数理统计

Mathematical Statistics

Yong YANG

Beijing University of Posts and Telecommunications

July 27, 2021

目　　次

Schedule

1. 概率空间: 15% 学时
2. 随机变量 (向量) 和概率分布: 11% 学时、14% 学时
3. 数学期望和方差: 15% 学时
4. 特征函数和概率极限定理: 15% 学时
5. 样本与抽样分布: 9% 学时
6. 参数估计与假设检验: 9% 学时、12% 学时

Exercise

参考习题, 浙大书:

- Chap1:1,2,3,4,7,10,11,14,16,20,22,27,30,38,40
- Chap2:1,6,13,15,16,19,23,24,25,28,29,30,33,34(1),35(1)
- Chap3:1,2,3,5,6,9,10,13,14,16,21,24,30,36
- Chap4:4(1),7(2),11,20,24,26(2),33,36
- Chap5:2,5,12
- Chap6:2,4,6
- Chap7:2(3),3(3),8,9,12,16,21,23,25
- Chap8:2,6,13,15,18

选择课后习题的 $1/2$ 至 $3/4$, 即可达到训练的目的.

Bibliography

参考书籍:

- Probability:Theory and Examples, Rick Durrett
- 高等概率论. 程士宏. 北京: 北京大学出版社, 1996.
- 测度论讲义. 严加安. 第二版, 北京: 科学出版社, 2004.
- A First Course in Probability. Sheldon M. Ross. 8th Edition.
- A Course in Probability Theory,Chung K L. Second Edition.
- 概率论与数理统计. 陈希孺. 合肥, 中国科学技术大学出版社, 2009.
- 概率论与数理统计教程. 茆诗松等. 第二版, 北京, 高等教育出版社,2011.2

描述性统计

数理统计

数理统计学是一门较年轻的学科, 它主要的发展是从 20 世纪初开始的. 在早期发展中, 起领导作用的是以 R.A.Fisher 和 K.Pearson 为首的英国学派. 特别是 Fisher 在本学科的发展中起到了独特的作用, 目前许多常用的统计方法以及教科书中的内容, 都与他的名字有关. 其他一些著名的学者, 如 W.S.Gosset(笔名 Student),J.Neyman,E.S.Pearson(K.Pearson 的儿子),A.Wald 以及我国的许宝🄵先生等, 都做出了根本性的贡献. 他们的工作奠定了许多统计学分支的基础, 提出了一系列有重要应用价值的统计方法以及一些列的基本概念和重要理论问题. 有些人认为, 瑞典统计学家 H.Cramer 在 1946 年发表的著作《Mathematical Methods of Statistics》标志着这门学科达到了成熟的阶段. 相比 R.A.Fisher 的《Experimental Design》和《Statistical Methods for Research Workers》等,Cramer 的上述著作是人们第一次用严谨的数学方法总结数理统计学的主要成就.

数理统计

收集和记录种种数据的活动, 在人类历史上十分久远. 翻开我国的二十四史, 可以看到上面有很多关于钱粮、人口及地震、洪水等自然灾害在记录. 在西方, ‘statistics’ 一词源出于 ‘state’(国家), 意指国家收集的国情资料.19 世纪中叶以后, 包括政治统计、人口统计、经济统计、犯罪统计、社会统计等多方面内容的” 社会统计学” 一词在西方开始出现, 与此相应的社会调查也有了较大发展. 人们试图通过社会调查, 搜集、整理、分析数据, 以揭示社会现象和问题, 并提出具体解决问题的方法.
这种情况延续了许多年, 研究方法属于描述统计学的范畴. 这是因为, 没有一定的数学工具特别是概率论的发展, 无法建立现代意义下的数理统计学. 也因为这方面的需求还没达到那么迫切, 足以构成一股强大的推动力. 到了十九世纪末和 20 世纪初情况才起了较大的变化. 部分人认为二十世纪初 K.Pearson 关于 χ^2 统计量的极限分布的论文可以作为数理统计学诞生的一个标志; 也有人认为, 直到 1922 年,Fisher 关于统计学的数学基础这篇文章的发表, 数理统计学才正式诞生.

综上所述，我们可以得到如下粗略的结论：收集和整理乃至使用观察和试验数据的工具由来已久，这类活动对于数理统计学的产生，可算是一个源头．十九世纪，特别是十九世纪后半期发展速度加快，且有了质的变化．十九世纪末到二十世纪初这一阶段，出现了一系列的重要工作．无论如何，至迟到二十世纪二十年代，这门学科已经稳稳地站住了脚跟．二十世纪前四十年有了迅速而全面的发展，到二十世纪四十年代时，已形成了一个成熟的数学分支．

从二战后到现在可以说是第二阶段．在这个时期中，许多战前开始形成的数理统计学分支，在战后得到了纵深的发展，理论上的深度也比以前大大加强了．同时还出现了带根本性的发展，如 Wald 的统计判决理论和 Bayes 学派的兴起．在数理统计的应用方面，也给人印象深刻．这不仅是战后工农生产和科学技术迅速发展所提出的要求，也是由于电子计算机这一有力工具的出现和飞速发展推动了数理统计学的进步．战前由于计算工具跟不上，许多需要大量计算的统计方法很难得以使用．战后有了高速计算机便变得很容易．这就大大推广了统计方法的应用．目前，统计学的研究仍然方兴未艾．在一些统计学发达的国家中，例如美国，这方面的人才数以十万计，并在大多数大学中建立了统计系．近三十年来，数理统计学在我国的发展也是令人瞩目的．

总体和参数

日常生活中，我们总是自觉或不自觉地和总体与样本打交道. 买桔子时，先要尝尝这批桔子甜不甜. 这时这批桔子是一个总体，单个的桔子是个体.
在仅关心桔子的甜度时，我们可以称单个桔子的甜度是个体，称所有桔子的甜度为总体. 这样就可以把桔子甜不甜数量化.
要了解一批桔子的甜度情况，你只需品尝一两个，然后通过这一两个桔子的甜度判断这批桔子的甜度. 这就是用个体推断总体.
为把上面的实际情况总结出来，需要引入一些术语.

总体和参数

在统计学中，我们把所要调查对象的全体叫做总体(population)，把总体中的每个成员叫做个体(individual).
总体中的个体可以用数量表示. 为了叙述的简单和明确，我们把个体看成数量，把总体看成数量的集体. 我们要调查的是总体的性质.
总体中的个体数目有时是确定的，有时较难确定，但是往往不影响总体的确定，也不影响问题的解决. 在判断一批桔子甜不甜时，你没有必要知道一共有多少个桔子.

总体和参数

总体平均是总体的平均值, 也称为总体均值(mean). 在统计学中, 常用 μ 表示总体均值. 当总体中有 N 个个体时, 第 k 个个体是 y_k 时, 总体均值

$$\mu = \frac{y_1 + \cdots + y_N}{N} \tag{1}$$

当 $y_1, \ldots, y_N$ 是总体中的全部个体时,μ 是总体均值时, 称:

$$\sigma^2 = \frac{(y_1 - \mu)^2 + \cdots + (y_N - \mu)^2}{N} \tag{2}$$

为总体方差或方差 (variance).
总体方差描述了总体中的个体向总体均值 μ 的集中程度. 方差越小, 个体向 μ 集中得越好. 总体方差 σ^2 也描述了总体中个体的分散程度或波动幅度, 总体方差越小, 个体就越整齐.

总体和参数

总体参数是描述总体特性的指标，简称为参数(parameter).
参数表示总体的特征，是要调查的指标. 总体均值、总体方差、总体标准差等都是参数. 讲到参数时，我们要明确它是哪个总体的参数.

样本和估计

考虑某大学一年级 2000 个同学的平均身高 μ. 要得到这 2000 个同学的平均身高并不是一件很困难的事情, 只要了解了每个同学的身高就可以利用公式

$$\mu = \frac{\text{这 2000 个同学的身高之和}}{2000} \tag{3}$$

计算得到.

但是在同一时刻要了解每个同学的准确身高也不是很容易的事情. 如果让各班长在班上点名登记全班同学的身高, 然后汇总, 可能一些同学一时不能给出准确的回答, 也可能有些同学受到其它同学的影响后, 偏向于把自己的身高报高或报低. 用这样的数据进行计算后得到的结果可能会产生偏差.

同一天对每个同学进行一次身高测量可以得到均值 μ 的准确值, 但是要花费同学们较多的精力. 统计上解决这类问题的最好方法时进行抽样调查, 例如在 2000 个同学中只具体测量 50 个同学的身高, 用这 50 个同学的平均身高作为总体平均身高的近似. 这时, 我们称这 50 个同学的身高为总体的样本, 称 50 为样本量.

样本和估计

从总体中抽取一部分个体, 称这些个体为样本(sample), 样本也称为观测数据(observation data).
称构成样本的个体数目为样本容量, 简称为样本量.
称从总体中抽取样本的工作为抽样(sampling).

在考虑身高问题时, 对于前述被选中的 50 个同学, 用 $x_1, \ldots, x_{50}$ 分别表示第 $1, 2, \ldots, 50$ 个同学在调查日的身高, 则这 50 个同学的身高

$$x_1, \ldots, x_{50} \tag{4}$$

是样本, 用 n 表示样本量, 则 $n = 50$.

样本和估计

样本均值是样本的平均值，用 $\overline{x}$ 表示.
给定 n 个观测数据 $x_1, \ldots, x_n$，称

$$s^2 = \frac{1}{n-1}[(x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2] \tag{5}$$

为这 n 个数据的样本方差.
样本方差 s^2 是描述观测数据关于样本均值 $\overline{x}$ 分散程度的指标，也是描述数据的分散程度或波动程度的指标.
样本标准差是样本方差的算数平方根 $s = \sqrt{s^2}$.
和总体均值 μ 相比较后知道，只要抽样合理. 对于较大的样本量 n，样本均值 $\overline{x}$ 会接近 μ. 于是，$\overline{x}$ 是总体均值 μ 的近似，所以称为 μ 的估计 (estimator).

样本和估计

估计是利用样本计算出的对参数的估计值. 估计能从观测数据直接计算出来.
对相同的观测数据, 不同的方法可以给出不同的估计结果, 所以估计不是唯一的, 这种不唯一性恰恰为统计学家们寻找更好的估计留下了余地.
实际问题中, 总体的容量往往是非常大的, 这时从数据本身无法看清总体的情况, 样本均值和样本方差等可以提供必要的信息.

样本和估计

例子 1.1　比赛中甲、乙两位射击运动员分别进行了 10 次射击，成绩分别如下:

甲	9.5	9.9	9.9	9.9	9.8	9.7	9.5	9.3	9.6	9.6
乙	9.4	9.3	9.5	9.0	9.1	9.8	9.7	9.5	9.3	9.4

问哪个运动员平均水平高，哪个运动员水平更稳定.

解: 用 $\overline{x}, s_x, \overline{y}, s_y$ 分别表示甲和乙成绩的样本均值和样本标准差，经过计算得到

$$\overline{x} = 9.67, s_x = 0.2058, \overline{y} = 9.4, s_y = 0.2449. \tag{6}$$

甲的平均水平和稳定性都比乙好.
此题表明，知道样本标准差后，可以作出更好的比较结果.

抽样调查

在日常生活中，人们总是自觉或不自觉地应用抽样方法. 例如在市场上买花生和瓜子时总要先尝几个看看是否饱满和新鲜，在烧菜的过程中经常要取一点尝尝味道.

在考察汤的味道的时候，没有必要把汤喝完，只要把汤" 均匀搅拌"，从中品尝一勺就可以了，注意无论这汤有多多，只要一勺就够了. 记住上面的例子是大有好处的，因为它提供了抽样调查方法的最重要信息.

第一，把汤" 搅拌均匀" 是说明抽样的随机性，没有抽样的随机性，样本进不能很好地反映总体的情况. 把刚加盐的地方舀出的汤做样本，你会得出汤太咸了的错误结论；第二，品尝一勺指出了选取的样本量不能太少，太少了不足以品出味道，品尝一大碗也没有必要；第三，" 无论这锅汤有多多，只要一勺就够了"，这里体现抽样调查的如下基本性质：总体个数增大时，样本量不必随着增大.

有人认为，总体数目很大时，样本量也必须跟着增大. 这种认识带有片面性. 实际的情况时这样的：在随机抽样下，一开始增大样本量会很快地增加估计量的准确度，但是当样本量达到一定的时候，继续增加样本量效果就不明显了.

抽样调查————抽样调查的必要性

抽样调查是相对于普查而言的，其含义是从总体中按一定的方式抽出样本进行考察，然后用样本的情况来推断总体的情况.
在评价 1000 个同型号的微波炉的平均工作寿命 μ 时，预备从中抽取 n 个进行工作寿命的测量试验，用这 n 个微波炉的平均工作寿命估计总体的平均寿命 μ.
这里，总体是 1000 个微波炉的工作寿命，样本量是 n，被选中的微波炉的工作寿命构成样本. 样本平均 $\overline{x}$ 是总体均值 μ 的估计.
在正确选择的前提下，样本量越大，$\overline{x}$ 越接近总体均值 μ. 但是，较大的样本量造成的花费也很大，因为这 n 个微波炉做完寿命试验后就报废了.
在本题中想要得到真正的总体均值 μ 是不可能的，除非把这 1000 个微波炉都拿来做工作寿命试验，报废掉这 1000 个微波炉.

抽样调查————抽样调查的必要性

在很多实际问题中，采用抽样调查的方法来确定总体性质不仅是必要的，也是必须的.
总体很大时，抽样调查往往可以提高调查的质量. 有人认为抽样调查不如全面调查来得结论准确，这是不客观的. 看到抽样调查是用局部推断全体，带有抽样的误差，这只是看到了问题的一个方面. 实际上调查数据的质量更加重要，总体很大进行全面调查，往往因为工作量过大，时间过长等而影响数据的质量. 一项经过科学设计并严格实施的抽样调查可能得到比全面调查更可靠的结果.

抽样调查————随机抽样

如果总体中的每个个体都有相同的机会被抽中，就称这样的抽样方法为随机抽样方法. 人们经常用"任取"、"随机抽取"、"等可能抽取" 等来表示随机抽样.

从概率论的知识知道，如果从总体中任选一个个体，这个个体是随机变量，这个变量的数学期望是总体均值，方差是总体方差.

随机抽样又分为无放回的随机抽样和有放回的随机抽样. 放回的随机抽样指在总体中随机抽出一个个体后，下次在余下的个体中再进行随机抽样. 有放回的随机抽样指抽出一个个体后，记录下抽到的结果后放回，摇匀后再进行下一次随机抽样.

抽样调查————随机抽样

例子 2.1　设 N 件产品中有 M 件次品,N, M 都是未知的. 估计这类产品的次品率 $p = M/N$.
解：无放回地从中依次取 n 件，用 Y 表示取得的次品数，则 $Y \sim H(N, M, n)$，根据概率论的知识，有

$$\mathrm{E}Y = np, \mathrm{var}(Y) = np(1-p)\frac{N-n}{N-1}. \tag{7}$$

用样本次品率 $\hat{p} = Y/n$ 估计 p 时，有

$$\mathrm{E}\hat{p} = p, \mathrm{var}(\hat{p}) = \frac{1}{n}p(1-p)\frac{N-n}{N-1}. \tag{8}$$

抽样调查————随机抽样

例子 2.1　设 N 件产品中有 M 件次品,N, M 都是未知的. 估计这类产品的次品率 $p = M/N$.
如果采用有放回的随机抽样, 用 X 表示取得的次品数, 则 $X \sim \mathcal{B}(n, p)$, 这时有

$$\mathrm{E}X = np, \mathrm{var}(Y) = np(1-p). \tag{9}$$

用这时的样本次品率 $\tilde{p} = Y/n$ 估计 p 时, 有

$$\mathrm{E}\tilde{p} = p, \mathrm{var}(\tilde{p}) = \frac{1}{n}p(1-p). \tag{10}$$

$\mathrm{E}\hat{p} = \mathrm{E}\tilde{p} = p$, 说明这两种方法都是较好的估计方法, 没有系统偏差.

抽样调查————随机抽样

例子 2.1　设 N 件产品中有 M 件次品,N, M 都是未知的. 估计这类产品的次品率 $p = M/N$.

由于方差 $\mathrm{E}(\hat{p} - p)^2$ 描述的是 $\hat{p}$ 向真实参数 p 的集中程度, 因而是描述估计精度的量. 方差越小, 说明估计的精度越高.$\mathrm{var}(\hat{p}) < \mathrm{var}(\tilde{p})$ 说明无放回随机抽样的估计精度好于有放回随机抽样的估计精度. 但是当 N 比 n 大很多时,$(N - n)/(N - 1)$ 接近于 1, 说明两种抽样方法差不大.

另外 $\mathrm{var}(\tilde{p})$ 与 N 无关, 说明达到一定的估计精度, 只需要适当地增加 n. 并不是说总体数目 N 越大, 就需要多抽样. 无放回随机抽样的情况也是类似的, 因为随机问题中 N 通常都很大, 而相比之下 n 较小.

抽样调查————随机抽样

在相同的总体中和相同的样本量下，无放回随机抽样得到的结果比有放回的随机抽样得到的结果要好。但是当总体的数量很大，样本量相对总体的数量又很小时，这两种抽样方法得到的结果是相近的.

试验和理论都表明: 在随机抽样下，样本均值 $\overline{x}$ 是总体均值 μ 很好的估计，样本标准差 s 是总体标准差 σ 很好的估计. 在样本量不大时，增加样本量可以比较好地提高估计的精度.

考虑某大学一年级 2000 个同学的平均身高 μ 时，需要调查 50 同学的身高. 实现无放回的随机抽样的方法是先将 2000 个同学的学号分别写在 2000 张小纸片上，然后放入一个大纸箱进行充分地摇匀，最后从纸箱中无放回地抽取 50 个纸片，纸片上的学号就是被选中的同学的学号.

抽样调查————随机抽样的无偏性

样本均值是对总体均值的估计. 在总体中任取一个个体 X,X 是随机变量, 从数学期望的定义知道 $\mathrm{E}X = \mu$ 是总体均值. 这说明随机抽样是无偏的. 如果用 $X_1, \ldots, X_n$ 表示依次随机抽取的样本, 则样本均值

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \tag{11}$$

是总体均值 μ 的估计. 下面证明 $\mathrm{E}\overline{X} = \mu$. 在有放回的随机抽样下,$X_1, \ldots, X_n$ 有相同的数学期望 μ, 于是有

$$\mathrm{E}\overline{X} = \frac{1}{n}\sum_{j=1}^{n} \mathrm{E}X_j = \frac{1}{n}\sum_{j=1}^{n} \mu = \mu. \tag{12}$$

抽样调查————随机抽样的无偏性

下面是没有采取正确的抽样方案导致调查结论严重失真的著名案例.

例子 2.2　1936 年是美国总统选举年. 这年罗斯福 (Roosevelt) 任美国总统期满, 参加第二届的连任选举, 对手是堪萨斯州州长兰登 (Landon). 当时美国刚从经济大萧条中恢复过来, 失业人数仍高达 900 多万, 人们的经济收入下降了三分之一后开始逐步回升, 当时, 观察家们普遍认为罗斯福会当选. 而美国的《文学摘要》杂志的调查却预测兰登会以 57% 对 43% 的压倒优势获胜.
《文学摘要》的预测是基于对 240 万选民的民意调查得出的. 自 1916 年以来, 在历届美国总统的选举中《文学摘要》都做了正确的预测.《文学摘要》的威信有力地支持着它的这次预测.
但是选举的结果是罗斯福以 62% 对 38% 的压倒优势获胜, 此后不久《文学摘要》杂志就破产了.

抽样调查————随机抽样的无偏性

要了解《文学摘要》预测失败的原因就必须检查他们的抽样调查方案.
《文学摘要》是将问卷寄给了 1000 万个选民, 基于收回的 240 万份问卷得出的判断. 这些选民的地址是在诸如电话簿、俱乐部会员名单等上查到的.
分析:19366 年只有大约四分之一的家庭安装了电话. 由于有钱人才更有可能安装家庭电话和参加俱乐部, 所以《文学摘要》的调查方案漏掉了那些不属于俱乐部的穷人和没有安装电话的穷人, 这就导致了调查结果有排除穷人的偏向.
在 1936 年, 由于经济开始好转, 穷人普遍有赞同罗斯福当选的倾向, 富人有赞同兰登当选的倾向.《文学摘要》的调查结果更多地代表了富人的意愿, 导致了预测的失败.
抽样的方案应该公平地对待每一位选民和每一个群体, 以便得到选民的真实情况. 将哪一个群体排除在外的抽样方案都可能导致有偏的样本, 从而导致错误的结论.

抽样调查————分层抽样方法

例子 2.3　2000 年, 某市进行家庭年收入调查时, 分别对城镇家庭和农村家庭进行调查. 在全部城镇的 85679 户中无放回随机抽取了 350 户, 在全部农村的 275692 户中无放回抽取了 360 户. 调查结果如下:

城镇家庭年平均收入是 35612 元, 农村家庭年平均收入是 5623 元.

这里遇到了两个子总体 A_1 和 A_2, 第一个子总体 A_1 是所有城镇家庭的年收入, 第二个子总体 A_2 是所有农村家庭的年收入. 用 A 表示该市所有家庭的年收入时, 总体 A 时两个子总体 A_1 和 A_2 的并.

用 $\overline{x}_1$ 表示来自总体 A_1 的样本均值, 用 $\overline{x}_2$ 表示来自总体 A_2 的样本均值, 则

$$\overline{x}_1 = 35612, \overline{x}_2 = 5623.$$

A_1 在 A 中所占的比例是

$$W_1 = \frac{85679}{85679 + 275692} = 0.2371.$$

抽样调查————分层抽样方法

A_2 在 A 中所占的比例是

$$W_2 = \frac{275692}{85679 + 275692} = 0.7629.$$

A 的总体均值 μ 的估计是

$$W_1\overline{x}_1 + W_2\overline{x}_2 = 0.2371 \times 35612 + 0.7629 \times 5623 = 12733(\text{元}).$$

于是该市平均家庭年收入的估计是 12733 元.

上面的抽样调查问题中, 还可以把全部家庭再细分成城镇中的工人、公务员、教师等; 将农村家庭分成农民家庭、农村干部家庭等.

于是引出下面的分层抽样方法.

抽样调查————分层抽样方法

分层抽样就是把总体 A 分成 L 个互不相交子总体, 即

$$A = A_1 + A_2 + \cdots + A_L, \tag{13}$$

称这些子总体为层, 称 A_i 为第 i 层, 然后在每层中独立地进行随机抽样.

用 N 表示总体 A 的个体总数, 用 N_i 表示第 i 层的个体总数时, 有

$$N = N_1 + N_2 + \cdots + N_L. \tag{14}$$

这时称

$$W_i = \frac{N_i}{N}, (i = 1, \ldots, L) \tag{15}$$

为第 i 层的层权 (weight).

抽样调查————分层抽样方法

用 μ 表示 A 的总体均值. 对 $i = 1, \ldots, L$, 用 $\overline{x}_i$ 表示从第 i 层抽出样本的样本均值, 我们称

$$\overline{x}_{st} = W_1\overline{x}_1 + \cdots + W_L\overline{x}_L \tag{16}$$

是总体均值 μ 的简单估计. 称

$$\mathrm{var}(\overline{x}_{st}) = W_1^2\mathrm{var}(\overline{x}_1) + \cdots + W_L^2\mathrm{var}(\overline{x}_L) \tag{17}$$

是简单估计 $\overline{x}_{st}$ 的抽样方差.

简单估计的抽样方差 $\mathrm{var}(\overline{x}_{st})$ 是评价简单估计 $\overline{x}_{st}$ 的估计精度的指标.$\mathrm{var}(\overline{x}_{st})$ 越小, 说明 $\overline{x}_{st}$ 越好.

在例子 2.3 中, 如果从城镇家庭中只抽取一个个体, 在农村中也只抽取一个个体, 这两个个体的平均值是不能估计总体均值的. 同样的道理, 如果不对例子 2.3 中的样本进行加层权平均, 将得到错误的估计.

抽样调查————分层抽样方法

分层抽样是一种常用的抽样方法，有如下的特点:
(1) 分层抽样在获得总体均值估计的同时，也得到了各层的均值估计. 在例子 2.3 中，不但得到了总体 A 的均值估计，还得到了子总体 A_1 和 A_2 的均值估计;
(2) 将差别不大的个体放在同一层，使得分层抽样得到的样本更具有代表性，从而提高估计的准确度;
(3) 抽样调查的实施更加方便，调查数据的收集、处理也更加方便.

抽样调查————系统抽样方法

例子 2.4　在调查某居民住宅区的 999 个住户对住宅区的环境满意程度时，要按照 1:14 的比例进行抽样调查，将这 999 户按门牌号码的顺序依次编号. 下面的每个数对应一户的门牌号码.

1	2	3	4	5	6	...	13	14
15	16	17	18	19	20	...	27	28
29	30	31	32	33	34	...	41	42
...	...	...	...	...	...	...	...	...
981	982	983	984	985	986	...	993	994
995	996	997	998	999				

现在 $1 \sim 14$ 中随机抽取一个数字，如果抽到 7，就调查排在第 7 列的所有家庭，请这些家庭对小区环境的满意程度打分，分数分为 1,2,3,4,5 级. 第 7 列有 71 户，所以样本量 $n = 71$. 这 71 户的平均分是样本均值，用样本均值作为全体住户对小区环境的平均分的估计.

抽样调查————系统抽样方法

用 x_i 表示这 71 户中第 i 户的打分，样本均值是

$$\overline{x} = \frac{x_1 + \cdots + x_{71}}{71}.$$

我们称上面的方法为系统抽样.

如果总体中的个体按一定的方式排列，在规定的范围内随机抽取一个个体，然后按照制定好的规则确定其他个体的抽样方法称为系统抽样.
最简单的系统抽样是取得一个个体后，按照相同的间隔抽取其他个体.
系统抽样的主要优点是实施简单，只需要先随机抽取第一个个体，以后按规定抽取就可以了. 系统抽样不像随机抽样，随机抽样每次都要随机抽取个体. 如果了解总体中个体排列的规律，设计合适的系统抽样规则可以增加估计的精度.

用样本估计总体分布

数据中的大量信息都可以概括在图表内，图表使人一目了然. 在实际问题中，样本量往往是比较大的，这时数据中的主要信息隐藏在背后. 要从数据中得到这些信息，必须对观测数据进行整理. 下面是几种常用的数据整理方法.

A、频率分布表　　制作频率分布表时，先将数据从小到大排列，然后将排列后的数据进行分段. 每段中的数据被称为一组数据，所以又把分段称为分组. 一般来讲，当样本量是 n，可以参照下面的经验公式将数据分成大约

$$K = 1 + 4 \lg n \tag{18}$$

段. 这里的经验公式支队分段起参考作用，实际应用时，应当根据样本量的大小和数据的特点以及分析的要求灵活确定.

用样本估计总体分布

例子 3.1　下面是某城市公共图书馆在一年中通过随机抽样调查得到的 60 天的读者借书数, 数据已经从小到大排列, 制作频率分布表:

213	230	239	289	291	301	308	310	311	312
318	318	337	343	344	348	349	351	360	362
368	372	374	379	383	385	390	393	396	399
400	404	406	425	429	430	436	438	440	441
444	446	450	453	456	458	471	473	475	483
484	495	498	498	521	524	549	556	568	584

解　数据中的最小值是 213, 最大值是 584. 这 60 个数据就散布在闭区间 $[213, 584]$ 中. 取一个略大的区间 $(200, 600]$, 它的端点都是整数. 用经验公式计算得出

$$K = 1 + 4 \lg n = 1 + 4 \lg 60 = 8.1126.$$

我们将 $(200, 600]8$ 等分, 排在下表的第一列. 计算出数据落入各段的个数 n_i, 填入第二列. 计算出数据落入各段的频率

用样本估计总体分布

$$f_1 = \frac{3}{60} = 5\%, f_2 = \frac{2}{60} = 3.3\%, \ldots, f_8 = \frac{3}{60} = 5\%$$

依次填入第三列. 最后将各列之和填入最后一行, 得到频率分布表

借出书数 i	发生次数 n_i	$f_i =$ 发生频率
$(200, 250]$	3	5%
$(250, 300]$	2	3.3%
$(300, 350]$	12	20%
$(350, 400]$	14	23.3%
$(400, 450]$	12	20%
$(450, 500]$	11	18.3%
$(500, 550]$	3	5%
$(550, 600]$	3	5%
总计	60	99.9%

用样本估计总体分布

由于计算频率时四舍五入引起计算误差, 频率之和可能是 1 的近似.
从上述频率分布表可以方便地分析出以下结果:
有 8.3% 的工作日借出的突出少于等于 300 册;
有 63.3% 的工作日借出图书的数量在 301 至 450 册之间;
有 48.3% 的工作日借出的图书在 400 册以上;
只有 10% 的工作日借出的图书多于 500 册.
当总体是全年每个工作日的借书数量时, 上述结果可以作为对总体的推测.
Rmk: 由于频率分布表的制作没有统一的数据分段方法, 所以对相同的数据, 可以作出不同的频率分布表. 但是好的频率分布表应当是简单明了的.

用样本估计总体分布

B、频率分布直方图

数据的频率分布表初步展示了数据分布的一些规律. 如果用图形来表示频率分布就会更加形象和直观. 从文献上看, 直方图在 1895 年由著名的英国统计学家皮尔逊 (Pearson) 作了描述, 这可能是直方图的第一次使用. 他作为伦敦皇家协会发表的讲话中, 当谈及 1885-1886 年英格兰房地产估价的时候使用了直方图.

有了数据的频率分布表, 很容易做出频率分布的直方图. 将观测数据按照制作频率分布表的方法进行分段, 计算出数据落入各段的频率 f_i. 将各段的端点画在直角坐标系的横坐标上, 用

$$g_i = \frac{f_i}{\text{本段的区间长度}} \tag{19}$$

作为纵坐标的高, 就得到了相连接长方形构成的图形. 我们把所得到的图形称为数据的频率分布直方图, 简称为直方图(histogram).

例子 3.2　绘制例 3.1 中图书馆借出图书数据的频率分布直方图.

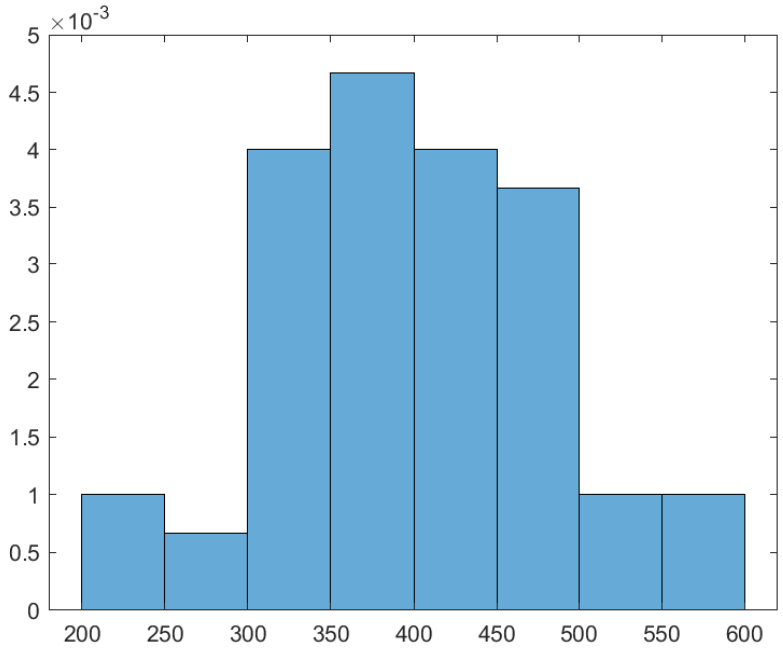
解　在横坐标上标出数据分段的端点 $200, 250, \ldots, 550, 600$.

在区间 $[200, 250]$ 上绘制以 $g_2 = 0.05/50$ 为高的矩形;

在区间 $[250, 300]$ 上绘制以 $g_2 = 0.033/50$ 为高的矩形;

$\ldots\ldots\ldots$;

在区间 $[550, 600]$ 上绘制以 $g_8 = 0.05/50$ 为高的矩形,



就得到了需要的频率分布直方图. 从直方图可以更直观地看到图书馆每日借出图书册数的分布情况.

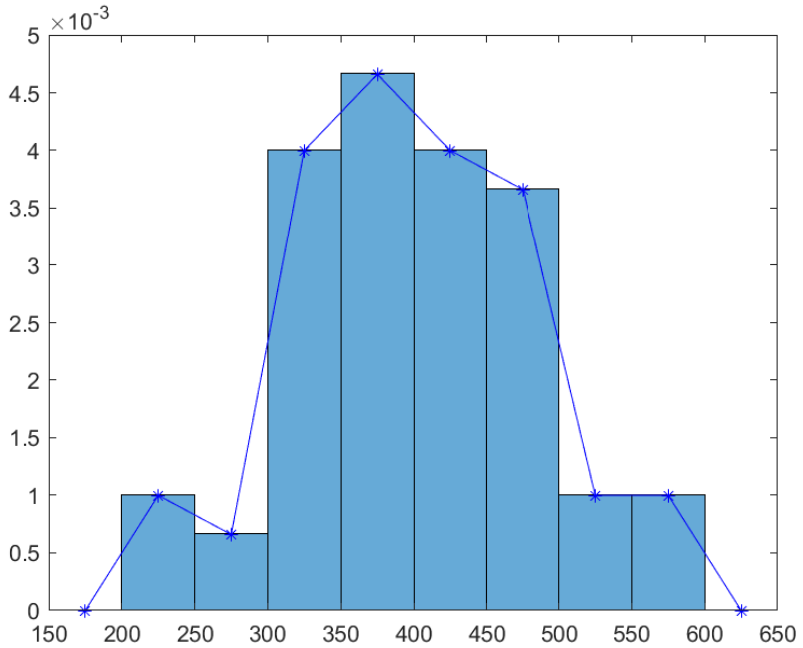
用样本估计总体分布

C、频率折线图

用 $d_1, d_2, \ldots, d_k$ 分别表示频率分布直方图中各矩形上边的中点，在直方图的左边延长出一个分段，分段的中点用 d_0 表示. 在直方图的右边也延长出一个分段，分段的中点用 d_{k+1} 表示.

用直线连接 $d_0, d_1, \ldots, d_{k+1}$ 就得到了一条折线，这条折线叫做频率折线图. 频率折线图也反映出数据频率分布的规律.

下图是例子 3.1 中图书馆借出图书数据的频率折线图.



用样本估计总体分布

D、数据茎叶图

直方图主要用于展示分段数据的频率分布，对于没有分段的观测数据还可以用数据的茎叶图展示它的特性.

数据的茎叶图由”茎”和”叶”两部分组成，在制作茎叶图的时候要先确定数据的”茎”和”叶”. 从数据的茎叶图可以看出数据的分布形状以及数据是否对称，是否集中等分布特性. 我们通过举例说明茎叶图的制作方法.

例: 下面是上海市 2004 年 7 月 10-31 日空气中可吸入颗粒物的监测数据，以这批数据制作茎叶图

85 85 66 71 62 52 55 59 52 62 59 70 80 96 97 94 62 51 57 67 96 93

用样本估计总体分布

解: 先将数据从小到大排列得到
51 52 52 55 57 59 59 62 62 62 66 67 70 71 80 85 85 93 94 96 96 97
数据的十位上的数是 5,6,7,8,9, 把他们叫做" 茎", 排列在下表的第一列, 茎 5 后面的个位数分别是 1,2,2,5,7,9,9, 把它们叫做茎 5 的" 叶", 排在茎 5 的后面;
按相同的方法把茎 6 的叶 2,2,2,6,7 排在茎 6 的右边;
. . . . . .
把茎 9 的叶 3,4,6,6,7 排在茎 9 的右边.
就得到了如下的茎叶图

用样本估计总体分布

茎	叶
5	1225799
6	22267
7	01
8	055
9	34667

从茎叶图中看出，尽管这 22 天中可吸入颗粒物都是处于良的水平，但是有较多的时间接近于优，也有较多的时间接近于轻度污染.

数据茎叶图的优点是利用了数据的每个信息，从茎叶图中可以直观地看到数据的分布情况. 但是数据量很大时，茎叶图的效果就不好了，因为这是的茎叶图会很长很粗.

众数和中位数

数据的频率分布表、频率分布直方图和茎叶图都可以展示出数据的分布形状，从中可以对数据有一个大致的了解. 为了更好地掌握数据的特性和规律，还需要进一步考虑代表数据特征的其他指标.

A、众数

我们称观测数据中出现次数最多的数是众数(mode)，用 M_0 表示.
按照这个定义，在抽样调查中，样本中出现次数最多的数据是样本的众数.

如果观测数据中每个数出现的次数都相同，它就没有众数. 如果观测数据中有两个或两个以上的数出现次数相同，且出现次数超过其它数的出现次数，这几个数都是众数.

众数是观测数据的代表值，它受数据中极大或极小值的影响较小. 从分布的角度看，众数出现的频率最高.

(样本) 众数和 (样本) 中位数

例子: 某超市用随机抽样的方式调查了 30 个顾客购买商品的件数, 结果从小到大排列如下:
0,0,1,1,1,2,2,2,3,3,4,5,6,6,8,9,9,10,10,10,10,12,12,13,15,16,18,20,23,29.
求众数和样本均值.
解: 样本中 10 出现的次数最多, 是 4 次, 所以 10 是众数. 样本均值是

$$\overline{x} = \frac{1+1+\cdots+23+29}{30} = 8.667. \tag{20}$$

在此例中, 如果购买件数最多的那个顾客购买件数从 29 增加到 40, 众数不变, 样本均值增加到 9.03.
从这个例子中, 数据中最大值的变化对众数没有影响, 对样本均值的影响较大.
在统计学上, 我们将数据的最大值和最小值统一称为极值. 称最大值和最小值之差为极差,极差 = 最大值 − 最小值.
上例中的极差为 $29-0=29$, 说明所有数据的变化范围或波动幅度不超过 29.

(样本) 众数和 (样本) 中位数

B、中位数

设观测数据已经从小到大排列为 $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$. (1) 如果样本量 n 是奇数, 我们称中间的数据是中位数(median), 记作 M_d.

$$M_d = x_m, \text{其中} m = \frac{n+1}{2}. \tag{21}$$

例如样本 1,5,9,12,13 的中位数是 9. 样本 4,26,45,67,96,98,112 的中位数是 67.
(2) 如果样本量 n 是偶数, 我们称两个中间数据的平均值是中位数, 也记作 M_d.

$$M_d = \frac{x_m + x_{m+1}}{2}, \text{其中} m = \frac{n}{2}. \tag{22}$$

例如 34,36,45,67,96,98,112,134 的中位数是 $M_d = (67+96)/2 = 81.5$

由于中位数位于顺序数据的中间, 所以有以下性质:
小于等于中位数的数据不少于样本量的二分之一, 大于等于中位数的数据不少于样本量的二分之一.

例: 2018 年, 在对 A 城市工作的某大班同学进行年人均收入 (单位: 万元) 调查时, 采用随机抽样的方法得到了以下 10 个数据:
7.9,9.8,11.7,14.6,16.7,17.9,18.2,19.8,22.6,97.8.
计算中位数和样本均值.
解: 数据已经从小到大排列. 中位数是 $M_d = (16.7 + 17.9)/2 = 17.3$.
样本均值 $\bar{x} = (7.9 + 9.8 + \cdots + 97.8) = 23.7$.
在本例中, 因为 97.8 万元的年收入比其他人的年收入高得太多了, 所以 97.8 拉高了样本均值. 但是, 即使将 97.8 改为 27.8, 中位数也不变化.

随机对照实验

近年来, 不断发生的公共安全事故已经成为公众关心的热点. 在大洋的彼岸, 美国食品药品监督管理局 (简称 FDA) 一直恪守成规, 遵照美国的食品和药品法规行事, 最大程度地将食品药品安全事故降到最低. 尽管如此, 在 20 世纪 80 年代上市的抗心律失常药 Tambocor(氟卡胺) 和 Enkaid(英卡胺) 等还是引发了一次重大药害事件. 短短的几年内, 估计有 5 万人因服用这类药物导致心脏骤停而死亡. 下面的内容选自 [Thomas J. Moore. 《致命的药物》]

随机对照实验

在 20 世纪 80 年代，3M 是美国最大的公司之一，曾在财富 500 强中名列第 47 位。它因买下一个小型制药公司，再经过 10 多年的研发，终于有了自己的新药 Tambocor(氟卡胺)。新药的上市，成了 3M 公司的摇钱树。13 年来，Tambocor 曾在小鼠、大鼠、兔子、猪、狗、猫和狒狒体内进行了试验研究。1975 年，FDA 批准了 Tambocor 的临床研究申请，1975-1978 年，对于 Tambocor 的 I 期和 II 期临床试验结果比较令人满意。与此同时，百时美公司也加快了同类药品 Enkaid(英卡胺) 的研发速度。

20 世纪 70 年代，用药物预防心率失常的理论开始流行，尽管缺乏充分的科学依据，1979 年医生开出了 1200 万次处方药，这些事实强烈刺激着 3M 的 Tambocor 研发。

随机对照实验

1982 年，德国批准了 Tambocor 的销售。接下来，斯坦福大学的温克医生发表文章讲述了 Enkaid 导致了一些重症患者的死亡。Tambocor 和 Enkaid 的化学结构相似，但是不一定又相同的问题。温克医生和 3M 公司讨论 Tambocor 时，提出在重症患者中试验 Tambocor，而 3M 公司只想在有室性早搏症，且身体健康的人群中试验 Tambocor。
1982 年 11 月，3M 公司邀请了医学界的专家分析临床试验数据。在 45 家医院服用 Tambocor 的患者中，Tambocor 没有表现出良好的疗效，反而导致了部分患者的死亡。Enkaid 也遇到了同样的问题。可惜的是，因为没有统一的试验方案，研究人员很难对于这些数据加以统计分析。尽管如此，在进一步的试验结果出来之前，3M 公司还是按原计划向 FDA 提交了新药上市的申请。

随机对照实验

1983 年，3M 公司在旅游圣地百慕大召开了 Tambocor 的研讨会，3M 支付了全部费用，除了温克医生继续表示担忧外，大部分收了报酬的专家对 Tambocor 给予热情的支持。
1985 年 10 月 8 日，在缺乏严密的试验数据的情况下，Tambocor 及其同类药物获得上市批准。新一代抗心律失常药开始大量涌向市场。之后，3M 借助推销、广告、研讨会等大力推广 Tambocor。1988 年春，美国的医生每月平均开出 57000 张 Tambocor 处方。
1986 年 12 月，FDA 批准了 Enkaid 上市。在一个以市场促销为目的的临床试验中，6 周内 Enkaid 导致了 39 人死亡。至此，FDA、制药公司和医生都注意到这类药会导致心脏骤停，但是他们认为这些药利大于弊。百时美和 3M 公司继续开动着赚钱机器。

随机对照实验

幸运的是，在抗心律失常药被审批和销售的同时，美国心肺血液研究所开始对这类药物进行大规模的心律失常抑制试验（简称 CAS 试验）。试验包括 Tambocor 和 Enkaid 等，涉及 4400 位患者，27 个研究中心和 100 多家医院，耗时 5 年，耗资 4400 万美元。
CAS 试验严格按照随机对照双盲试验的原则进行：把每个患者随机地分在 X 组或 Y 组，为其中一组提供药品，为另一组提供安慰剂（貌似药物，但无任何药效）。这是随机对照的含义。不管是研究人员还是患者，除了生物统计学家霍尔斯乔姆一人，没人知道哪些是真药，哪些是安慰剂。这是双盲的含义。谁也没有权利过问治疗组（服用真药的组）和对照组（服用安慰剂的组）的身份情况，由于霍尔斯乔姆默默地保守着秘密，使得想为药品讲话的医生也无从开口，大家只能默默地等待试验的进展。

随机对照实验

到 1988 年 9 月 1 日，CAS 试验的内部结果如下（其中的猝死人数包括心脏骤停但是抢救成功的患者）。

	X 组	Y 组
患者总数	576	571
猝死人数	3	19

可以看出，Y 组的猝死率是 X 组的 6.38 倍。这样的差异大大超出了随机因素所能解释的范围。事实上，对该数据进行如下的假设检验，设原假设 H0: 该药物与猝死无关 vs 备择假设 H1: 该药物与猝死有关，由调查数据得到：

统计量	值	P 值
卡方	12.0068	0.0005
似然比卡方检验	13.3432	0.0003
连续调整卡方	10.5612	0.0012
Mantel-Haenszel 卡方	11.9963	0.0005

随机对照实验

可以看到，无论是上述的哪一种检验方法，其 P 值都远小于 0.01，所以这个检验是高度显著的，我们可以肯定地说，根据 CAS 试验，认为该药物与猝死有关时，所犯错误的概率不会超过 $=0.01$，我们应当否认该药物与猝死无关。随着试验的继续，两组间的差异没有缩小，趋势也没有发生转变。因为患者是被随机分配到 X 组和 Y 组的，所以结论表明，不是该药品有效，就是该药品致命。
现在，我们已经知道了，X 组是安慰剂组，Y 组是试验组。随机对照双盲试验清楚地揭示了 Tambocor 和 Enkaid 的确是致命的药物。CAS 试验中止后，CAS 试验项目长官弗里德曼在给 3M 公司的信中写道：“采取这个行为是因为 Tambocor 经证明基本不可能存在疗效，反而它很可能对该患者群体有危害”。

随机对照实验

在 CAS 试验中，称使用真药的人在试验组，使用安慰剂的人在对照组。通常，实验组由随机选择出的对象构成，试验组的成员接受特殊的待遇或治疗。而对照组由那些没有接受过这种特殊待遇的对象构成，通常为他们提供的是安慰剂。任何好的试验设计应当有一个实验组和对照组。在 CAS 试验中，如果没有对照组，为所有的患者提供药品，就无法确认 Tambocor 和 Enkaid 会造成心脏骤停的结论。如果试验组和对照组不是随机选择的，由于两组人群的差异，也无法分析出正确的结论。同样，不让医生和患者知道患者在哪一组，甚至不让他们知道安慰剂的存在，是为了得到没有偏见的数据。

随机对照实验

当然，随机对照试验也会遇到道德方面的谴责。在决定停止对 Tambocor 和 Enkaid 的 CAS 试验时，就有医生愤怒指责 CAS 试验“不道德”。因为一旦试验证明药品有效，那么分在对照组的一半人就没有得到治疗。有这样想法的医生并不少。因为当时有一半的医生在治疗心脏早搏，都以为自己在帮助患者。但是，CAS 试验的结论证明，他们正在无意中杀害自己的患者。

Rmk: 上述抗心律失常药对于部分轻度心律失常患者有效.

随机选择试验对象是英国统计学家 Fisher 的贡献，在 20 世纪初，他用此方法致力于农业试验对象的研究。从此，随机选择试验组成为安排试验的基本原则。下面的例子来自 [Freedman D.《统计学》]、[Iverson G R, Gergen M.《统计学》]、[Grace N D,et al. The present status of shunts for protal hypertension in cirrhosis. Journal of Gastroenterology.1966,50,686-691]、[Sacks H, Chalmer T C and Smith H. Randomized versus historical controls for clinical trials. American Journal of Medicine,1982,72,233-240]

随机对照实验

例 (静脉吻合分流术) 在一些肝硬化病例中，许多患者会从肝出血直至死亡. 历史上有一种称为" 静脉吻合分流术" 的外科手术用于治疗肝硬化，其原理是用外科手术的方法使血流改变方向. 这种手术花费很大并且有很高的危险性. 值得做这样的手术吗？

为了解决上述问题，一共有三批共 51 次手术试验. 第一批进行了 32 次无对照组的试验. 结果如下:

设计方法	试验次数	显著有效	中等有效	无效
无对照组	32	24	7	1
所占比例		75%	21.9%	3.1%

试验说明 75% 的手术显著有效,21.9% 的手术中等有效，看来手术是值得做的.

随机对照实验

第二批共进行了 15% 次手术试验, 这批试验有对照组, 但是对照组的患者不是随机选取的. 医生根据患者的临床诊断情况决定是将患者编入试验组做手术, 还是编入对照组不做手术. 结果如下:

设计方法	试验次数	显著有效	中等有效	无效
非随机对照	15	10	3	2
所占比例		66.7%	20%	13.3%

这次试验的结果是 66.7% 的手术显著有效,20% 的手术中等有效,13.3% 的手术无效. 这个试验结果也是对" 静脉吻合分流术" 的肯定. 这次的结果与无对照组的试验结果差别不是很大.

随机对照实验

再看有随机选取的对照组的第三批试验，这批试验只有 4 次手术. 随机选取的方式类似于掷硬币，如果硬币正面朝上就将患者选入试验组做手术. 这次试验的结果如下:

设计方法	试验次数	显著有效	中等有效	无效
随机对照	4	0	1	3
所占比例		0%	25%	75%

随机对照试验的结果显著地否定了外科手术”静脉吻合分流术”的价值. 经过认真设计的试验研究显示”静脉吻合分流术”几乎没有什么价值.

随机对照实验

为什么会出现如此大的差别呢?
在无对照组和非随机选取对照组的试验中, 试验者根据患者的临床诊断决定是否将他编入试验组进行手术. 这样就做出现一种自然的倾向: 试验人员更倾向于将那些身体状态较好的患者选入试验组, 以减少手术风险. 其结果有利于对手术的肯定评价, 这种结果是不真实的.
对上述试验的跟踪观测发现, 做手术的 51 个患者中 3 年后大约有 60% 仍然活着, 随机对照组中 (没做手术的患者)3 年后大约也有 60% 仍然活着. 这说明手术基本是无效的. 而在非随机对照组中, 只有 45% 的患者存活期超过三年, 这说明了非随机对照组中的患者健康情况较差, 验证了健康情况较好的患者更容易被选入试验组做手术.
随机安排对照组是十分必要的, 否则可能得出错误的结论. 我们称随机选取试验组的对照试验为随机对照试验. 在随机对照试验中, 为了得到更真实的结果, 有时还需要其他的手段配合.

随机对照实验

在人类历史上，还有许多成功使用随机对照试验的例子，也有许多惨痛的教训。例如，随机对照试验否定了治疗冠状动脉病的冠状动脉旁道外科手术（该手术费用昂贵），否定了用抗凝剂治疗心脏病突发，否定了用 5-FU 对结肠癌进行化疗，否定了用己烯雌酚预防流产. 具体情况如下:

医疗方法	随机对照试验		非随机对照试验	
结论	有效	无效	有效	无效
冠状旁道外科手术	1	7	16	5
抗凝剂治疗	1	9	5	1
5-FU 结肠癌化疗	0	5	2	0
己烯雌酚预防流产	0	3	5	0

随机对照实验

特别需要指出的是有关己烯雌酚的试验，随机对照试验完全否定了这种预防流产的药物。而历史上糟糕的非随机对照试验缺赞同药物的疗效，这是一个医学的悲剧。在 20 世纪 60 年代末的美国，医生每年大约为 5 万名孕妇发放这种药。后来揭示，怀孕期间的母亲服用己烯雌酚，20 年后将给她们的女儿带来灾害性的副作用，可能引发她们的女儿得一种罕见的癌症。该药于 1971 年被禁止使用。
人们从太多的悲剧中总结了教训：对一种新药不作随机对照试验是非常危险的。

例: 1916 年小儿麻痹症 (脊髓灰质炎) 袭击了美国, 以后的 40 年间, 受害者成千上万.20 世纪 50 年代, 人们开始发现预防疫苗. 当时萨凯 (Salk) 培育的疫苗最有希望. 他的疫苗在试验室表现良好: 安全, 产生对脊髓灰质炎病毒的抗体. 但是在大规模使用前必须进行现场人体试验, 通过试验最后确定疫苗是否有效. 只有这样才能达到保护儿童的目的.
当时采用了随机对照的研究方案, 对每个儿童用类似投掷一个硬币的方法决定是否将他编入试验组: 正面朝上分在试验组, 否则分在对照组. 除了试验的设计人员, 连医生也不知道哪个儿童分在试验组, 哪个儿童分在对照组.
然后给分在试验组的儿童注射疫苗, 给分在对照组的儿童注射生理盐水, 让他们认为也被注射了疫苗. 得到的结果如下:

	试验人数	试验后的发病率
试验组	20 万	28/10 万
对照组	20 万	71/10 万

试验结果显示, 疫苗将小儿麻痹症的发病率从 10 万分之 71 降低到 10 万分之 28. 由于 71 和 28 的差别超出了随机性本身所能解释的范围, 所以宣布疫苗是成功的. 进一步的分析指出, 可以以近 100% 的概率保证疫苗是有效的 (见后面的显著性检验一节).

随机对照实验

上例中的安慰剂是注射生理盐水, 给对照组的儿童使用安慰剂是为了避免儿童的心理作用影响试验的结果. 尽管可以认为仅靠精神作用不能抵抗小儿麻痹症, 但是为了确认试验结果的可靠性, 使用安慰剂是必要的.
上例中的随机对照试验是双盲的. 双盲的之一是指儿童自己不知道自己是在试验组还是在对照组, 也就是说不知道自己被注射的是疫苗还是生理盐水 (安慰剂), 甚至不知道有安慰剂, 这就有效地避免了潜在的心理影响. 另外一盲是指医生不了解他诊断的患者是在对照组还是在试验组, 这就避免了医生对疫苗的主观看法带来的可能影响. 在可能的场合, 随机对照双盲试验可以最大程度地避免心理因素的影响.
在许多场合, 心理因素是不能忽视的. 有资料显示在医院中给那些手术后产生剧痛的患者服用用淀粉制成的" 止痛片" 后, 大约有 1/3 的患者感觉剧痛减轻.

练习: 用 s_x^2 表示 $x_1, \ldots, x_n$ 的样本方差, 用 b 表示常数, 用 s_y^2 表示 $y_1, \ldots, y_n$ 的样本方差. 当 $y_1 = x_1 + b, y_2 = x_2 + b, \ldots, y_n = x_n + b$ 时, 验证 $s_y^2 = s_x^2$.

练习: 某学苑超市销售部收到甲乙两厂送来的质地相同的可乐各 10 瓶, 测量后得到甲乙两厂可乐的净含量 (单位:ml) 分别是:

甲厂	501	500	499	500	502	500	500	501	499	498
乙厂	497	501	500	502	499	501	503	500	500	497

问: 销售部应当销售哪家的冰可乐.

练习: 用自己的话叙述什么是对照组、什么是试验组、什么是随机对照试验.

参数的估计方法

样本均值和样本方差

如果 X 是从总体中随机抽出的个体, 则 X 是随机变量,X 的分布就是总体的分布. 如果对总体进行有放回的抽样, 则得到独立同分布且和 X 同分布的随机变量 $X_1, X_2, \cdots, X_n$. 这时称 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的简单随机样本, 简称为总体 X 的样本.

在观测放射性钋 (pō) 放射 α 粒子的试验中, 用 X 表示 7.5s 内观测到的粒子数. 独立重复观测时, 用 X_i 表示第 i 次的观测结果, 则 $X_1, X_2, \cdots, X_n$ 独立同分布且和 X 同分布,$X_1, X_2, \cdots, X_n$ 是总体 X 的样本.

Def1.1 如果 $X_1, X_2, \cdots, X_n$ 独立同分布且和 X 同分布, 则称 X 是总体, 称 $X_1, X_2, \cdots, X_n$ 是总体 X 的样本, 称观测数据的个数 n 为样本量.

样本均值和样本方差

在实际问题中得到的总是简单随机样本 $X_1, X_2, \cdots, X_n$ 的观测值 $x_1, x_2, \cdots, x_n$, 人们也称 $x_1, x_2, \cdots, x_n$ 是总体 X 的样本. 在统计学中, 常常不把 $X_1, X_2, \cdots, X_n$ 与它们的观测值 $x_1, x_2, \cdots, x_n$ 严格区分, 这是为了符号使用的方便. 当对数据进行统计分析时, 用大写的 $X_1, X_2, \cdots, X_n$, 实际计算时更多地用小写的 $x_1, x_2, \cdots, x_n$.

在统计问题中, 总体 X 的分布形式往往是已知的. 例如重复测量一个物体的重量时, 认为总体 X 服从正态分布 $N(\mu, \sigma^2)$, 未知参数是 μ, σ^2, 问题是根据总体 X 的样本估计总体参数 μ, σ^2. 观测放射物钋放射 α 粒子时, 总体 X 服从泊松分布 $\mathcal{P}(\lambda)$, 未知参数是 λ, 问题是根据总体 X 估计参数 λ.

设 $X_1, \cdots, X_n$ 是总体 X 的样本,θ 是总体 X 的未知参数. 如果

$$g_n(x_1, x_2, \cdots, x_n) \tag{23}$$

是已知函数, 则称

$$\widehat{\theta}_n = g_n(X_1, X_2, \cdots, X_n) \tag{24}$$

是 θ 的估计量, 简称为估计(estimator). 换句话说, 估计或估计量是从观测数据 $X_1, X_2, \cdots, X_n$ 能够直接计算的量. 计算以后得到的值称为估计值. 估计量也称为统计量(statistic). 为了符号使用的简便, 以后会把统计量 $\widehat{\theta}_n$ 简写称 $\widehat{\theta}$, 于是

$$\widehat{\theta} \equiv \widehat{\theta}_n. \tag{25}$$

设 $\widehat{\theta}$ 是总体参数 θ 的估计, 作为随机变量 $X_1, X_2, \cdots, X_n$ 的函数, 估计量也是随机变量. 估计量是样本的函数. 关于估计量有以下的一些定义.

Def 1.2 设 $\widehat{\theta}$ 是 θ 的估计.

- 如果 $\mathrm{E}\widehat{\theta} = \theta$, 则称 $\widehat{\theta}$ 是 θ 的无偏估计.
- 如果当样本量 $n \to \infty$,$\widehat{\theta}$ 依概率收敛到 θ, 则称 $\widehat{\theta}$ 是 θ 的相合估计.
- 如果当样本量 $n \to \infty$,$\widehat{\theta}$ 以概率 1 收敛到 θ, 则称 $\widehat{\theta}$ 是 θ 的强相合估计.

由于以概率 1 收敛可以推出依概率收敛, 所以强相合估计一定是相合估计. 一个估计起码应当是相合的, 否则我们不知道这个估计有什么用, 也不知道它到底估计谁, 咱也不敢问.

样本均值

设总体均值 $\mu = \mathrm{E}X$ 存在,$X_1, X_2, \cdots, X_n$ 是总体 X 的样本. 均值 μ 的估计定义为

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}. \tag{26}$$

由于 $\overline{X}_n$ 是从样本计算出来的, 所以是样本均值. 样本均值 $\overline{X}_n$ 有如下的性质:

(1)$\overline{X}_n$ 是 μ 的无偏估计, 这是因为 $\mathrm{E}\overline{X}_n = \mu$;

(2)$\overline{X}_n$ 是 μ 的强相合估计, 从而是相合估计. 这是因为从 Kolmogorov 强大数律得到:

$$\lim_{n\to\infty} \overline{X}_n = \mu, a.s. \tag{27}$$

样本均值

给定总体 X 的样本 $X_1, X_2, \cdots, X_n$, 也可以用

$$\overline{X}_{n-1} = \frac{X_1 + X_2 + \cdots + X_{n-1}}{n-1} \tag{28}$$

估计总体均值 μ, 这时仍然有

$$\mathrm{E}\overline{X}_{n-1} = \mu, \lim_{n\to\infty} \overline{X}_{n-1} = \mu, a.s. \tag{29}$$

说明 $\overline{X}_{n-1}$ 也是 μ 的无偏估计和强相合估计.
但是因为少用了一个数据, 所以

$$\mathrm{var}(\overline{X}_{n-1}) = \frac{\sigma^2}{n-1} > \mathrm{var}(\overline{X}_n) = \frac{\sigma^2}{n} \tag{30}$$

说明在均方误差的意义下,$\overline{X}_{n-1}$ 没有 $\overline{X}_n$ 的估计精度高. 这时称 $\overline{X}_n$ 比 $\overline{X}_{n-1}$更有效.

样本方差

给定总体 X 的样本 $X_1, X_2, \cdots, X_n$, 以下用 $\widehat{\mu}$ 表示样本均值, 于是

$$\widehat{\mu} = \overline{X}_n. \tag{31}$$

总体方差 $\sigma^2 = \mathrm{var}(X)$ 的估计由

$$S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \widehat{\mu})^2 \tag{32}$$

定义. 由于 S^2 是由样本计算出来的, 所以称为样本方差.

样本方差

取定 j. 因为 $\mathrm{E}(X_j - \widehat{\mu}) = \mu - \mu = 0$, 所以从 $X_1, X_2, \cdots, X_n$ 的独立性得到

$$
\begin{aligned}
\mathrm{E}(X_j - \widehat{\mu})^2 &= \mathrm{var}\left[X_j - \frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \mathrm{var}\left[(1 - \frac{1}{n})X_j - \frac{1}{n}\sum_{i\neq j} X_i\right] \\
&= \left(\frac{n-1}{n}\right)^2 \sigma^2 + \frac{1}{n^2}\sum_{i\neq j}\sigma^2 \\
&= \left[\left(\frac{n-1}{n}\right)^2 + \frac{n-1}{n^2}\right]\sigma^2 = \frac{n-1}{n}\sigma^2.
\end{aligned}
\tag{33}
$$

样本方差

于是得到

$$\mathrm{E}S^2 = \frac{1}{n-1}\sum_{j=1}^{n}\mathrm{E}(X_j - \widehat{\mu})^2 = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2. \tag{34}$$

说明样本方差 S^2 是总体方差 σ^2 的无偏估计.

样本方差

利用强大数律得到 $\widehat{\mu} \to \mu, a.s.$ 和

$$\frac{1}{n-1}\sum_{j=1}^{n} X_j^2 \to \mathrm{E}X^2, a.s. \tag{35}$$

于是有

$$\begin{aligned}\frac{1}{n-1}\sum_{j=1}^{n}(X_j-\widehat{\mu})^2 &= \frac{1}{n-1}\left(\sum_{j=1}^{n} X_j^2 - n\widehat{\mu}^2\right) \\ &\to \mathrm{E}X^2 - \mu^2 = \sigma^2, a.s.\end{aligned} \tag{36}$$

说明样本方差 S^2 是总体方差 σ^2 的强相合估计.

样本标准差

由于 S^2 是 σ^2 的估计, 所以定义标准差 σ 的估计为

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_j - \widehat{\mu})^2}, \tag{37}$$

称 S 为样本标准差. 由于 $S^2 \to \sigma^2, a.s.$, 所以 $S \to \sigma, a.s.$ 成立, 说明 S 是 σ 的强相合估计.

样本标准差

当 $\sigma > 0$ 时,S 不是 σ 的无偏估计, 也就是说 $\mathrm{E}S = \sigma$ 不成立. 这是因为没有不全为零的常数 a, b 使得 $P(aS + b = 0) = 1$, 所以由内积不等式得到

$$\mathrm{E}S = \mathrm{E}(S \cdot 1) < \sqrt{\mathrm{E}S^2 \cdot \mathrm{E}1^2} = \sqrt{\sigma^2} = \sigma. \tag{38}$$

于是 $\mathrm{E}S < \sigma$, 这时称 S 低估了 σ.

样本均值和样本方差

我们把上面的结果总结如下:

Thm 1.1 设 $X_1, \cdots, X_n$ 是总体 X 的样本, $\mu = \mathrm{E}X$, $\sigma^2 = \mathrm{var}(X) > 0$.

- 样本均值 $\overline{X}_n$ 是总体均值 μ 的强相合无偏估计;
- 样本方差 S^2 是总体方差 σ^2 的强相合无偏估计;
- 样本标准差 S 是总体标准差 σ 的强相合估计, 但是 $\mathrm{E}S < \sigma$.

样本均值和样本方差

例 设 $X_1, \cdots, X_n$ 是总体 X 的样本. 当 $\mu_k = \mathrm{E}X^k$ 存在时, 试给出 μ_k 的强相合无偏估计.

解 因为 $X_1^k, \cdots, X_n^k$ 独立同分布, 且和 X^k 同分布, 所以是总体 X^k 的样本, 并且

$$\widehat{\mu}_k = \frac{1}{n}\sum_{j=1}^{n} X_j^k \tag{39}$$

是 μ_k 的估计. 从 Thm1.1 知道 $\widehat{\mu}_k$ 是 μ_k 的强相合无偏估计.

在例 1.1 中, 称 $\mu_k = \mathrm{E}X^k$ 为总体 X 的 k 阶原点矩, 称 $\widehat{\mu}_k$ 为 k 阶样本原点矩.

样本均值和样本方差

在实际数据的计算中，常用 $\overline{x}_n, s^2$ 和 s 分别表示样本均值、样本方差和样本标准差:

$$\overline{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j, s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x}_n)^2, s = \sqrt{s^2}. \tag{40}$$

如果用 $x_1, x_2, \cdots, x_n$ 表示总体 X 的样本的一次观测，则 Thm1.1 保证了如下事实:

$$\lim_{n\to\infty} \overline{x}_n = \mu, \lim_{n\to\infty} s^2 = \sigma^2, \lim_{n\to\infty} s = \sigma. \tag{41}$$

样本均值和样本方差

为了了解样本均值的表现，下面用计算机产生 10^7 个来自正态总体 $N(\mu, \sigma^2)$ 的样本 $x_1, x_2, \cdots, x_n$，其中真值 $\mu = 1.8, \sigma^2 = 0.04, \sigma = 0.2$. 利用前面公式和前 n 个观测数据计算的 $\overline{x}_n, s^2, s$ 如下:

n	10	10^2	10^3	10^4	10^5	10^6	10^7
$\overline{x}_n$	1.7679	1.8196	1.8039	1.8006	1.7987	1.7997	1.7999
s^2	0.0452	0.0377	0.0407	0.0398	0.0402	0.0400	0.0400
s	0.1044	0.1942	0.2019	0.1997	0.2006	0.2001	0.2000

计算结果支持强相合结论:$\overline{x}_n \to \mu, s^2 \to \sigma^2, s \to \sigma$.

有一天, 有一个警察在巡逻, 警察看到有一个人在街上东张西望, 他就上去问他:" 先生, 你需要我帮助么?", 这个人说," 我的手表掉了, 所以我在找我的手表". 于是这个警察就帮帮这个人一起找. 这个警察帮他认真看了半天, 一点也没看到有手表在街上. 这个警察后来说," 先森, 你能确定你的手表就在这个附近掉的么?". 这个人说:" 我的手表不是在这里掉的"." 你的手表在哪里掉的?"." 在那边那条街上掉的"." 那你在这里找干什么?"." 那边的街没有路灯, 黑漆麻乌, 什么也看不见, 更别说找了; 这边路灯很亮, 什么都好, 所以我在这里找."

这个例子告诉我们, 统计模型就是为了处理方便建立得简简单单, 不与实际问题建立联系, 仅仅为了做出来很快, 结果很" 显著"、很" 漂亮", 实际上是没有用的.

矩估计

矩估计方法由 K.Pearson 在 1894 年正式提出. 矩估计的理论根据是大数定律. 设 $X_1, X_2, \cdots, X_n$ 是总体 X 的样本. 通常,

$$\mu_k = \mathrm{E}X^k \tag{42}$$

称为 k 阶总体 (原点) 矩, 而

$$\widehat{\mu}_k = \frac{1}{n}\sum_{j=1}^{n} X_j^k \tag{43}$$

称为 k 阶样本 (原点) 矩.
根据大数定律, 可以用各阶样本矩去估计相应的总体矩. 由此而得矩估计(moment estimator) 这个名称.

矩估计

Def 2.2 设 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的一个样本 (总体的未知参数是 θ). 若涉及的矩存在, 则

- k 阶样本矩 $\widehat{\mu_k} = \frac{1}{n}\sum_{j=1}^{n} X_j^k$ 为相应的总体矩 $\mu_k = \mathrm{E}X^k$ 的矩估计.
- 若存在连续函数 ϕ 使 $g(\theta) = \phi(\mu_1, \mu_2, \cdots, \mu_k)$ 成立, 则 $g(\theta)$ 的矩估计就定义成 $\widehat{g(\theta)} = \phi(\widehat{\mu_1}, \widehat{\mu_2}, \cdots, \widehat{\mu_k})$

为了方便理解这个定义, 下面通过举例介绍参数的矩估计.

例 某院在一个月中组织了 12 次讲座, 讲座的听众数依次如下:

169 183 167 157 163 151 154 157 163 154 162 165.

如果每次讲座的听众数相互独立, 都服从泊松分布 $\mathcal{P}(\lambda)$, 试估计参数 λ.

解 用 X 表示服从泊松分布 $\mathcal{P}(\lambda)$ 的随机变量, 则

$$\lambda = \mathrm{E}X = \mu_1. \tag{44}$$

因为 μ_1 的矩估计是 $\widehat{\mu}_1$, 所以 λ 的矩估计也是 $\widehat{\mu}_1$. 将上面的数据代入 $\widehat{\mu}_1$, 得到 λ 的估计

$$\widehat{\lambda} = \widehat{\mu}_1 = \frac{1}{12}\sum_{j=1}^{12} x_j = 162.083. \tag{45}$$

本例中称 $\widehat{\lambda}$ 为 λ 的矩估计.$\widehat{\lambda}$ 正是这 12 次讲座的平均听众数.

例 键盘侠甲在论坛上发帖后就开始观察跟帖情况. 在 1 个小时内, 他记录了以下的跟帖间隔时间 (单位:s):

1	2	9	33	28	62	17	46	1	0.12
35	11	53	33	2	15	62	7	18	81
63	20	18	40	22	32	126	145	47	176
38	15	96	135	19	20	67	166	67	21

假设跟帖时间的间隔时间服从指数分布 $\mathcal{E}(\lambda)$, 试估计参数 λ.

解 设 X 服从指数分布 $\mathcal{E}(\lambda)$. 由 $\mu = \mathrm{E}X = 1/\lambda$ 得到 $\lambda = 1/\mu$. 因为 $\widehat{\mu}_1$ 是 μ_1 的矩估计, 于是可以用

$$\widehat{\lambda} = 1/\widehat{\mu}_1 = 1/\overline{x}_n. \tag{46}$$

估计参数 λ. 经计算得到

$$\overline{x}_n = 46.525, \widehat{\lambda} = 0.0215. \tag{47}$$

在本例中, 称 $\widehat{\lambda} = 0.00215$ 为参数 λ 的矩估计.

例 单晶硅太阳能电池以高纯度单晶硅棒棒为原料. 制作时需要将单晶硅棒棒进行切片, 每片的厚度在 0.3mm 左右. 现在用随机抽样的方法测量了某厂家的 n 片单晶硅的厚度, 得到测量数据 $x_1, x_2, \cdots, x_n$. 假设这批单晶硅厚度的总体分布是正态分布, 试估计这批单晶硅的总体均值和总体方差.

解 设样本 $x_1, x_2, \cdots, x_n$ 来自总体 X, 则 $X \sim N(\mu, \sigma^2)$, 并且

$$\mu = \mathrm{E}X, \sigma^2 = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \mu_2 - \mu^2, \tag{48}$$

因为 $\widehat{\mu}_1, \widehat{\mu}_2$ 分别是 μ,μ_2 的矩估计, 于是分别用

$$\widehat{\mu} = \widehat{\mu}_1, \widehat{\sigma}^2 = \widehat{\mu}_2 - \widehat{\mu}_1^2 \tag{49}$$

估计 μ, σ^2.

在本例中, 称 $\widehat{\mu}$ 为 μ 的矩估计, 称 $\widehat{\sigma}^2$ 为 σ^2 的矩估计. 容易看出,$\widehat{\mu}$ 就是样本均值. 但是从

$$\begin{aligned}\widehat{\sigma}^2 &= \widehat{\mu}_2 - \widehat{\mu}_1^2 \\ &= \frac{1}{n}\sum_{j=1}^{n} x_j^2 - \overline{x}_n^2 \\ &= \frac{1}{n}\sum_{j=1}^{n}(x_j - \widehat{\mu})^2\end{aligned} \tag{50}$$

知道 σ^2 的矩估计 $\widehat{\sigma}^2$ 比样本方差 s^2 略小.

从上面的例子看出, 如果总体 X 的分布函数 $F(x;\theta)$ 只有一个未知函数 θ, 则 $\mu_1 = \mathrm{E}X$ 和 θ 有关. 如果能从

$$\mu_1 = \mathrm{E}X \text{得到} \theta = \phi(\mu_1), \tag{51}$$

其中的 ϕ 是已知函数, 则 $\widehat{\theta} = \phi(\widehat{\mu}_1)$ 是 θ 的矩估计, 其中 $\widehat{\mu}_1$ 是样本均值.

如果总体 X 的分布函数 $F(x;\theta_1,\theta_2)$ 有两个未知参数 θ_1,θ_2, 则 $\mu_1 = \mathrm{E}X$ 和 $\mu_2 = \mathrm{E}X^2$ 都和 θ_1,θ_2 有关. 如果能从

$$\begin{cases} \mu_1 = \mathrm{E}X, \\ \mu_2 = \mathrm{E}X^2 \end{cases} \text{得到} \begin{cases} \theta_1 = g_1(\mu_1,\mu_2), \\ \theta_2 = g_2(\mu_1,\mu_2) \end{cases} \tag{52}$$

其中 g_1, g_2 是已知函数, 则

$$\widehat{\theta}_1 = g_1(\widehat{\mu}_1,\widehat{\mu}_2), \widehat{\theta}_2 = g_2(\widehat{\mu}_1,\widehat{\mu}_2) \tag{53}$$

分别是 θ_1,θ_2 的矩估计. 其中的 $\widehat{\mu}_1,\widehat{\mu}_2$ 分别是 1,2 阶样本矩.

例 设 $X_1, X_2, \cdots, X_n$ 是总体 $\mathcal{U}[a,b]$ 的样本, 其中 a, b 是未知参数. 求 a, b 的矩估计.

解 设 X 在 $[a,b]$ 中均匀分布, 则 X 是所述的总体. 因为 X 的密度关于 $(a+b)/2$ 对称, 所以

$$\mu_1 = \mathrm{E}X = \frac{a+b}{2}. \tag{54}$$

容易计算出 X 的方差为

$$\mathrm{var}(X) = \frac{(b-a)^2}{12}. \tag{55}$$

由公式 $\mathrm{var}(X) = \mathrm{E}X^2 - (\mathrm{E})^2$ 得到

$$\mu_2 - \mu_1^2 = \frac{(b-a)^2}{12}. \tag{56}$$

解得:

$$a = \mu_1 - \sqrt{3(\mu_2 - \mu_1^2)}, b = \mu_1 + \sqrt{3(\mu_2 - \mu_1^2)} \tag{57}$$

于是,a, b 的矩估计分别为

$$\widehat{a} = \widehat{\mu}_1 - \sqrt{3(\widehat{\mu}_2 - \widehat{\mu}_1^2)}, \widehat{b} = \mu_1 + \sqrt{3(\widehat{\mu}_2 - \widehat{\mu}_1^2)} \tag{58}$$

为了了解这个矩估计的表现, 用计算机产生 10^7 个独立同分布的都在 $(0.8, 5.2)$ 中均匀分布的随机变量的观测值, 利用该式和前 n 个观测数据计算的矩估计如下:

n	10	10^2	10^3	10^4	10^5	10^6	10^7
$\widehat{a}$	1.0336	0.9891	0.8053	0.8124	0.8001	0.8016	0.8001
$\widehat{b}$	5.4547	5.2623	5.2132	5.2330	5.2067	5.2018	5.1997

计算结果支持强相合的结论:$\widehat{a} \to 0.8, \widehat{b} \to 5.2.$

根据上面的讨论, 可以给出矩估计的一般定义.Rmk: 此处能用低阶矩就不去用高阶矩.

Def 2.2* 设 X 的分布含有参数 $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)$,$X_1, \cdots, X_n$ 是总体 X 的样本.

- 如果能得到表达式

$$\begin{cases} \theta_1 = g_1(\mu_1, \cdots, \mu_m), \\ \cdots\cdots\cdots \\ \theta_m = g_m(\mu_1, \cdots, \mu_m) \end{cases} \text{则称由} \begin{cases} \widehat{\theta}_1 = g_1(\widehat{\mu}_1, \cdots, \widehat{\mu}_m), \\ \cdots\cdots\cdots \\ \widehat{\theta}_m = g_m(\widehat{\mu}_1, \cdots, \widehat{\mu}_m) \end{cases} \tag{59}$$

  定义的 $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \cdots, \widehat{\theta}_m)$ 为 $\boldsymbol{\theta}$ 的矩估计.
- 由于总体 X 的分布中含有 $\boldsymbol{\theta}$ 的信息, 所以 μ_k 往往是 $\boldsymbol{\theta}$ 的函数, 上面的矩估计通常可由如下的估计方程解出.

估计方程

$$\begin{cases} \mu_1 = h_1(\theta_1, \cdots, \theta_m), \\ \cdots\cdots\cdots \\ \mu_m = h_m(\theta_1, \cdots, \theta_m) \end{cases} \tag{60}$$

因为矩估计没有充分利用总体分布的信息, 所以在已知 X 的分布信息时, 矩估计一般不如下面给出的极大似然估计来得好.

极大似然估计

A、离散分布的情况

如果袋中有红球和黑球共三个, 现从中任取一个, 得到红球. 你会判断袋中有 2 个红球,1 个黑球. 这是因为当红球多于黑球时, 才更可能取得红球.

若用 A 表示取得红球, 用 p 表示袋中红球的比例, 则 $P(A) = p$. 从已知条件知道 $p = 2/3$ 或 $p = 1/3$. 因为现在 A 发生了, 所以判断 $p = 2/3$. 这时称 $\widehat{p} = 2/3$ 为 p 的最大似然估计. 这种思考问题的方法被称为最大似然方法.

极大似然估计

例 yy 和 HL 两人下棋, 用 p 表示 yy 在每局中获胜的概率. 如果 5 局中 yy 胜了 3 局, 你会判断 $p > 1/2$. 这是因为当 $p > 1/2$ 时,yy 才更可能 5 局 3 胜. 但是 p 到底应当是多大呢？让我们把这个问题数学化.
设 5 局中 yy 胜 X 局, 并且假设各局的胜负相互独立, 则 $P(X = 3) = \binom{5}{3} p^3 (1-p)^2$. 现在已知 $X = 3$, 所以 p 应当使 $X = 3$ 发生的概率

$$L(p) = \binom{5}{3} p^3 (1-p)^2$$

达到最大. 对于 $L(p)$ 求导数, 令

$$L'(p) = \binom{5}{3} \left[3p^2(1-p)^2 - 2p^3(1-p)\right] = \binom{5}{3} p^2(1-p)[3(1-p)-2p] = 0.$$

得到 $p = 3/5 (p = 0, 1$ 不合题意). 这时称 $\widehat{p} = 3/5$ 为 p 的最大似然估计.

极大似然估计

Def 3.1　设离散随机变量 $X_1, X_2, \cdots, X_n$ 有联合分布

$p(x_1, x_2, \cdots, x_n; \boldsymbol{\theta}) = P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$, 其中 $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)$ 是未知参数.

- 给定观测数据 $x_1, x_2, \cdots, x_n$ 后, 称 $\boldsymbol{\theta}$ 的函数

$$L(\boldsymbol{\theta}) = p(x_1, x_2, \cdots, x_n; \boldsymbol{\theta}) \tag{61}$$

  为似然函数

- 称 $L(\boldsymbol{\theta})$ 的最大值点 $\widehat{\boldsymbol{\theta}}$ 为 $\boldsymbol{\theta}$ 的最大似然估计.

极大似然估计

最大似然估计常常被缩写成 MLE(maximum likelihood estimator), 并且用

$$\arg\sup L(\boldsymbol{\theta}) \tag{62}$$

表示. 因为 $\ln x$ 是严格单调的增函数, 所以 $l(\boldsymbol{\theta}) = \ln(L(\boldsymbol{\theta}))$ 和 $L(\boldsymbol{\theta})$ 有相同的最大值点. 通常称 $l(\boldsymbol{\theta})$ 为对数似然函数, 称

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_j} = 0, j = 1, 2, \cdots, m. \tag{63}$$

为似然方程 (组). 在许多情况下, 最大似然估计可以由似然方程组解出. 注意, 在此定义中,$x_1, x_2, \cdots, x_n$ 是观测数据, 不再变动, 所以 $L(\boldsymbol{\theta})$ 和 $l(\boldsymbol{\theta})$ 都是 $\boldsymbol{\theta}$ 的函数.

极大似然估计

例 设 $X_1, X_2, \cdots, X_n$ 独立同分布, 都服从泊松分布 $\mathcal{P}(\lambda)$.
(1) 给定观测 $X_1 = 169$, 计算 λ 的 MLE;
(2) 给定 $X_1, X_2, \cdots, X_{12}$ 的观测值:
169 167 157 196 163 151 154 157 163 154 162 165,
计算 λ 的 MLE.
解 (1)X_1 有概率分布

$$P(X_1 = x) = \frac{\lambda^x}{x!} e^{-\lambda}. \tag{64}$$

对数似然函数为

$$l(\lambda) = \ln L(\lambda) = x \ln \lambda - \lambda - \ln(x!). \tag{65}$$

解似然方程 $l'(\lambda) = x/\lambda - 1 = 0$, 得到 MLE 为 $\widehat{\lambda} = x = 169$.

极大似然估计

(2)λ 的似然函数为

$$\begin{aligned} L(\lambda) &= P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) \\ &= \frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\frac{\lambda^{x_2}}{x_2!}e^{-\lambda}\cdots\frac{\lambda^{x_n}}{x_n!}e^{-\lambda} \\ &= \frac{\lambda^{x_1+x_2+\cdots+x_n}}{x_1!x_2!\cdots x_n!}e^{-n\lambda}. \end{aligned} \tag{66}$$

对数似然函数为

$$l(\lambda) = \ln L(\lambda) = (x_1 + x_2 + \cdots + x_n)\ln\lambda - n\lambda - c_0, \tag{67}$$

其中的 c_0 和参数 λ 无关, 解似然方程 $l'(\lambda) = 0$, 得到 λ 的 MLE 为

$$\widehat{\lambda} = \frac{x_1 + x_2 + \cdots + x_n}{n} = 163.167. \tag{68}$$

例 设 $X_1, X_2, \cdots, X_n \overset{i.i.d}{\sim} \mathcal{B}(1,p)$, 给定样本观测值 $x_1, x_2, \cdots, x_n$ 后, 计算 p 的 MLE.

解 因为 $P(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}$, 所以 p 的似然函数为:

$$\begin{aligned} L(p) &= P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) \\ &= P(X_1 = x_1)P(X_2 = x_2)\cdots P(X_n = x_n) \\ &= p^{n\overline{x}_n}(1-p)^{n-n\overline{x}_n} \end{aligned} \tag{69}$$

对数似然函数为

$$l(p) = \ln L(p) = n\overline{x}_n \ln p + (n - n\overline{x}_n)\ln(1-p). \tag{70}$$

解似然方程:

$$l'(p) = n\overline{x}_n/p + (n - n\overline{x}_n)/(1-p) = 0, \tag{71}$$

得到 p 的 MLE 为:

$$\widehat{p} = \overline{x}_n. \tag{72}$$

B、连续分布的情况

例 设 $X \sim N(\mu, 1)$, 给定 $X = x$, 试通过极大似然估计 μ.

分析 根据最大似然的思路,μ 应当使 $\{X = x\}$ 发生的概率

$$P(X = x) = f(x; \mu)\mathrm{d}x \tag{73}$$

达到最大.

解 μ 应当使

$$f(x; \mu) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2}. \tag{74}$$

达到最大. 于是得到 μ 的极大似然估计是 $\widehat{\mu} = x$.

Def 3.2 设随机向量 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 有联合密度 $f(\boldsymbol{x}; \boldsymbol{\theta})$, 其中 $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)$ 是未知参数.

- 得到 $\boldsymbol{X}$ 的观测值 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ 后, 称

$$L(\boldsymbol{\theta}) = f(\boldsymbol{x}; \boldsymbol{\theta}) \tag{75}$$

  为 $\boldsymbol{\theta}$ 的似然函数.
- 称 $L(\boldsymbol{\theta})$ 的最大值点 $\widehat{\boldsymbol{\theta}}$ 为 $\boldsymbol{\theta}$ 的最大似然估计(MLE).

设总体 X 有概率密度 $f(x;\boldsymbol{\theta})$,$X_1, X_2, \cdots, X_n$ 是总体 X 的样本, 则 $(X_1, X_2, \cdots, X_n)$ 的联合密度是

$$f(x_1, x_2, \cdots, x_n; \boldsymbol{\theta}) = \prod_{j=1}^{n} f(x_j; \boldsymbol{\theta}), \tag{76}$$

基于观测值 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ 的似然函数是

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{n} f(x_j; \boldsymbol{\theta}). \tag{77}$$

由于对数似然函数 $l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta})$ 和 $L(\boldsymbol{\theta})$ 有相同的最大值点. 许多问题中, 求 $L(\boldsymbol{\theta})$ 的最大值可以转化为解似然方程 (组)

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_j} = 0, j = 1, 2, \cdots, m. \tag{78}$$

例 设 $x_1, x_2, \cdots, x_n$ 是总体 $\mathcal{E}(\lambda)$ 的样本, 求 λ 的 MLE. **解** 因为指数分布 $\mathcal{E}(\lambda)$ 的概率密度是

$$f(x;\lambda) = \lambda e^{-\lambda x}\mathbf{1}_{(x>0)}. \tag{79}$$

基于观测值 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ 的似然函数是

$$L(\lambda) = \lambda^n \exp\left(-\lambda \sum_{j=1}^{n} x_j\right). \tag{80}$$

对数似然函数是

$$l(\lambda) = n \ln \lambda - \lambda \sum_{j=1}^{n} x_j. \tag{81}$$

由

$$\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum_{j=1}^{n} x_j = 0, \tag{82}$$

得到参数 λ 的 MLE 是 $\widehat{\lambda} = 1/\overline{x}_n$.

例 设 $X_1, X_2, \cdots, X_n$ 是正态总体 $N(\mu, \sigma^2)$ 的样本. 给定观测数据 $x_1, x_2, \cdots, x_n$, 求 μ, σ^2 的 MLE. **解** 不妨引入 $a = \sigma^2$.
因为正态分布 $N(\mu, a)$ 的概率密度是

$$f(x; \mu, a) = \frac{1}{\sqrt{2\pi a}} \exp\left[-\frac{(x-\mu)^2}{2a}\right]. \tag{83}$$

所以基于观测值 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ 的似然函数是

$$L(\mu, a) = \frac{1}{(\sqrt{2\pi a})^n} \exp\left[-\sum_{j=1}^{n} \frac{(x_j-\mu)^2}{2a}\right]. \tag{84}$$

对数似然函数是

$$l(\mu, a) = \frac{n}{2} \ln a - \sum_{j=1}^{n} \frac{(x_j-\mu)^2}{2a} + c. \tag{85}$$

其中,c 是常数. 求 $l(\mu, a)$ 的最大值点可以通过解似然方程组

$$\begin{cases} \dfrac{\partial l}{\partial \mu} = \dfrac{1}{a}\displaystyle\sum_{j=1}^{n}(x_j - \mu) = 0 \\ \dfrac{\partial l}{\partial a} = -\dfrac{n}{2a} + \dfrac{1}{2a^2}\displaystyle\sum_{j=1}^{n}(x_j - \mu)^2 = 0 \end{cases} \tag{86}$$

得到, 从该方程组解得 μ, σ^2 的 MLE 为

$$\begin{cases} \widehat{\mu} = \overline{x}_n, \\ \widehat{\sigma}^2 = \widehat{a} = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n}(x_j - \overline{x}_n)^2. \end{cases} \tag{87}$$

从该例中看到, 对于正态总体来讲,MLE 和矩估计也是一致的.

例 设 $x_1, x_2, \cdots, x_n$ 是总体 $\mathcal{U}[a,b]$ 的样本, 求 a, b 的 MLE. **解** 均匀分布 $\mathcal{U}[a,b]$ 的概率密度函数是

$$f(x;a,b) = \frac{1}{b-a}\mathbf{1}_{(a\leqslant x\leqslant b)}, \tag{88}$$

其中,$\mathbf{1}_A$ 是示性函数:

$$\mathbf{1}_A = \begin{cases} 1, A\text{发生}, \\ 0, A\text{不发生}. \end{cases} \tag{89}$$

给定观测数据 $x_1, x_2, \cdots, x_n$ 定义

$$x_{(1)} = \min\{x_1, x_2, \cdots, x_n\}, x_{(n)} = \max\{x_1, x_2, \cdots, x_n\}. \tag{90}$$

可以把 a, b 的似然函数写成:

$$
\begin{aligned}
L(a,b) &= \frac{1}{(b-a)^n}\prod_{j=1}^{n}\mathbf{1}_{(a\leqslant x_j\leqslant b)} \\
&= \frac{1}{(b-a)^n}\mathbf{1}_{(a\leqslant x_{(1)}\leqslant x_{(n)}\leqslant b)}.
\end{aligned}
\tag{91}
$$

要使 $L(a,b)$ 达到最大, 首先要使示性函数 $\mathbf{1}_{(a\leqslant x_{(1)}\leqslant x_{(n)}\leqslant b)}=1$, 这也就是说 $a\leqslant x_{(1)}\leqslant x_{(n)}\leqslant b$. 然后, 再要求 $\frac{1}{(b-a)^n}$ 最大, 不难看出,

$$
\widehat{a}=x_{(1)},\widehat{b}=x_{(n)}. \tag{92}
$$

小练习: 在本题中, 证明 $\widehat{a}$ 与 $\widehat{b}$ 都不是无偏估计
(Hint: $\mathrm{E}X_{(1)}>a$,$\mathrm{E}X_{(n)}<b$.)

练习: 设 $Y_1, Y_2, \cdots, Y_n$ 是来自**对数正态分布**的样本, 有共同的概率密度

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right) \tag{93}$$

计算参数 μ, σ^2 的最大似然估计.
Ans:

$$\widehat{\mu} = \frac{1}{n}\sum_{j=1}^{n} \ln y_j, \widehat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n} (\ln y_j - \widehat{\mu})^2. \tag{94}$$

练习: 设 $Y_1, Y_2, \cdots, Y_n$ 是来自**几何分布** $\mathrm{Ge}(p)$ 的样本, 计算参数 p 的矩估计和极大似然估计. Ans:

$$1/\overline{X}_n, 1/\overline{X}_n. \tag{95}$$

练习: 举例说明矩估计、极大似然估计都不一定唯一.

抽样分布

抽样分布

χ^2 分布、t 分布、F 分布是统计学中最重要的抽样分布，被称为统计学中的三大分布. 本节的任务是推导这三个分布.
抽样分布是统计推断的基础，其理论证明是概率统计课中不可绕过的一个难点.
(1)χ^2 分布
设 $X_1, ..., X_n$ 是总体 $X \sim N(0,1)$ 的样本，则平方和

$$\xi_n^2 = X_1^2 + ... + X_n^2 \tag{96}$$

服从 n 个自由度的 χ^2 分布，有概率密度

$$f_n(z) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} \mathbf{1}_{(z>0)} \tag{97}$$

记作: $\xi_n^2 \sim \chi^2(n)$

证明: 往证 $Y_k = X_k^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$.
注意到: $\{Y_k = y\} = \{X_k = \sqrt{y}\} \cup \{X_k = -\sqrt{y}\}$,
故 Y_k **的密度函数是**:

$$\begin{aligned} f_{Y_k}(y) &= \varphi(\sqrt{y})\frac{1}{2\sqrt{y}} + \varphi(-\sqrt{y})\frac{1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi y}} e^{-y/2} \mathbf{1}_{(y>0)}. \end{aligned} \tag{98}$$

这正是 $\Gamma(\frac{1}{2}, \frac{1}{2})$ **的概率密度函数, 故** $Y_k = X_k^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$.
再由 Γ **分布的可加性知道**:

$$\xi_n^2 \sim \Gamma(\frac{n}{2}, \frac{1}{2}) \tag{99}$$

因此, 它有概率密度函数 $f_n(z) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} \mathbf{1}_{(z>0)}$

另解: 先求分布函数, 之后对分布函数求导得概率密度.
Y_k 有分布函数

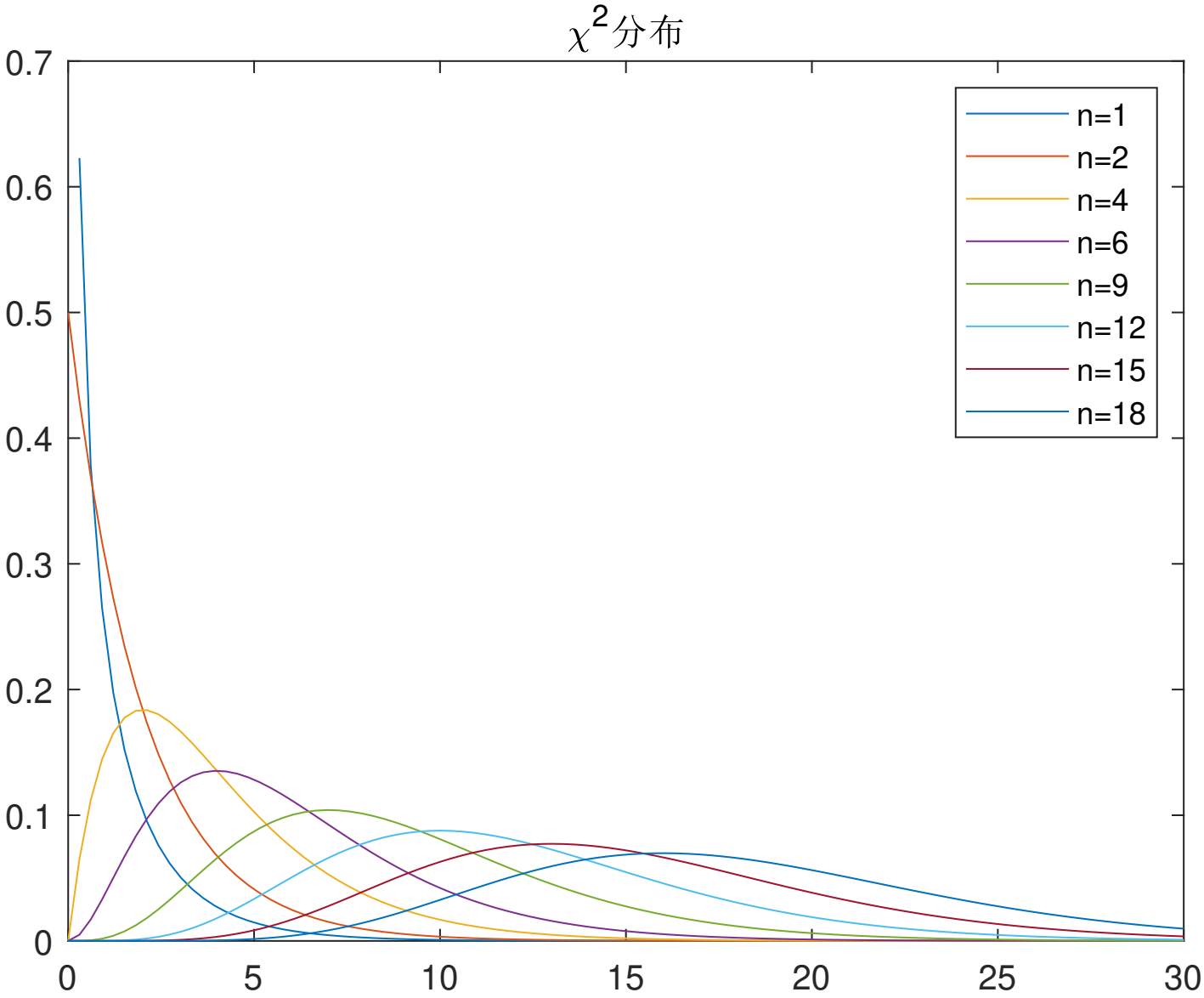
$$\begin{aligned} F_{Y_k}(y) &= P(X_k^2 \leqslant y) = P(-\sqrt{y} \leqslant X_k \leqslant \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1. \end{aligned} \tag{100}$$

因此,

$$\begin{aligned} f_{Y_k}(y) &= F'_{Y_k}(y) \\ &= 2 \times \frac{1}{2\sqrt{y}} \varphi(\sqrt{y}) \\ &= \frac{1}{\sqrt{2\pi y}} e^{-y/2} \mathbf{1}_{(y>0)}. \end{aligned} \tag{101}$$

下同第一种解法.

χ^2 分布


χ^2分布
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
n=1
n=2
n=4
n=6
n=9
n=12
n=15
n=18

(2)t 分布

设随机变量 $X \sim N(0,1), Y \sim \chi^2(n)$, 其中 X, Y 独立, 则随机变量

$$T = \frac{X}{\sqrt{Y/n}} \tag{102}$$

服从 n 个自由度的 t 分布, 并且有概率密度

$$f_T(t) = a_n \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \tag{103}$$

其中 $a_n = \frac{1}{\sqrt{n}\mathrm{B}(n/2,1/2)}$ 是归一化的系数.

t 分布

证明:(法一) 增补变量 $S = Y$, 来求 (T, S) 的联合概率密度函数 $g(t, s)$.
(X, Y) **有联合密度函数** $f(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2} \mathbf{1}_{(y>0)}$
注意到: $\{(T, S) = (t, s)\} = \{(X, Y) = (t\sqrt{s/n}, s)\}$,
Jacobi **阵为:**$J = \begin{pmatrix} \sqrt{s/n} & * \\ 0 & 1 \end{pmatrix}$,$|\det(J)| = \sqrt{s/n}$, **故有:**

$$\begin{aligned} g(t, s) &= f(t\sqrt{s/n}, s)\sqrt{s/n} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2 s}{2n}} \frac{1}{2^{n/2}\Gamma(n/2)} s^{n/2-1} e^{-s/2} \mathbf{1}_{(s>0)} \sqrt{s/n} \\ &= c_n s^{\frac{n-1}{2}} \exp\left(-\frac{1}{2}(\frac{t^2}{n} + 1)s\right) \mathbf{1}_{(s>0)} \end{aligned} \tag{104}$$

因此,T **的概率密度为** $f_T(t) = \int_0^\infty g(t, s)\mathrm{d}s$

为计算该式, 替换 $u = \frac{1}{2}(\frac{t^2}{n} + 1)s$, 便有:

$$
\begin{aligned}
f_T(t) &= \int_0^{\infty} g(t,s)\mathrm{d}s \\
&= c_n \int_0^{\infty} s^{\frac{n-1}{2}} \exp\left(-\frac{1}{2}(\frac{t^2}{n}+1)s\right)\mathrm{d}s \\
&= 2^{\frac{n-1}{2}} c_n \left[\int_0^{\infty} u^{\frac{n-1}{2}} \exp(-u)\mathrm{d}s\right]\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \\
&= a_n \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}.
\end{aligned}
\tag{105}
$$

其中 $a_n = 2^{\frac{n-1}{2}} c_n \Gamma(\frac{n+1}{2}) = \frac{1}{\sqrt{2\pi}} \frac{2^{\frac{n-1}{2}}}{2^{n/2}\Gamma(n/2)} \frac{1}{\sqrt{n}} \Gamma(\frac{n+1}{2}) = \frac{1}{\sqrt{n}\mathrm{B}(n/2,1/2)}$.

a_n 的计算也可以由 $\int_{-\infty}^{\infty} f_T(t) = 1$ 得到:

$$\begin{aligned}\frac{1}{a_n} &= 2\int_0^{\infty}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}\mathrm{d}t \\ &= \sqrt{n}\int_0^1 s^{n/2-1}(1-s)^{1/2-1}\mathrm{d}s\left[\text{取变换}1+\frac{t^2}{n}=\frac{1}{s}\right] \\ &= \sqrt{n}\mathrm{B}(n/2,1/2)\end{aligned} \tag{106}$$

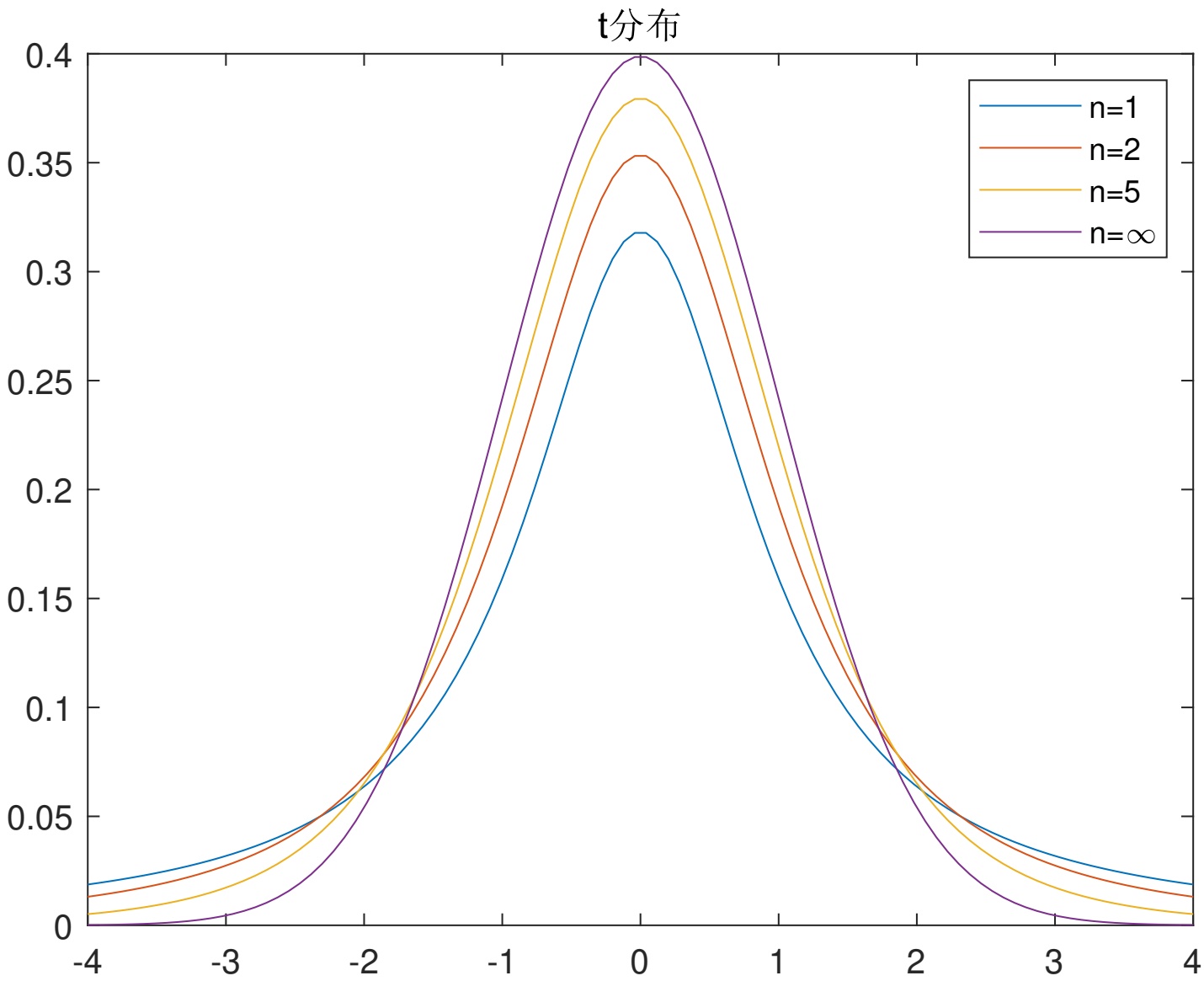
证明:(法二) 先求概率分布, 再求导得概率密度.
(X,Y) **有联合密度函数** $f(x,y)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\frac{1}{2^{n/2}\Gamma(n/2)}y^{n/2-1}e^{-y/2}\mathbf{1}_{(y>0)}$

$$
\begin{aligned}
F_T(t)=P(T\leqslant t)&=\iint\limits_{\{x\leqslant t\sqrt{y/n}\}} f(x,y)\mathrm{d}x\mathrm{d}y \\
&=\int_0^{\infty}\Phi(t\sqrt{y/n})\frac{1}{2^{n/2}\Gamma(n/2)}y^{n/2-1}e^{-y/2}\mathrm{d}y
\end{aligned}
\tag{107}
$$

对上式求导得到:

$$
\begin{aligned}
f_T(t)&=F_T'(t) \\
&=\int_0^{\infty}\varphi(t\sqrt{y/n})\sqrt{y/n}\frac{1}{2^{n/2}\Gamma(n/2)}y^{n/2-1}e^{-y/2}\mathrm{d}y \\
&=\frac{1}{\sqrt{2\pi n}2^{n/2}}\int_0^{\infty}y^{\frac{n-1}{2}}\exp\left(-\frac{1}{2}(\frac{t^2}{n}+1)y\right)\mathrm{d}y \quad \textbf{下同法一}
\end{aligned}
\tag{108}
$$

t 分布



F 分布, 注意: 分子的自由度在前

设 X, Y 独立, 分别服从自由度是 n, m 的 χ^2 分布, 则

$$Z = \frac{X/n}{Y/m} \tag{109}$$

服从自由度为 n, m 的 F 分布, 有概率密度

$$f_Z(z) = cz^{n/2-1}\left(1 + \frac{nz}{m}\right)^{-(n+m)/2} \mathbf{1}_{(z>0)}. \tag{110}$$

其中, 归一化常数 $c = (n/m)^{n/2}[\mathrm{B}(n/2, m/2)]^{-1}$.

F 分布, 注意: 分子的自由度在前

证明:(法一) 先求概率分布, 再求导得概率密度.
由于 $Z = \frac{m}{n}\frac{X}{Y}$, **故先计算** $U = \frac{X}{Y}$ **的密度** $f_U(u)$,
自然得到 Z **的密度** $f_Z(z) = \frac{n}{m} f_U(\frac{n}{m}z)$.
(X, Y) **有联合密度:**$f(x,y) = c_1 x^{n/2-1} e^{-x/2} y^{m/2-1} e^{-y/2} \mathbf{1}_{(x>0, y>0)}$

$$
\begin{aligned}
F_U(u) &= P(U \leqslant u) = P(X \leqslant uY) \\
&= c_1 \int_0^{\infty} \left(\int_0^{uy} x^{n/2-1} e^{-x/2} y^{m/2-1} e^{-y/2} \mathrm{d}x \right) \mathrm{d}y
\end{aligned} \tag{111}
$$

求导得:

$$
\begin{aligned}
f_U(u) &= c_1 \int_0^{\infty} y (uy)^{n/2-1} e^{-(uy)/2} y^{m/2-1} e^{-y/2} \mathrm{d}y \\
&= c_1 \int_0^{\infty} u^{n/2-1} y^{(n+m)/2-1} \exp\left(-\frac{1}{2}(1+u)y \right) \mathrm{d}y
\end{aligned} \tag{112}
$$

替换 $t = \frac{1}{2}(1+u)y$, 计算得到:

$$\begin{aligned} f_U(u) &= c_2[\int_0^\infty t^{(n+m)/2-1}\exp(-t)\mathrm{d}t][\frac{1}{2}(u+1)]^{-(n+m)/2}u^{n/2-1} \\ &= c_3[(u+1)]^{-(n+m)/2}u^{n/2-1}, u>0 \end{aligned} \tag{113}$$

于是知道 Z 的密度

$$\begin{aligned} f_Z(z) &= \frac{n}{m}f_U(\frac{n}{m}z) \\ &= c_4[(\frac{n}{m}z+1)]^{-(n+m)/2}(\frac{n}{m}z)^{n/2-1}, z>0 \end{aligned} \tag{114}$$

常数 c_4 的计算可以按照前面的计算过程一步一步计算下来, 也可以利用归一化来计算:

$$\begin{aligned}
\frac{1}{c_4} &= \int_0^{\infty} [(\frac{n}{m}z+1)]^{-(n+m)/2}(\frac{n}{m}z)^{n/2-1}\mathrm{d}z \\
&= \int_0^1 (\frac{1}{s}-1)^{n/2-1}s^{(n+m)/2}\frac{m}{ns^2}\mathrm{d}s, [1+nu/m=\frac{1}{s}] \\
&= \frac{m}{n}\int_0^1 (1-s)^{n/2-1}s^{m/2-1}\mathrm{d}s \\
&= \frac{m}{n}\mathrm{B}(n/2,m/2).
\end{aligned} \tag{115}$$

(法二): 增补变量 $W = X + Y$, 来求 (Z, W) 的联合概率密度函数 $g(z, w)$.

事件

$$\begin{aligned}\{Z = z, W = w\} &= \{\frac{m}{n}\frac{X}{Y} = z, X + Y = w\} \\ &= \{X = \frac{\frac{zn}{m}w}{1 + \frac{zn}{m}}, Y = \frac{w}{1 + \frac{zn}{m}}\}\end{aligned} \tag{116}$$

$$\begin{aligned}|\det(J)| &= \left|\det\left(\frac{\partial(z, w)}{\partial(x, y)}\right)\right|^{-1} \\ &= \left|\det\begin{pmatrix}\frac{m}{ny} & -\frac{m}{n}\frac{x}{y^2} \\ 1 & 1\end{pmatrix}\right|^{-1} \\ &= \frac{n}{m}\frac{y^2}{x + y} = \frac{n}{m}\frac{w}{(1 + zn/m)^2}.\end{aligned} \tag{117}$$

因此, 得到 (Z,W) 的联合概率密度函数

$$g(z,w)=f(\frac{\frac{zn}{m}w}{1+\frac{zn}{m}},\frac{w}{1+\frac{zn}{m}})\frac{n}{m}\frac{w}{(1+zn/m)^2}, \tag{118}$$

其中,$f(x,y)$ 是 (X,Y) 的概率密度函数. 将上式经过化简, 得到:

$$g(z,w)=\frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m+n}{2}\right)}w^{\frac{m+n}{2}-1}e^{-\frac{w}{2}}\frac{(n/m)^{n/2}}{\mathrm{B}(m/2,n/2)}z^{n/2-1}(1+\frac{zn}{m})^{-(m+n)/2} \tag{119}$$

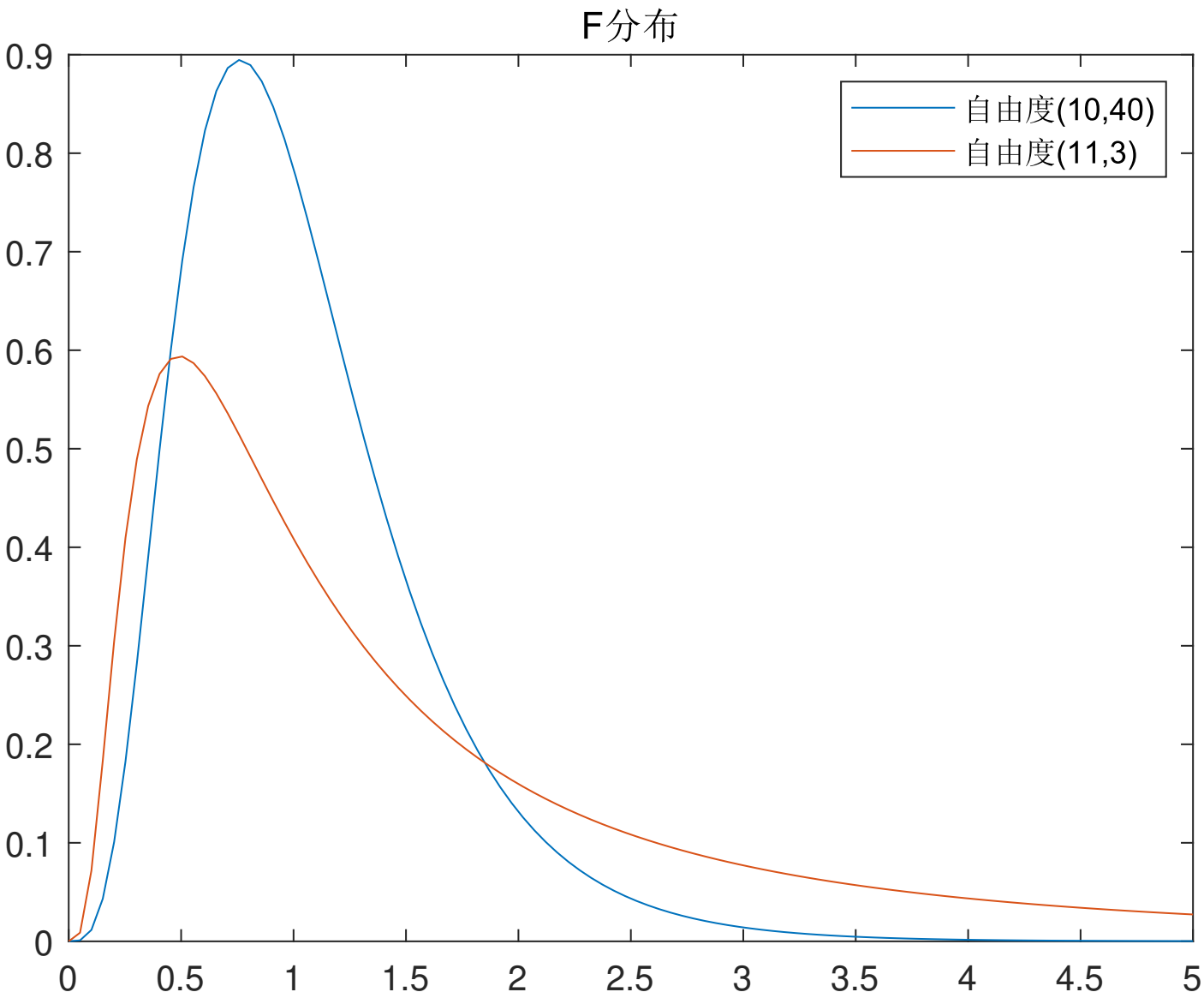
巧了! 观察上式发现 $g(z,w)$ 可分离变量, 且 (Z,W) 的取值区域为矩形区域.

故随机变量 W 和 Z 独立,

W 有密度 $\frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{m+n}{2}\right)}w^{\frac{m+n}{2}-1}e^{-\frac{w}{2}}$.

Z 有密度 $\frac{(n/m)^{n/2}}{\mathrm{B}(m/2,n/2)}z^{n/2-1}(1+\frac{zn}{m})^{-(m+n)/2}$, 证毕.

F 分布, 注意: 分子的自由度在前



抽样分布

例: 如果 $X \sim \chi^2(n), Y \sim \chi^2(m), X, Y$ 独立, 则 $X+Y \sim \chi^2(n+m)$
证明: 取 $X_1, X_2, \ldots, X_{n+m}$ 独立同分布, 都服从标准正态分布. 此时有 (X, Y) 和 $\left(\sum_{i=1}^{n} X_i^2, \sum_{j=n+1}^{n+m} X_i^2\right)$ 同分布, 于是 $X+Y$ 和 $\sum_{i=1}^{n+m} X_i^2$ 同分布.
即有 $X+Y \sim \chi^2(n+m)$. 证毕.

例: 设 $X_1, \ldots, X_n$ 是来自总体 $N(0,1)$ 的样本.

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j, \quad S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2 \tag{120}$$

分别是样本均值和样本方差, 则
(1) $\overline{X}$ 和 S^2 独立
(2) $(n-1)S^2 \sim \chi^2(n-1)$.

抽样分布

证明: 引入正交矩阵

$$\boldsymbol{T}=\frac{1}{\sqrt{n}}\begin{pmatrix}1 & 1 & \dots & 1\\ * & * & \dots & *\\ \vdots & \vdots & & \vdots\\ * & * & \dots & *\end{pmatrix}. \tag{121}$$

由 $\boldsymbol{X}=(X_1,\dots,X_n)^{\mathrm{T}}\sim N(\boldsymbol{0},\boldsymbol{I})$ 知道 $\boldsymbol{Y}=\boldsymbol{T}\boldsymbol{X}\sim N(\boldsymbol{0},\boldsymbol{I})$, 于是 $Y_1,\dots,Y_n$ 独立同分布且都服从于标准正态分布. 利用 $\overline{X}=Y_1/\sqrt{n}$ 和

$$\sum_{j=1}^{n}X_j^2=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}=\sum_{j=1}^{n}Y_j^2 \tag{122}$$

抽样分布

得到:

$$
\begin{aligned}
(n-1)S^2 &= \sum_{j=1}^{n} X_j^2 - n\overline{X}^2 \\
&= \sum_{j=1}^{n} Y_j^2 - Y_1^2 \\
&= \sum_{j=2}^{n} Y_j^2 \sim \chi^2(n-1)
\end{aligned}
\tag{123}
$$

并且和 $\overline{X} = Y_1/\sqrt{n}$ 独立.

抽样分布

练习: 如果 $X_1, X_2, \ldots, X_n$ 是总体 $N(\mu, \sigma^2)$ 的样本, 证明:
(1) $\overline{X}$ 和 S^2 独立
(2) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. 证明:
令

$$Y_j = \frac{X_j - \mu}{\sigma}, \tag{124}$$

则 $Y_1, \ldots, Y_n$ 独立同分布于标准正态分布.

$$\overline{X} = \mu + \sigma\overline{Y}, \frac{(n-1)S^2}{\sigma^2} = \sum_{j=1}^{n}(Y_j - \overline{Y})^2 \tag{125}$$

由上题的结论知道 $\overline{X}$ 和 S^2 独立, 且 $\sum_{j=1}^{n}(Y_j - \overline{Y})^2 \sim \chi^2(n-1)$.

练习 *(**非中心** χ^2 **分布**): **如果** $X_1, X_2, \ldots, X_n$ **相互独立**,X_i **服从** $N(\mu_i, 1)$, **令** $Z = X_1^2 + X_2^2 + \cdots + X_n^2$. **证明**:
(1) Z **的概率密度函数为**:

$$f_Z(z) = e^{-\delta/2} e^{-z/2} \sum_{j=0}^{+\infty} \frac{1}{j!} \frac{(\delta/2)^j}{2^{j+n/2}\Gamma(j+n/2)} z^{j+n/2-1}, z > 0; \qquad (126)$$

(2)Z **的特征函数** $\mathrm{E}e^{itZ} = \frac{1}{(1-2it)^{n/2}} \exp\left(\frac{it\delta}{1-2it}\right)$,
其中,$\delta = \mu_1^2 + \cdots + \mu_n^2 > 0$.

Rmk: **在统计中称** Z **服从自由度为** n、**非中心参数为** δ **的非中心** χ^2 **分布**, **记作** $Z \sim \chi^2(n, \delta)$.

证明:

(1) **作正交矩阵**

$$\boldsymbol{T}=\frac{1}{\sqrt{\delta}}\begin{pmatrix}\mu_1 & \mu_2 & \dots & \mu_n\\ * & * & \dots & *\\ \vdots & \vdots & & \vdots\\ * & * & \dots & *\end{pmatrix}. \tag{127}$$

由于 $\boldsymbol{X}=(X_1,\dots,X_n)^{\mathrm{T}}\sim N(\boldsymbol{\mu},\boldsymbol{I})$, **其中** $\boldsymbol{\mu}=(\mu_1,\dots,\mu_n)^{\mathrm{T}}$.
令 $\boldsymbol{Y}=(Y_1,\dots,Y_n)=\boldsymbol{TX}$, **则** $\boldsymbol{Y}\sim N(\boldsymbol{T\mu},\boldsymbol{I})$, **因此, 有** $Y_1,\dots,Y_n$ **独立,** $Y_i\sim N(b_i,1)$, **按** $\boldsymbol{T}$ **的取法, 有:** $b_1=\sqrt{\delta},b_2=\cdots=b_n=0$.
记 $W=\sum_{j=2}^{n}Y_j^2$, **则** $Z=W+Y_1^2$, **其中,**W **与** Y_1^2 **独立, 而**
$W\sim\chi^2(n-1),Y_1\sim N(\sqrt{\delta},1)$.

容易算出: Y_1^2 **有密度**:

$$
\begin{aligned}
g(x) &= \frac{1}{2\sqrt{2\pi x}}\left[\exp\left(-\frac{(\sqrt{x}-\sqrt{\delta})^2}{2}\right)+\exp\left(-\frac{(-\sqrt{x}-\sqrt{\delta})^2}{2}\right)\right]\mathbf{1}_{(x>0)} \\
&= \frac{1}{2\sqrt{2\pi x}}e^{-\delta/2}e^{-x/2}(e^{\sqrt{\delta x}}+e^{-\sqrt{\delta x}})\mathbf{1}_{(x>0)} \\
&= \frac{1}{\sqrt{2\pi}}e^{-\delta/2}e^{-x/2}\sum_{j=0}^{+\infty}\frac{1}{(2j)!}\delta^j x^{j-1/2}\mathbf{1}_{(x>0)}
\end{aligned}
\tag{128}
$$

回忆 $W\sim\chi^2(n-1)$ **的密度函数**:

$$
k_{n-1}(x)=\frac{1}{2^{(n-1)/2}\Gamma(\frac{n-1}{2})}x^{(n-1)/2-1}e^{-x/2}\mathbf{1}_{(x>0)} \tag{129}
$$

按和的密度公式, 定出 Z 的概率密度函数为

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} g(u) k_{n-1}(z-u)\mathrm{d}u \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\delta/2} e^{-u/2} \sum_{j=0}^{+\infty} \frac{1}{(2j)!} \delta^j u^{j-1/2} \mathbf{1}_{(u>0)} \\
&\times \frac{1}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)} (z-u)^{(n-1)/2-1} e^{-(z-u)/2} \mathbf{1}_{(z-u>0)} \mathrm{d}u \qquad (130) \\
&= \int_{0}^{z} \frac{1}{\sqrt{2\pi}} e^{-\delta/2} e^{-u/2} \sum_{j=0}^{+\infty} \frac{1}{(2j)!} \delta^j u^{j-1/2} \\
&\times \frac{1}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)} (z-u)^{(n-1)/2-1} e^{-(z-u)/2} \mathrm{d}u \mathbf{1}_{(z>0)}
\end{aligned}
$$

将上式稍作整理得到:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\delta}{2}} \frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} e^{-z/2} \sum_{j=0}^{+\infty} \frac{1}{(2j)!}\delta^j \times \int_0^z u^{j-1/2}(z-u)^{(n-1)/2-1}\mathrm{d}u\mathbf{1}_{(z>0)}. \tag{131}$$

利用 Gamma 函数和 Beta 函数的关系:

$$\begin{aligned}\int_0^x y^a(x-y)^b\mathrm{d}y &= x^{a+b+1}\int_0^1 t^a(1-t)^b\mathrm{d}t \\ &= x^{a+b+1}\mathrm{B}(a+1,b+1) \\ &= x^{a+b+1}\frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)}.\end{aligned} \tag{132}$$

化简得到:

$$f_Z(z) = e^{-\delta/2} e^{-z/2} \sum_{j=0}^{+\infty} \frac{1}{j!} \frac{(\delta/2)^j}{2^{j+n/2}\Gamma(j+n/2)} z^{j+n/2-1} \mathbf{1}_{(z>0)} \tag{133}$$

(2) 要计算 Z 的特征函数 $\mathrm{E}e^{itZ}$
若用定义

$$\mathrm{E}e^{itZ} = \int_{-\infty}^{\infty} f_Z(z) e^{itz} dz \tag{134}$$

$\longrightarrow$ 糟糕!

注意到:$Z = X_1^2 + X_2^2 + \cdots + X_n^2$, 并且 $X_1, X_2, \ldots, X_n$ 相互独立, 于是有:

$$\mathrm{E}e^{itZ} = \mathrm{E}e^{itX_1^2} \cdots \mathrm{E}e^{itX_n^2}. \tag{135}$$

注意到：X_k^2 有特征函数

$$\mathrm{E}e^{itX_k^2} = \int_{-\infty}^{+\infty} e^{itx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_k)^2}{2}} \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{M(x)} \mathrm{d}x \tag{136}$$

其中 $M(x) = itx^2 - \frac{(x-\mu_k)^2}{2} = -\frac{(1-2it)}{2}(x - \frac{\mu_k}{1-2it})^2 + \frac{i\mu_k^2 t}{1-2it}$.
因此,

$$\begin{aligned}
\mathrm{E}e^{itX_k^2} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(1-2it)}{2}(x-\frac{\mu_k}{1-2it})^2 + \frac{i\mu_k^2 t}{1-2it}} dx \\
&= \frac{1}{\sqrt{2\pi}} e^{\frac{i\mu_k^2 t}{1-2it}} \int_{-\infty}^{+\infty} e^{-\frac{(1-2it)}{2}y^2} dy \\
&= \frac{1}{(1-2it)^{1/2}} \exp\left(\frac{it\mu_k^2}{1-2it}\right),
\end{aligned} \tag{137}$$

最后,

$$\begin{aligned} \mathrm{E}e^{itZ} &= \mathrm{E}e^{itX_1^2}\cdots\mathrm{E}e^{itX_n^2} \\ &= \frac{1}{(1-2it)^{n/2}}\exp\left(\frac{it\delta}{1-2it}\right). \end{aligned} \tag{138}$$

练习 *(非中心 t 分布): 设 $X\sim N(\delta,1), Y\sim\chi^2(n)$, 求 $T=\frac{X}{\sqrt{Y/n}}$ 的密度函数, 其中 $\delta\in\mathbb{R}$. Ans:

$$\begin{aligned} h_{n,\delta}(x) = &\frac{n^{n/2}}{\sqrt{\pi}\Gamma(\frac{n}{2})}e^{-\delta^2/2}(n+x^2)^{-(n+1)/2} \\ &\times\sum_{j=0}^{\infty}\Gamma(\frac{n+j+1}{2})\frac{(\delta x)^j}{j!}(\frac{2}{n+x^2})^{j/2}. \end{aligned} \tag{139}$$

练习 *(非中心 F 分布): 设 X, Y 独立,$X \sim \chi^2(n), Y \sim \chi^2(m, \delta)$, 求 $Z = \frac{Y/m}{X/n}$ 的密度函数.

Ans:

$$
\begin{aligned}
f_{m,n,\delta}(x) = e^{-\delta^2/2} \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} n^{n/2} m^{m/2+j} \\
\times \frac{\Gamma(\frac{1}{2}(m+n)+j)}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2}+j)} \frac{x^{m/2-1+j}}{(n+mx)^{(m+n)/2+j}} \mathbf{1}_{(x>0)}.
\end{aligned} \tag{140}
$$

练习 *: **设随机变量** U_1, U_2 **独立**,$U_1 \sim \chi^2(n_1, \delta), U_2 \sim \chi^2(n_2)$. **则** $F = (U_1/n_1)/(U_2/n2)$ **是自由度为** (n_1, n_2), **非中心参数为** δ **的** F **分布**.

证明:

(i)$\mathrm{E}F = \frac{n_2(n_1+\delta)}{n_1(n_2-2)}$,(**其中** $n_2 > 2$);

(ii)$\mathrm{var}(F) = \frac{2n_2^2\left[(n_1+\delta)^2+(n_2-2)(n_1+2\delta)\right]}{n_1^2(n_2-2)^2(n_2-4)}$,(**其中** $n_2 > 4$.)

练习: 设 $X \sim \chi^2(2n), n$ 为正整数, 设 $a > 0, Y \sim \mathcal{P}(\frac{a}{2})$(Poisson 分布).
证明:

$$P(X < a) = P(Y \geqslant n). \tag{141}$$

证明:

$$P(X < a) = \frac{1}{2^n \Gamma(n)} \int_0^a x^{n-1} e^{-x/2} \mathrm{d}x \tag{142}$$

反复利用分部积分公式:

$$\begin{aligned} P(X < a) &= \frac{1}{2^n (n-1)!} \int_0^a x^{n-1} e^{-x/2} \mathrm{d}x \\ &= \frac{1}{2^n (n-1)!} \int_{x=0}^a e^{-x/2} \mathrm{d}\left(\frac{x^n}{n}\right) \\ &= \frac{a^n}{2^n n!} e^{-a/2} + \frac{1}{2^{n+1} n!} \int_0^a x^n e^{-x/2} \mathrm{d}x \\ &= \sum_{k=n}^{n+r} \frac{a^k}{2^k k!} e^{-a/2} + \frac{1}{2^{n+r+1}(n+r)!} \int_0^a x^{n+r} e^{-x/2} \mathrm{d}x. \end{aligned} \tag{143}$$

放缩:

$$
\begin{aligned}
0 &< \frac{1}{2^{n+r+1}(n+r)!}\int_0^a x^{n+r}e^{-x/2}\mathrm{d}x \\
&< \frac{1}{2^{n+r+1}(n+r)!}\int_0^a x^{n+r}\mathrm{d}x \\
&= \frac{1}{2^{n+r+1}(n+r)!}\frac{a^{n+r+1}}{n+r+1} \\
&= \frac{(a/2)^{n+r+1}}{(n+r+1)!} \to 0(r\to\infty)
\end{aligned}
\tag{144}
$$

(回忆: 数学分析中的" 阶指幂对" 原理.)
令 $r \to \infty$ 即得结论

练习: 把 χ^2 分布推广到 n 非整数的情况:Z 服从 $\chi^2(a)$ 分布, 则 Z 是有密度

$$f_Z(z) = \frac{1}{2^{a/2}\Gamma(a/2)} z^{a/2-1} e^{-z/2} \mathbf{1}_{(x>0)} \tag{145}$$

的随机变量. 证明:

$$\mathrm{E}Z = a, \mathrm{var}Z = 2a; \tag{146}$$

且当 $a \to \infty$ 时,

$$\frac{Z-a}{\sqrt{2a}} \xrightarrow{d} N(0,1). \tag{147}$$

证明: 由 Z 的概率密度

$$f(z)=\frac{1}{2^{a/2}\Gamma(a/2)}z^{a/2-1}e^{-z/2},z>0 \tag{148}$$

得到:

$$\begin{aligned}\mathrm{E}(Z^k)&=\int_{-\infty}^{\infty}z^k f(z)\mathrm{d}z\\&=\int_0^{\infty}\frac{1}{2^{a/2}\Gamma(a/2)}z^{a/2+k-1}e^{-z/2}\mathrm{d}z\\&=\int_0^{\infty}\frac{2^k}{\Gamma(a/2)}t^{a/2+k-1}e^{-t}\mathrm{d}t\\&=\frac{2^k\Gamma(a/2+k)}{\Gamma(a/2)}\end{aligned} \tag{149}$$

因此,$\mathrm{E}Z=a,\mathrm{E}Z^2=4(\frac{a}{2}+1)(\frac{a}{2})=a(a+2),\mathrm{var}Z=2a.$

还注意到:Z 有特征函数

$$\phi_Z(t) = \mathrm{E}e^{itZ} = \frac{1}{(1-2it)^{a/2}} \tag{150}$$

因此, 随机变量 $\frac{Z-a}{\sqrt{2a}}$ 的特征函数为

$$\begin{aligned}
\mathrm{E}e^{it\frac{Z-a}{\sqrt{2a}}} &= e^{-it\frac{a}{\sqrt{2a}}}\mathrm{E}e^{i\frac{t}{\sqrt{2a}}Z} = e^{-it\frac{a}{\sqrt{2a}}}\frac{1}{\left(1-2i\frac{t}{\sqrt{2a}}\right)^{a/2}} \\
&= \exp\left(-it\frac{a}{\sqrt{2a}} - \frac{a}{2}\ln\left(1-2i\frac{t}{\sqrt{2a}}\right)\right) \\
&= \exp\left(-it\frac{a}{\sqrt{2a}} + it\frac{a}{\sqrt{2a}} - \frac{a}{2}\frac{1}{2}\frac{4t^2}{2a} + o(\frac{1}{\sqrt{a}})\right) \\
&= \exp\left(-\frac{1}{2}t^2 + o(\frac{1}{\sqrt{a}})\right) \to e^{-\frac{t^2}{2}}
\end{aligned} \tag{151}$$

证毕.

例: 如果 $X_1, \ldots, X_n$ 是总体 $N(\mu, \sigma^2)$ 的样本, 则:

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1). \tag{152}$$

证明:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1), \xi = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1). \tag{153}$$

Z 和 ξ 独立, 于是:

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{Z}{\sqrt{\xi/(n-1)}} \sim t(n-1). \tag{154}$$

例: 设 $X_1, \ldots, X_n$ 是来自总体 $N(\mu_1, \sigma_1^2)$ 的样本,$Y_1, \ldots, Y_m$ 是来自总体 $N(\mu_2, \sigma_2^2)$ 的样本, 又设这两个总体是相互独立的, 则当 $m, n \geqslant 2$ 时,

$$\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F(n-1, m-1). \tag{155}$$

其中

$$S_X^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2, S_Y^2 = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \overline{Y})^2 \tag{156}$$

证明: 由于 $(n-1)S_X^2/\sigma_1^2 \sim \chi^2(n-1), (m-1)S_Y^2/\sigma_2^2 \sim \chi^2(m-1)$ 且二者独立,
由 F 分布的定义知结论成立.

例: 设 $X_1, \ldots, X_n$ 是来自总体 $X \sim N(\mu_1, \sigma_1^2)$ 的样本,$Y_1, \ldots, Y_m$ 是来自总体 $Y \sim N(\mu_2, \sigma_2^2)$ 的样本, 又设这两个总体是相互独立的,$\{X_i\}$ 和 $\{Y_j\}$ 的样本均值、样本方差分别为: $\overline{X}, \overline{Y}, S_X^2, S_Y^2$. 定义

$$S_W^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}, \tag{157}$$

则,

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \sim N(0,1). \tag{158}$$

如果还有条件 $\sigma_1^2 = \sigma_2^2 = \sigma^2$, 则

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_W\sqrt{1/n + 1/m}} \sim t(n+m-2). \tag{159}$$

证明：由于总体 X, Y 独立，所以 $X_1, \ldots, X_n$ 与 $Y_1, \ldots, Y_m$ 独立. 根据概率论的知识知道:

$$\overline{X} \sim N(\mu_1, \sigma_1^2/n), \overline{Y} \sim N(\mu_2, \sigma_2^2/m) \tag{160}$$

利用 $\overline{X}, \overline{Y}$ 独立得到:

$$\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \sigma_1^2/n + \sigma_2^2/m) \tag{161}$$

于是得到 $Z \sim N(0,1)$. 再证 $T \sim t(n+m-2)$. 利用

$$\xi_1 = \frac{(n-1)S_X^2}{\sigma^2} \sim \chi^2(n-1), \xi_2 = \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2(m-1), \tag{162}$$

及 ξ_1, ξ_2 独立，得到: $\xi_1 + \xi_2 \sim \chi^2(n+m-2)$.

简单计算, 有:

$$S_W^2 = \frac{(\xi_1 + \xi_2)\sigma^2}{n+m-2} \tag{163}$$

又由于 Z, ξ_1, ξ_2 相互独立, 于是 Z 与 $\xi_1 + \xi_2$ 独立. 因此,

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_W\sqrt{1/n + 1/m}} = \frac{Z}{\sqrt{(\xi_1 + \xi_2)/(n+m-2)}} \sim t(n+m-2). \tag{164}$$

证毕.

例: 设 T 服从自由度是 n 的 t 分布, 证明: 若整数 $k < n$, 则 $\mathrm{E}T^k$ 存在, 若 $k \geqslant n$, 则 $\mathrm{E}T^k$ 不存在.

证明:

研究积分

$$\begin{aligned} I_k &= \int_{-\infty}^{\infty} |t|^k f_T(t) \mathrm{d}t \\ &= \int_{-\infty}^{\infty} a_n |t|^k \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \mathrm{d}t \end{aligned} \tag{165}$$

由于 $|t|^k \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \sim (\frac{1}{n})^{-(n+1)/2} |t|^{-(n+1)+k}, (t \to \infty)$ 根据广义积分的比较审敛法知道: 当且仅当 $-(n+1) + k < -1$ 时, 上式收敛, 由此便得结论.

例: 设 T 服从自由度是 n 的 t 分布, 计算:
(1)$\mathrm{E}T, (n \geqslant 2)$;
(2)$\mathrm{var}T, (n \geqslant 3)$.
解:(1) 由 $f_T(t)$ 关于 $t = 0$ 对称得到$\mathrm{E}T = 0, (n \geqslant 2)$;
(2) 由于 $\mathrm{E}T = 0$, 所以 $\mathrm{var}T = \mathrm{E}T^2$.
将 T 表示成 $T = \frac{X}{\sqrt{Y/n}}$, 其中 $X \sim N(0,1), Y \sim \chi^2(n)$,($X, Y$ 独立) 就有:

$$\mathrm{E}T^2 = n\mathrm{E}X^2\mathrm{E}Y^{-1} = n\mathrm{E}Y^{-1}. \tag{166}$$

前面曾经计算过:

$$\mathrm{E}Y^k = \frac{2^k\Gamma(a/2+k)}{\Gamma(a/2)} \tag{167}$$

取 $k = -1$ 得到:

$$\mathrm{E}Y^{-1} = \frac{2^{-1}\Gamma(n/2-1)}{\Gamma(n/2)} = (n-2)^{-1}. \tag{168}$$

因此,$\mathrm{var}T = \frac{n}{n-2}, (n \geqslant 3)$.

例: (1) 如果 $X \sim F(n,m)$, 则 $X^{-1} \sim F(m,n)$
(2) 对 $F(n,m)$ 分布的上 α 分位数 $F_\alpha(n,m)$, 有:

$$F_\alpha(n,m) = \frac{1}{F_{1-\alpha}(m,n)} \tag{169}$$

证明留作习题.
(3) 如果 $X \sim F_{n,m}$, 当 $m > 2$ 时,$\mathrm{E}X = \frac{m}{m-2}$.(仅与第二自由度有关)
证明: 表示 $X = \frac{\xi_1/n}{\xi_2/m}$, 其中 $\xi_1 \sim \chi^2(n), \xi_2 \sim \chi^2(m)$ 相互独立,

$$\mathrm{E}X = \frac{m}{n}\mathrm{E}\xi_1 \mathrm{E}\frac{1}{\xi_2} = \frac{m}{n} n \frac{1}{m-2} = \frac{m}{m-2} \tag{170}$$

The End