpha<1$，称满足条件

$$\int_{\chi_\alpha^2(n)}^{+\infty} f(y)\,\mathrm{d}y = \alpha \tag{1-16}$$

的点 $\chi_\alpha^2(n)$ 为 $\chi^2(n)$ 分布的上 100α 百分位点（式中的 $f(y)$ 是 $\chi^2(n)$ 分布的概率密度函数），如图 1－5 所示。对于不同的 α 及 n，上 100α 百分位点 $\chi_\alpha^2(n)$ 的值已制成表格（见附录二）。例如 $\alpha=0.1$，$n=25$，查得 $\chi_{0.1}^2(25)=34.382$。但该表只列到 $n=40$ 为止，对于 $n>40$ 时，可用下列近似公式计算

$$\chi_\alpha^2(n) \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2 \tag{1-17}$$

式中，z_α 值就是标准正态分布的上 100α 百分位点（见附录一）。由详细的表格可以查得

$$\chi_{0.05}^2(50) = 67.5, \chi_{0.05}^2(75) = 96.2, \chi_{0.05}^2(100) = 124.3$$

而按式（1－17）算得的近似值为

$$\chi_{0.05}^2(50) = 67.2, \chi_{0.05}^2(75) = 95.9, \chi_{0.05}^2(100) = 124.1$$

可见二者相差甚微。

注意：这时的 z_α 为上 100α 百分位点，如 $\alpha=0.05$，则 z_α 为概率 $p=1-0.05$ 时的 z 值或 u 值，即 $z_\alpha=1.645$。

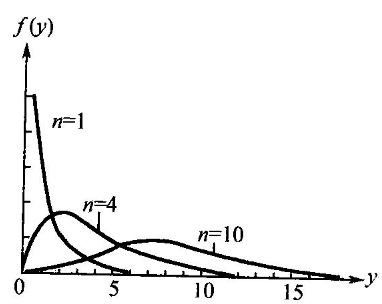


图 1－4 $\chi^2(n)$ 的概率密度函数图像

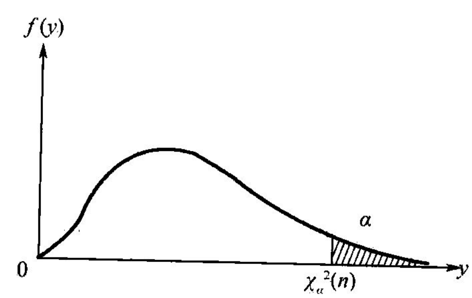


图 1－5 $\chi^2(n)$ 分布的上 100α 百分位点示意图

1.2.3 *t* 分布

设 $X\sim N(0,1)$，$Y\sim\chi^2(n)$，并且 X 和 Y 相互独立，则称随机变量

$$t = \frac{X}{\sqrt{Y/n}} \tag{1-18}$$

服从自由度为 n 的 t（Student）分布，记作 $t\sim t(n)$，其概率密度函数为

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, -\infty < t < +\infty \tag{1-19}$$

$f(t)$ 的图像如图 1－6 所示。它是关于 $t=0$ 对称的，并且形状类似于标准正态分布的图形，并且可以证明当 $n\to\infty$ 时

$$\lim_{n\to\infty} f(t) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}$$

故当 n 很大时，t 分布接近于 $N(0,1)$。对于较小的 n，已列成附录三（t 分布表）可供查阅。附录三和附录一相比，在 α 很小时，两者之间的差异很大。

t 分布的上 100α 百分位点 $t_\alpha(n)$ 是指满足

$$\int_{\alpha(n)}^{+\infty} f(t)\,\mathrm{d}t = \alpha, 0 < \alpha < 1$$

的点 $t_\alpha(n)$，式中 $f(t)$ 为 t 分布的概率密度函数，其数学意义如图 1-7 所示。

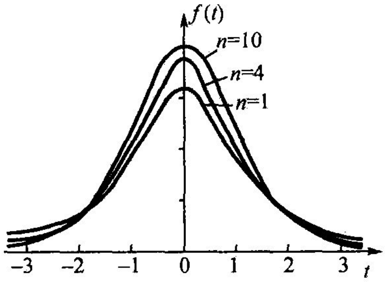


图 1-6　t 分布概率密度曲线

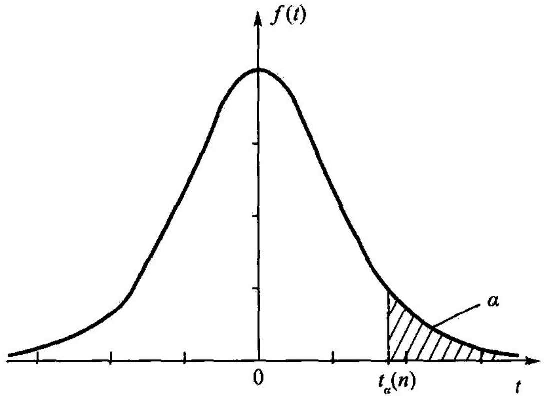


图 1-7　t 分布的上 100α 百分位点示意图

由式(1-19)可见，t 分布为偶函数，即

$$t_{1-\alpha}(n) = -t_\alpha(n)$$

又，数 $t_{\alpha/2}(n)$ 满足

$$P\{|t| > t_{\alpha/2}(n)\} = \alpha, t \sim t(n)$$

故也称 $t_{\alpha/2}(n)$ 为双侧 100α 百分位点。

t 分布的上 100α 百分位点可从附录三查得。但在 $n>35$ 时（有的表只列到 $n=30$），没有详细的表格可供查阅，就利用标准正态分布 $N(0,1)$ 来近似计算，即

$$t_\alpha(n) \approx z_\alpha, n > 35$$

1.2.4　F 分布

设 $U \sim \chi^2(n_1)$，$V \sim \chi^2(n_2)$，并且 U 和 V 相互独立，则称随机变量

$$F = \frac{U/n_1}{V/n_2} \tag{1-20}$$

服从自由度为 (n_1, n_2) 的 F 分布，记作 $F \sim F(n_1, n_2)$。

可以证明，F 分布的概率密度函数为

$$f(y) = \begin{cases} \dfrac{\Gamma\left[\dfrac{(n_1+n_2)}{2}\right]}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}\left(\dfrac{n_1}{n_2}\right)\left(\dfrac{n_1}{n_2}y\right)^{\frac{n_1}{2}-1}\left(1+\dfrac{n_1}{n_2}y\right)^{-\frac{n_1+n_2}{2}}, y \geq 0 \\ 0, y < 0 \end{cases} \tag{1-21}$$

$f(y)$ 的图像如图 1-8 所示。

F 分布的上 100α 百分位点 $F_\alpha(n_1, n_2)$ 是指满足

$$\int_{F_\alpha(n_1,n_2)}^{+\infty} f(y)\,\mathrm{d}y = \alpha, 0 < \alpha < 1$$

的点 $F_\alpha(n_1,n_2)$，式中，$f(y)$ 为 F 分布的概率密度函数。其数学意义与 χ^2 分布和 t 分布的上 100α 百分位点的意义相似，参见图 1－5和图 1－7。

$F_\alpha(n_1,n_2)$ 具有以下性质

$$F_{1-\alpha}(n_1,n_2) = \frac{1}{F_\alpha(n_2,n_1)} \quad (1-22)$$

式(1－22)常用来求 F 分布表中没有列出的某些值。例如设 $F \sim F_\alpha(15,12)$，$\alpha = 0.05$，则从附录四可查得

$$F_{0.05}(15,12) = 2.62$$

又若需要求 $F_{0.95}(15,12)$，则需应用式(1－22)，先查得

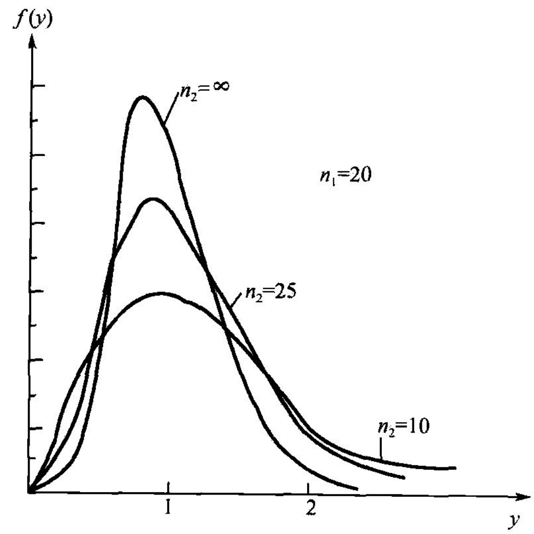


图 1－8 F 分布概率密度曲线

$$F_{0.05}(12,15) = 2.48$$

故

$$F_{0.95}(15,12) = \frac{1}{2.48} = 0.403$$

习 题 一

1. 用数学归纳法证明勒让德(Legendre)倍量公式(其中 n 为自然数)：

(1)$2 \cdot 4 \cdot 6 \cdot \cdots \cdot (2n) = 2^n\Gamma(n+1)$； (2)$1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2n-1) = \dfrac{\Gamma(2n)}{2^{n-1}\Gamma(n)}$；

(3)$\sqrt{\pi}\Gamma(2n) = 2^{2n-1}\Gamma(n)\Gamma\left(n+\dfrac{1}{2}\right)$

2. 查标准正态分布表，求：

(1)当 u 分别为 $-2, -1, 0, 1, 2$ 时的概率 p 分别是多少?

(2)当概率 p 分别为 0.001, 0.005, 0.010, 0.050, 0.100, 0.900, 0.950, 0.990, 0.995, 0.999时的 u 值分别是多少?

3. 查 χ^2 分布表或计算：

(1)$\chi^2_{0.05}(10)$；(2)$\chi^2_{0.025}(15)$；(3)$\chi^2_{0.01}(60)$；(4)$\chi^2_{0.001}(80)$；(5)$\chi^2_{0.05}(200)$

4. 查 t 分布表或正态分布表，计算：

(1)$t_{0.25}(15)$；(2)$t_{0.10}(30)$；(3)$t_{0.05}(60)$；(4)$t_{0.025}(100)$；(5)$t_{0.01}(400)$

5. 查 F 分布表或计算：

(1)$F_{0.01}(10,9)$；(2)$F_{0.90}(28,2)$；(3)$F_{0.95}(10,9)$；(4)$F_{0.995}(10,20)$