

《概率论与数理统计》复习笔记

由HDU-STEASyan2018整理于2020年1月3日

由HDU-STEAParticle_G校对

基本概念

随机测试

我们将能在相同条件下重复进行的、每次测试结果不止一个，并事先明确所有可能结果的、在测试之前不能预测出现结果的试验称作随机测试

样本空间

随机测试 E 所有可能结果的集合称作 E 的样本空间，每种可能结果称作样本点

随机事件

E 样本空间一部分样本点的集合称作 E 的随机事件

其中由单个样本点组成的事件称作基本事件，由全集组成的事件称作必然事件，空集组成的事件称作不可能事件

关系运算

包含、等价

$A \subset B$ ，称作 B 事件包含 A 事件，指事件 A 的发生必然导致事件 B 的发生

在选择题时，可以通过作图画圆来辅助求解

如果同时有 $B \subset A$ 则称事件 A 和事件 B 相等，即 $A=B$

和关系

事件 $A \cup B = \{x | x \in A \vee x \in B\}$ 称为事件 A 和事件 B 的和事件，当且仅当 AB 之一发生时，有事件 $A \cup B$ 发生。

额外的，称事件 $\bigcup_{k=1}^n A_k$ 为 n 个事件 A_i 的和事件，其中要求事件可列

积关系

事件 $A \cap B = \{x | x \in A \wedge x \in B\}$ 称为事件 A 和事件 B 的积事件，当且仅当 AB 同时发生时，有事件 $A \cap B$ 发生。

额外的，称事件 $\bigcap_{k=1}^n A_k$ 为 n 个事件 A_i 的积事件，其中要求事件可列

差关系

事件 $A - B = \{x | x \in A \wedge x \notin B\}$ 称为事件 A 和事件 B 的差事件，当且仅当 A 发生， B 不发生，事件 $A - B$ 发生。

互斥关系

若 $A \cap B = \varnothing$ 则称事件 A 与 B 互斥，即两事件不可能同时发生

逆关系

若 $A \cap B = \varnothing \wedge A \cup B = S$ ，称事件 A 与 B 为逆事件，即对于每次测试，必有 AB 之一发生

有 $\bar{A} = S - A$

关系运算律

交换律、结合律、分配律

德摩根律： $\overline{A \cup B} = \bar{A} \cap \bar{B}$ ， $\overline{A \cap B} = \bar{A} \cup \bar{B}$

频率（后验）

在相同条件下进行 n 次测试，在这 n 次测试中事件发生的次数称为事件发生的频数，频数和次数的比值被称作频率，记作 $f_n(A)$ 。频率表示一个事件在测试中发生的可能性

概率（先验）

设 E 是随机测试， S 为其对应样本空间，对 E 中的每一事件 A 赋予特定的实数 $P(A)$ ，将之称作事件 A 的概率。如果集合函数 $P(\cdot)$ 满足非负性($P \geq 0$)、规范性（对必然事件有 $P = 1$ ）、可列可加性（ $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$ ），则可以用 P 表示事件在一次测试中发生可能性的大小

$$P(A) = \lim_{n \rightarrow \infty} f_n(A)$$

性质

- （有限可加性）对于互不相容事件 A_i ，有 $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$
- $P(\varnothing) = 0$
- 设事件 A 、 B ，若有 $A \subset B$ ，则有 $P(B-A) = P(B) - P(A)$, $P(B) > P(A)$
- $P(A) \leq P(S) = 1$
- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B - AB) = P(A) + P(B) - P(AB)$

等可能概型

超几何分布模型：

$$S = \binom{N}{n}, A_k = \binom{M}{k} \cdot \binom{N-M}{n-k}$$
$$P(A_k) = \frac{A_k}{S}$$

其中组合数 $C(r, a)$ 的定义为

$$C_a^r = \binom{a}{r} = \frac{a(a-1) \cdots (a-r+1)}{r!}$$

（要求 a 、 r 为非负整数，且 $r \leq a$ ）

条件概率

$P(B|A) = \frac{P(AB)}{P(A)}$ 称作在 A 事件发生的情况下事件 B 发生的条件概率。条件概率大于零，且对于必然事件 B ，有条件概率等于一。

条件概率同样有可列可加性： $P(\bigcup_{i=1}^{\infty} B_i | A) = \sum_{i=1}^{\infty} P(B_i | A)$

推广上式可得相应的乘法公式 $P(AB) = P(B|A)P(A)$

划分

设 S 为试验 E 的样本空间， B_i 为 E 的一组事件，若

- $B_i B_j = \varnothing, i \neq j, i, j = 1, 2, \dots, n$
- $B_1 \cup B_2 \cup \dots \cup B_n = S$

则称 B_i 为样本空间 S 的一个划分，对于划分有，单次测试有且仅有一个事件发生

全概率公式

$$\text{若有 } \bigcup_{k=1}^n B_k = S$$
$$\text{且有 } B_i B_j = \varnothing (i \neq j, i, j = 1, 2, 3, \dots, n)$$
$$\text{且有 } P(B_k) > 0 (k = 1, 2, 3, \dots, n)$$
$$\text{则有 } P(A) = \sum_{k=1}^n P(B_k) \cdot P(A|B_k)$$

全概率公式通过积累条件概率来近似推算\$P(A)\$

贝叶斯公式

若有
$$\bigcup_{k=1}^n B_k = S$$

且有
$$B_i B_j = \varnothing (i \neq j, i, j = 1, 2, 3, \cdots, n)$$

且有
$$P(B_k) > 0 (k = 1, 2, 3, \cdots, n)$$

且有
$$P(A) > 0$$

则有
$$P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{i=1}^n P(B_i) \cdot P(A|B_i)}$$

可以简化作 $P(B_k|A)=\frac{P(B_k)\cdot P(A|B_k)}{P(A)}$

证明： $P(B_k|A)=\frac{P(B_kA)}{P(A)}=\frac{P(A|B_k)\cdot P(B_k)}{P(A)}$

独立性

满足\$P(AB)=P(A)P(B)\$的事件被称作独立事件

随机变量及其分布

随机变量

设定随机样本空间为\$S=\{e\}\$，\$X=X(e)\$是定义在样本空间上的实值单值函数（即仅与具体的元事件相关联），称\$X=X(e)\$为随机变量。随机变量是样本点事件到实数轴的一一映射。

离散型随机变量

离散型随机变量取值可列，且每个当值对应一个概率。描述离散型随机变量\$X\$的分布有

$$P\{X = x_k\} = p_k, k = 1, 2, \cdots$$

称之为离散型随机变量的分布率，对应的可以使用表格表示

X	x_1	x_2	\cdots	x_n	\cdots
p_k	p_1	p_2	\cdots	p_n	\cdots

要求 p_i 的累加和等于\$1\$或趋向于\$1\$（对于无穷可列离散型随机变量）

典型离散型随机变量

（0-1）分布

X	0	1
p_k	$1-p$	p

样本空间只有两种元素，仅进行一次测试的特殊分布

伯努利测试/二项分布

\$(0-1)\$分布进行多次重复试验的分布

进行\$n\$次测试，其中\$k\$次发生随机变量\$X\$对应事件的概率为：

$$P\{X = k\} = \binom{n}{k} p^k (1 - p)^{n-k}$$

二项分布记作\$X\sim b(n,p)\$

泊松分布

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \cdots$$

记作\$X \sim \pi(\lambda)\$

对于二项分布，存在特性 $\lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$ ，其中 $\lambda = np$

当 n 较大时可以通过二项分布近似泊松分布，以简化计算

分布函数

对于随机变量 X ，和任意实数 x ，有 X 的分布函数

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

对于任意 $x_1 \leq x_2$ ，有

$$\begin{aligned} P\{x_1 < X \leq x_2\} &= P\{X \leq x_2\} - P\{X \leq x_1\} \\ &= F(x_2) - F(x_1) \end{aligned}$$

其中对离散型随机变量 X 有分布函数 $F(x) = \sum_{x_k \leq x} p_k$

另外的，对连续的随机变量，累加可以表示为积分

连续型随机变量

对于随机变量 X 的分布函数 $F(x)$ ，如存在非负可积函数 $f(x)$ ，使得对任意实数 x 有

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

则称 X 为连续型随机变量， $f(x)$ 为 X 的概率密度函数，与离散型随机变量的分布率对应

当 x 趋向于无穷时， $F(x)$ 取值为1

若 $f(x)$ 在点 x 处连续，则有 $F'(x) = f(x)$ ，反之亦成立

特别的，有 $P\{X=a\} = 0$ ，即连续型随机变量在单值上概率密度为0

典型连续型随机变量

均匀分布

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & else \end{cases}$$

称 X 在区间 (a,b) 上均匀分布，记作 $X \sim U(a,b)$

显然，其对应的分布函数从 a 到 b 由0均匀增加至1

指数分布

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0 \\ 0 & else \end{cases}$$

易得概率密度函数有对应的分布函数

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{\theta}} & x > 0 \\ 0 & else \end{cases}$$

指数分布存在一个特殊的性质 $P\{X>s+t | X>s\} = P\{X>t\}$ ，被称作无记忆性

正态分布

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

正态分布也被称为高斯分布，记作 $X \sim N(\mu, \sigma^2)$

关于正态分布的处理，通常会通过换元法 $Z = \frac{X-\mu}{\sigma}$ 统一转化为 $\mu = 0, \sigma = 1$ 的标准正态分布，其中概率密度和分布函数特别的表示为 $\varphi(x), \Phi(x)$ ，有：

$$\begin{aligned} \varphi(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2}} \\ \Phi(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \end{aligned}$$

额外的，有 $\Phi(-x)=1-\Phi(x)$

随机变量的函数分布

已知当前随机变量概率分布时，可以求得变量的函数的概率分布，如果函数 g 处处可导，且 g' 大于 0 ，即函数严格单调时，我们可以求得 $Y=g(X)$ 的分布

$$f_Y(y)=\begin{cases} f_X[h(y)]|h'(y)| & y\in(\alpha,\beta) \\ 0 & else \end{cases}$$

其中 $\alpha=\min\{g(a),g(b)\},\beta=\max\{g(a),g(b)\}$ ， $h(y)$ 为 $g(x)$ 的反函数

证明：

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} = P\{g(X) \leq y\} \\ &= P\{X \leq h(y)\} = F_X[h(y)] \\ f_Y(y) &= \begin{cases} f_X[h(y)]|h'(y)| & y\in(\alpha,\beta) \\ 0 & else \end{cases} \end{aligned}$$

多维随机变量及其分布

二维随机变量

为了避免信息的过分丢失，我们将样本空间扩大到更高维的层面进行精确的描述。其中，将样本空间映射到一个二维平面上，即可获得二维的随机变量，即映射由两个坐标轴进行表示，写作 (X, Y) 。

我们同样用分布函数来对多维随机变量进行描述：

$$F(x,y)=P\{(X\leq x)\cap(Y\leq y)\} \overset{\text{记做}}{=} P\{X\leq x,Y\leq y\}$$

易知分布函数对于变量 x 和 y 都是不减函数，且当任一变量小于定义域时有 $F=0$

对于离散型二维随机变量，我们能获得变量的联合分布率 $P(X=x_i,Y=y_i)=p_{ij}$

这个联合分布率同样可以用表格表示，另外有联合分布函数 $F(x,y)=\sum\limits_{x_i\leq x}\sum\limits_{y_i\leq y}p_{ij}$

对于连续型随机变量，有 $F(x,y)=\int_{-\infty}^x\int_{-\infty}^y f(u,v)\mathrm{d}u\mathrm{d}v$

其中 $f(x,y)$ 被称为随机变量的联合概率密度

随机变量落在映射平面上固定范围的概率为概率密度函数的积分

特别的，若 f 在点 x,y 处连续，有 $\frac{\partial^2 F(x,y)}{\partial x\partial y}=f(x,y)$

边缘分布

虽然我们主要考虑的是二维随机变量 (X,Y) ，但随机变量 X 和 Y 固然存在其自身的概率分布，记作 $F_X(x)$ 和 $F_Y(y)$ ，他们被称为二维随机变量的边缘分布函数，可以由 $F(x,y)$ 确定，有：

离散型随机变量：

$$F_X(x)=F(x,\infty)=\sum_{x_i\leq x}\sum_{j=1}^{\infty}p_{ij}$$

$$P\{X=x_i\}=\sum_{j=1}^{\infty}p_{ij}$$

记 $p_{i\cdot}=\sum_{j=1}^{\infty}p_{ij}$ 为 (X,Y) 关于 X 的边缘分布率

连续型随机变量：

$$F_X(x)=F(x,\infty)=\int_{-\infty}^x\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y\mathrm{d}x$$

其中 (X,Y) 关于 X 的边缘概率密度为 $f_X(x)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y$

类似的，我们可以得到Y的分布

条件分布

条件分布基于条件概率定义，讨论在随机事件发生的前提下另一随机变量的分布，易知，有：（为了节省打字量采用了极不严谨的写法）

$$P\{X\leq x|Y=y\}=\frac{P\{X\leq x,Y=y\}}{P\{Y=y\}}$$

对于指定的\$y\$，我们称该式为\$Y=y\$下\$X\$的条件概率分布

对于离散型随机变量可得\$X\$的分布率\$\frac{p_{ij}}{p_{\cdot j}}\$,有对应分布函数：

$$\sum_{x_i\leq x}\frac{p_{ij}}{p_{\cdot j}}$$

对于连续型，有\$X\$的密度函数\$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}\$，对应有条件分布函数：

$$F_{X|Y}(x|y)=P\{X\leq x|Y=y\}=\int_{-\infty}^x\frac{f(x,y)}{f_Y(y)}\mathrm{d}x$$

相互独立

若两个事件相互独立，可以得出其对应的随机变量相互独立，即写作

$$\begin{aligned}P\{X\leq x,Y\leq y\}&=P\{X\leq x\}P\{Y\leq y\}\\F(x,y)&=F_X(x)F_Y(y)\\f(x,y)&=f_X(x)f_Y(y)\end{aligned}$$

对于\$n\$维随机变量，依旧如此推广

另外的，可以引申出一个结论，两个相互独立随机变量的函数也相互独立

函数分布

\$Z=X+Y\$

$$f_{X+Y}(z)=f_X*f_Y=\int_{-\infty}^{\infty}f(x,z-x)\mathrm{d}x$$

\$Z=XY\$

$$f_{XY}(z)=\int_{-\infty}^{\infty}\frac{1}{|x|}f(x,\frac{z}{x})\mathrm{d}x$$

\$Z=\max/\min\{X,Y\}\$

$$\begin{aligned}F_{max}(z)&=F_X(z)F_Y(z)\\F_{min}(z)&=1-[1-F_X(z)][1-F_Y(z)]\end{aligned}$$

随机变量的数字特征

数学期望

数学期望简称期望，又称作均值，记作\$E(x)\$，是一个用以表征分布重心的实数

数学表示为单值的获取概率与其值的积在整个分布上的和

对于离散型随机变量，有

$$E(x)=\sum_{k=1}^{\infty}x_kp_k$$

对于连续型随机变量，有

$$E(x)=\int_{-\infty}^{\infty}xf(x)\mathrm{d}x$$

数学期望\$E(x)\$完全由随机变量\$X\$的概率分布决定，若\$X\$服从特定分布，则称\$E(x)\$为该分布的数学期望

特殊分布的数学期望

指数分布：	$E(X) = \theta$
泊松分布：	$E(X) = \lambda$
平均分布：	$E(X) = \frac{a+b}{2}$
二项分布：	$E(X) = np$
高斯分布：	$E(X) = \mu$

函数的数学期望

设 Y 是随机变量 X 的函数： $Y=g(X)$

对于离散型随机变量

$$E(Y) = E[g(x)] = \sum_{k=1}^{\infty} g(x_k) p_k$$

对于连续型随机变量

$$E(Y) = E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

对于两者，均要求有期望值绝对收敛

证明：

前证：

$$\begin{aligned} f_Y(y) &= \begin{cases} f_X[h(y)]|h'(y)| & y \in (\alpha, \beta) \\ 0 & else \end{cases} \\ E(Y) &= \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{\alpha}^{\beta} y f_X[h(y)]|h'(y)| dy \\ &\because x = h(y) \\ E(Y) &= \int_{\alpha}^{\beta} y f_X[h(y)]|h'(y)| dy \\ &= \int_{\alpha}^{\beta} y f_X[h(y)] dh(y) \\ &= \int_{-\infty}^{\infty} g(x) f(x) dx \end{aligned}$$

该定理同样可以推广到多个随机变量函数的情况，有

$$E(Z) = E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy$$

该公式同样适用于二维随机变量仅考虑单变量期望的情况

数学期望的性质

1. 设 C 是常数，有 $E(C)=C$
2. 设 X 为随机变量， C 是常数，有 $E(CX)=CE(X)$
3. 设 X, Y 是两个随机变量，则有 $E(X+Y)=E(X)+E(Y)$
4. 设 X, Y 是相互独立随机变量，有 $E(XY)=E(X)E(Y)$

方差

方差是用来表征随机变量与其均值偏离程度的数字特征

设 X 为一个随机变量，有

$$D(X) = Var(X) = E\{[X - E(X)]^2\}$$

记 D/Var 为随机变量 X 的方差

另引入量 $\sqrt{D(X)}$ 记作 $\sigma(X)$ ，称为标准差/均方差

对于离散型随机变量，有

$$D(X) = \sum_{k=1}^{\infty} [x_k - E(X)]^2 p_k$$

对于连续型随机变量，有

$$D(X) = \int [x - E(X)]^2 f(x) dx$$

通常的，给出计算公式

$$\begin{aligned} D(X) &= E\{[X - E(X)]^2\} = E\{X^2 - 2XE(X) + [E(X)]^2\} \\ &= E(X^2) - 2E(X)E(X) + [E(X)]^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

特殊分布的方差

$$\begin{aligned} 0-1 \text{分布: } & D(X) = p(1-p) \\ \text{泊松分布: } & D(X) = \lambda \\ \text{均匀分布: } & D(X) = \frac{(b-a)^2}{12} \\ \text{指数分布: } & D(X) = \theta^2 \\ \text{二项分布: } & D(X) = np(1-p) \\ \text{高斯分布: } & D(X) = \sigma^2 \end{aligned}$$

方差的性质

1. 设\$C\$是常数，\$D(C)=0\$
2. 设\$X\$是随机变量，\$C\$是常数，\$D(CX)=C^2D(X)\$，\$D(X+C)=D(X)\$
3. 设\$X\$，\$Y\$是两个随机变量，则有

$$D(X+Y) = D(X) + D(Y) + 2E(X - E(X))(Y - E(Y))$$

当\$XY\$相互独立时，有去除余项

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

$$\begin{aligned} 2E\{[X - E(X)][Y - E(Y)]\} &= 2E\{XY - XE(Y) - YE(X) + E(X)E(Y)\} \\ &= 2\{E(XY) - E(X)E(Y)\} \end{aligned}$$

1. \$D(X)=0\$的充要条件是\$P\{X=E(X)\}=1\$

切比雪夫不等式

设随机变量\$X\$具有数学期望\$E(X)=\mu\$，方差\$D(X)=\sigma^2\$，则对任意正数\$\epsilon\$，有

$$P\{|X - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}$$

证明：

$$P\{|X - \mu| \geq \epsilon\} = \int_{|x-\mu| \geq \epsilon} f(x) dx \leq \int_{|x-\mu| \geq \epsilon} \frac{|x - \mu|^2}{\epsilon^2} f(x) dx \leq \frac{1}{\epsilon^2} \int (x - \mu)^2 f(x) dx = \frac{\sigma^2}{\epsilon^2}$$

通常的，切比雪夫不等式被用作概率估算的方法

协方差与相关系数

对于多维随机变量，除了要讨论期望和方差，有时还需要讨论随机变量间的关系特征

仅对二维随机变量，经前面的探讨，有如果随机变量相互独立\$E\{[X-E(X)][Y-E(Y)]\}=0\$

因此的，我们可以发现该式具有表现两随机变量关联性的特征，于是定义**协方差**

$$\text{Cov}(X, Y) = E\{[X - E(X)][Y - E(Y)]\}$$

由定义可知：

$$\begin{aligned} \text{Cov}(X, Y) &= \text{Cov}(Y, X) \\ \text{Cov}(X, X) &= D(X) \\ D(X+Y) &= D(X) + D(Y) + 2\text{Cov}(X, Y) \\ \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \end{aligned}$$

给出协方差的部分性质：

$$\begin{aligned}\text{Cov}(aX, bY) &= ab\text{Cov}(X, Y) \\ \text{Cov}(X_1 + X_2, Y) &= \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)\end{aligned}$$

相关系数

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

给定定理

1. $|\rho_{XY}| \leq 1$
2. $|\rho_{XY}| = 1$ 的充要条件是，存在常数a,b使

$$P\{Y = a + bX\} = 1$$

有相干系数表示两个随机变量间的联系，对于相干系数趋向于1时，有随机变量间线性关系紧密，而相关系数较小时，有关联程度弱。

其中特例为相关系数=0时，称随机变量不相关。

矩与协方差矩阵

矩

原点矩

$$E(X^k)$$

称为X的k阶原点矩

原点矩表示元素成分距离原点的分布情况，其中k阶为维度约束，表示在不同维度（层次）上变量距离中心的距离，可以结合范数进行理解。给出图片，这里的表示的是三变量混合分布下的不同范数约束

中心矩

$$E\{[X - E(X)]^k\}$$

称为X的k阶中心矩

中心矩和原点矩的概念其实类似，只是把原点位移到了变量期望的位置

混合矩

$$E(X^k Y^l)$$

称为X和Y的k+l阶混合矩

混合中心矩

$$E\{[X - E(X)]^k [Y - E(Y)]^l\}$$

称为X和Y的k+l阶混合中心矩

拾遗

显然的，我们可以看出数学期望即为一阶原点矩，而方差即为二阶中心矩，协方差则为二阶混合中心矩

协方差矩阵

协方差矩阵是为了讨论随机变量间的关联性及其方差引入的分布矩阵特性矩阵

假设有n维随机变量 X_i 的二阶混合中心矩均存在，即

$$c_{ij} = \text{Cov}(X_i, X_j) = E\{[X_i - E(X_i)][X_j - E(X_j)]\}$$

均存在，我们将其排列为矩阵形式，称矩阵

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

为协方差矩阵，有协方差矩阵为对称矩阵（储存结构了解一下（唐突数据结构草

统计基本定理

大数定律

辛钦大数定理

设随机变量序列 X_k 相互独立，服从同分布，且具有数学期望 $E(X_k)=\mu$ 。则作前 n 个变量的算术平均值 $\frac{1}{n} \sum_{k=1}^n X_k$ ，对于任意 $\epsilon > 0$ ，有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| < \epsilon \right\} = 1$$

证明：

$$\begin{aligned} E \left(\frac{1}{n} \sum_{k=1}^n X_k \right) &= \frac{1}{n} \sum_{k=1}^n E(X_k) = \frac{1}{n} (n\mu) = \mu \\ D \left(\frac{1}{n} \sum_{k=1}^n X_k \right) &= \frac{1}{n^2} \sum_{k=1}^n D(X_k) = \frac{1}{n^2} (n\sigma^2) = \frac{\sigma^2}{n} \\ 1 \geq P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| < \epsilon \right\} &\geq 1 - \frac{\sigma^2/n}{\epsilon^2} \end{aligned}$$

由切比雪夫不等式推导，令 n 趋向于 ∞ ，由夹逼定理得 $P=1$

同时称 X_n 依概率收敛于 μ ，现给出弱大数定理的另一种表示：

设随机变量 X_k 相互独立，服从同一分布且具有数学期望 μ ，则序列 $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ 依概率收敛于 μ ，即

$$\bar{X} \xrightarrow{P} \mu$$

伯努利大数定理

设 f_A 为 n 次独立重复试验中事件 A 发生的次数， p 是事件 A 在每次试验中发生的概率，则有：

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{f_A}{n} - p \right| < \epsilon \right\} = 1$$

中心极限定理

[关于这一块证明可以查看知乎的这篇文章:辉煌的中心极限定义](#)

感谢大崎佬同学的安利√

定理1：独立同分布的中心极限定理

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 独立同分布，设其具有数学期望 $E(X_k)=\mu$ ，方差 $D(X_k)=\sigma^2>0$

设随机变量之和 $\sum_{k=1}^n X_k$ 的标准化变量

$$Y_n = \frac{\sum_{k=1}^n X_k - E(\sum_{k=1}^n X_k)}{\sqrt{D(\sum_{k=1}^n X_k)}} = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma}$$

其分布函数 $F_n(x)$ 对于任意 x 满足

$$\begin{aligned}\lim_{n \rightarrow \infty} F_n(x) &= \lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \leq x \right\} \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ &= \Phi(x)\end{aligned}$$

定理2：李雅普诺夫定理（不考，暂略，待补充）

定理3：蒂莫夫-拉普拉斯定理

设随机变量 η_n 服从参数为 n ， p 的二项分布，则对任意 x ，有

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\eta_n - np}{\sqrt{np(1-p)}} \leq x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x)$$

数理统计

通过试验和观察获得的数据进行抽样，对概率分布进行分析

样本与抽样分布

随机样本

研究有关对象的某一数量指标视为统计对象，为考虑与该指标分布相关的随机变量，需要对该指标进行定量的观察，将试验的全体可观察值称作总体，将单个特定观察值视为个体，总体包含个体个数称作样本容量，通过个体数量划分有限总体和无限总体

在实际中，总体的分布通常是未知的，为此，我们需要从总体中进行抽样，取一系列独立同分布的个体作为样本，而对于其中抽取的样本数，被称作样本容量。其中每次的个体观察值被称作样本值

对于样本，有分布函数和概率密度

$$\begin{aligned}F^*(x_1, x_2, \dots, x_n) &= \prod_{i=1}^n F(x_i) \\ f^*(x_1, x_2, \dots, x_n) &= \prod_{i=1}^n f(x_i)\end{aligned}$$

抽样分布

样本同样存在均值、方差、均方差、原点矩、中心矩

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i \\ S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} (\sum_{i=1}^n X_i^2 - n\bar{X}^2) \\ S &= \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \\ A_k &= \frac{1}{n} \sum_{i=1}^n X_i^k \\ B_k &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k\end{aligned}$$

***关于样本方差中的n-1**

$$\begin{aligned}
E[s_{biased}^2] &= E\left[\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2\right] \\
&= E\left[\frac{1}{n} \sum_{i=1}^n [(x_i - \mu) - (\bar{x} - \mu)]^2\right] \\
&= E\left[\frac{1}{n} \sum_{i=1}^n [(x_i - \mu)^2 - 2(\bar{x} - \mu)(x_i - \mu) + (\bar{x} - \mu)^2]\right] \\
&= E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - \frac{2(\bar{x} - \mu)}{n} \sum_{i=1}^n (x_i - \mu) + \frac{1}{n} \sum_{i=1}^n (\bar{x} - \mu)^2\right] \\
&= E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - 2(\bar{x} - \mu)^2 + \frac{1}{n} \sum_{i=1}^n (\bar{x} - \mu)^2\right] \\
&= E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - 2(\bar{x} - \mu)^2 + \frac{1}{n} \cdot n \cdot (\bar{x} - \mu)^2\right] \\
&= E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - 2(\bar{x} - \mu)^2 + (\bar{x} - \mu)^2\right] \\
&= E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - (\bar{x} - \mu)^2\right] \\
&= E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2\right] - E[(\bar{x} - \mu)^2] \\
&= \frac{1}{n} \sum_{i=1}^n E[(x_i - \mu)^2] - E[(\bar{x} - \mu)^2] \\
&= \frac{1}{n} \sum_{i=1}^n \sigma^2 - E[(\bar{x} - \mu)^2] \\
&= \frac{1}{n} \cdot n \cdot \sigma^2 - E[(\bar{x} - \mu)^2] \\
&= \sigma^2 - E[(\bar{x} - \mu)^2] \\
&= \sigma^2 - Var[\bar{x}] \\
&= \sigma^2 - \frac{\sigma^2}{n} \\
&= \left(\frac{n-1}{n}\right) \sigma^2
\end{aligned}$$

这里给出的是推论公式，不管怎么说看到这里应该也不至于对这个结论过度纠结了

不过为了增进理解，这里结合知乎选手张英峰给出的理解作两句总结，有兴趣的同学可以去看看原链接

[为什么样本方差（sample variance）的分母是 n-1？](#)

即我们在样本方差的计算中引入了一个 \bar{X} 作为限制因素，缩小了分布的自由度 $n \rightarrow n-1$ ，因此估计量 $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 无法继续表示 σ^2 的无偏估计，遂经过严密的证明对样本方差进行了修正。

χ^2 分布

设 X_k 是来自总体 $N(0,1)$ 的样本，称统计量

$$\chi^2 = \sum_{i=1}^n X_k^2$$

服从自由度为 n 的 χ^2 分布，记作 $\chi^2 \sim \chi^2(n)$

χ^2 分布的性质

χ^2 分布有概率密度函数

$$f(y) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(n/2)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} & y > 0 \\ 0 & else \end{cases}$$

首先给出一下 Γ 函数的介绍吧，这里附上参考链接

[Gamma/伽马函数，伽马分布](#)

Γ 函数实际上是泊松分布的一种变体，一种较高纬度的连续化表示，有

$$\Gamma(x, k) = \frac{x^k e^{-x}}{k!}$$

$$X_i \sim N(0, 1) \rightarrow X_i^2 \sim \chi^2(1) \Leftrightarrow X_i^2 \sim \Gamma(\frac{1}{2}, 2) \rightarrow \chi^2 \sim \Gamma(\frac{n}{2}, 2)$$

可加性

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$$

数学期望\方差

$$E(X_i^2) = D(X_i) = 1$$

$$D(X_i^2) = E(X_i^4) - [E(X_i^2)]^2 = 3 - 1 = 2$$

$$E(\chi^2) = n$$

$$D(\chi^2) = 2n$$

分位点

定义分位点

对于给定整数 $\alpha \in (0,1)$ ，满足

$$P\{\chi^2 > \chi_\alpha^2(n)\} = \int_{\chi_\alpha^2(n)}^{\infty} f(y) dy = \alpha$$

则称点 $\chi_\alpha^2(n)$ 称为分布上的 α 分位点，对于充分大的 n ，卡方分布的 α 分位点存在

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

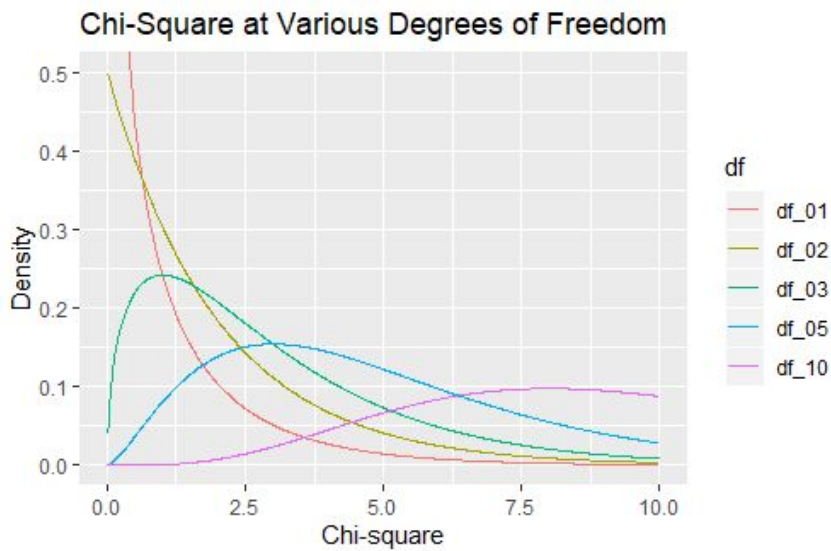
其中 z_α 为标准正态分布上的 α 分位点，有 $\Phi(z_\alpha) = 1 - \alpha$

通过关系可推， $-z_\alpha = z_{1-\alpha}$

χ^2 分布的现实意义

首先给出我认为是卡方分布的基本意义，即检验样本相较于期望的离散化程度

卡方分布的表达式是一系列服从标准正态分布随机变量的平方和，其中随机变量的个数被称作卡方分布的**自由度**，这里给出不同自由度下卡方分布的概率密度图像



自由度 n 表示存在独立变量的个数，亦表示单个变量对整体分布影响的大小，自由度越大，有分布图象越趋向于对称。

卡方分布的特性基于 X 的分布，自由度 n 的卡方分布表示的是累计抽样带来的， n 次试验结果与期望的累计偏差。

在图像上来看，卡方分布与泊松分布存在一定的相似性（由于 Γ 函数的性质）

t 分布

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

$$t = \frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t 分布，记作 $t \sim t(n)$

t 分布的性质

t 分布存在概率密度函数

$$h(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n} \Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

当 t 分布中的自由度 n 充分大时，有概率密度函数趋近于标准正态分布

$$\lim_{n \rightarrow \infty} h(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

有 t 分布的上分位点

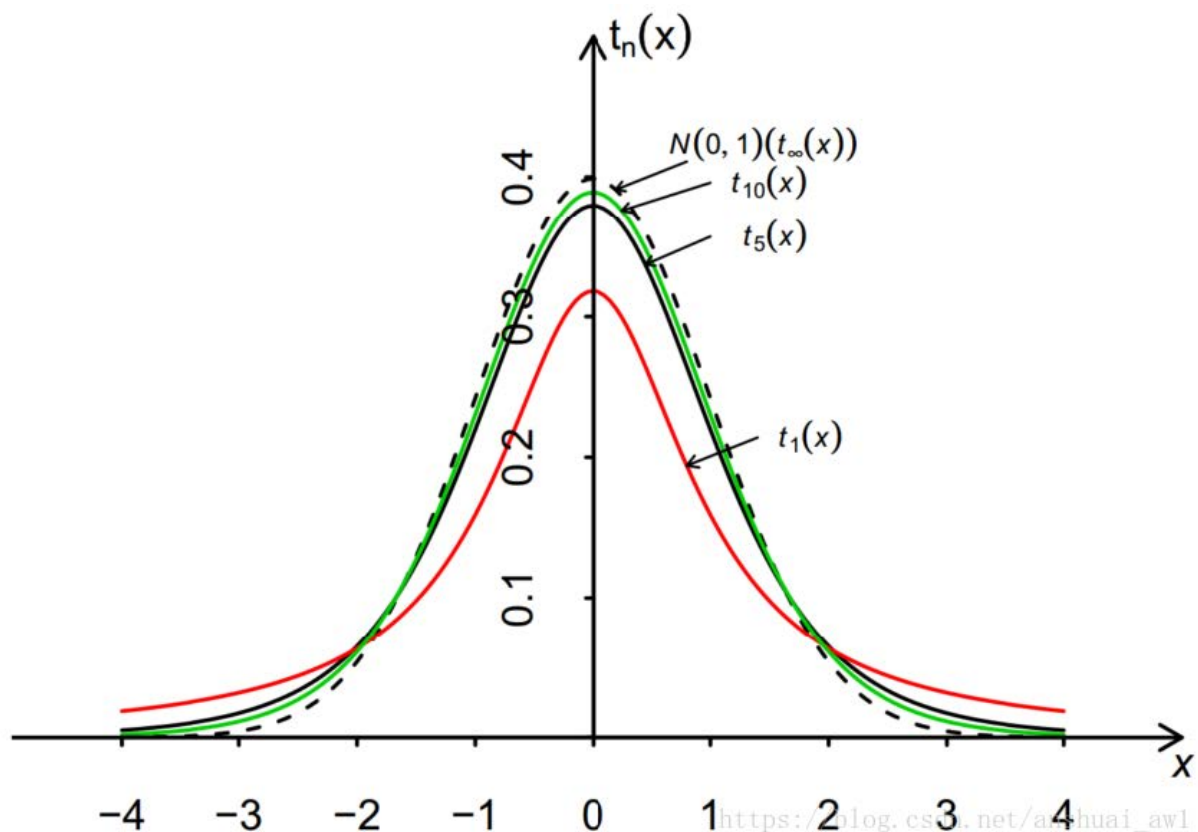
$$P\{t > t_\alpha(n)\} = \int_{t_\alpha(n)}^{\infty} h(t) dt = \alpha$$

当 $n > 45$ 时，可以由标准正态分布近似 $t_\alpha(n) \approx z_\alpha$

另由对称性可得 $t_{1-\alpha}(n) = -t_\alpha(n)$

t 分布的现实意义

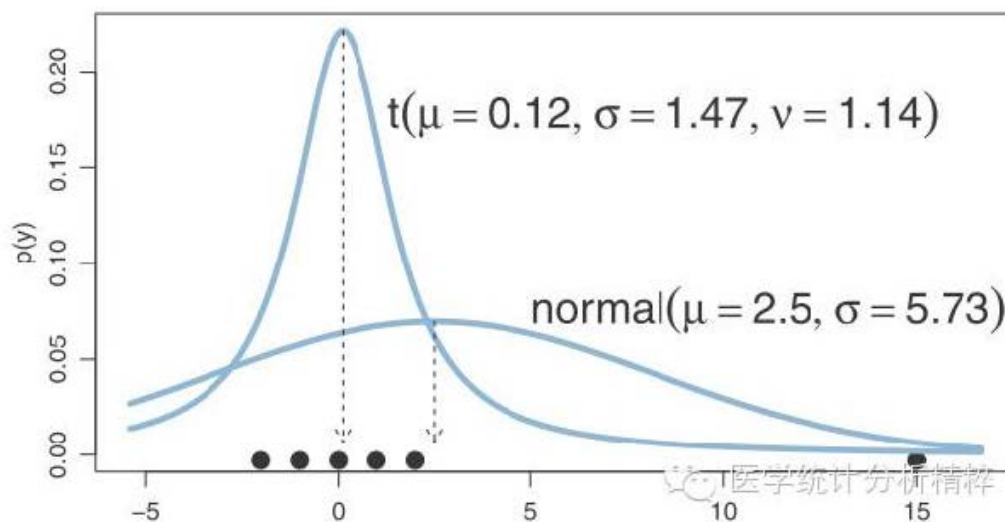
t 分布被广泛的用于小样本假设验证，用于排除异常数据的干扰，准确把握数据的集中趋势和离散趋势,即消除异常数据的干扰判断，判定模型均值与实际分布的关联性



在大样本情况下，易得有t检验与z检验等价

由于t分布的数学表示是一个标准正态分布除以一个卡方分布，前面有提到的，卡方分布表示的是随机变量相较期望偏移度的积累，因而其倒数也可以表示为对偏移的容忍度。

观察自由度趋近于无穷的t分布曲线，即标准正态分布曲线，我们可以发现，在我们要考虑问题的实际范围内，t分布相较正态分布曲线收敛度相对较低，因而有可以更好容纳偏移期望数据的特性，具有更强的鲁棒性和收敛力



这里给出的是面对同样样本的t分布和高斯分布检验结果，显然的，t分布可以较好地排除异常样本的干扰

F分布

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

$$F = \frac{U/n_1}{V/n_2}$$

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

F分布的性质

有 $F(n_1, n_2)$ 分布的概率密度

$$\psi(y) = \begin{cases} \frac{\Gamma[(n_1 + n_2)/2](n_1/n_2)^{n_1/2} y^{(n_1/2)-1}}{\Gamma(n_1/2)\Gamma(n_2/2)[1 + (n_1 y/n_2)]^{(n_1+n_2)/2}} & y > 0 \\ 0 & else \end{cases}$$

由定义知

$$F \sim F(n_1, n_2) \rightarrow \frac{1}{F} \sim F(n_2, n_1)$$

有F分布的上分位点

$$P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1, n_2)}^{\infty} \psi(y) dy = \alpha$$

另有性质

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

F分布的现实意义

F分布通常用于检验两种分布的方差之间的关系，即表现为两卡方分布对样本与期望的偏离度正则化后之比

正态总体下的分布关系

对于单分布下样本

- $\bar{X} \sim N(\mu, \sigma^2/n)$
- $\frac{S^2(n-1)}{\sigma^2} \sim \chi^2(n-1)$
- $t(n-1) \sim \frac{\bar{X} - \mu}{S/\sqrt{n}}$
- \bar{X} 和 S^2 相互独立

证明：

$$\begin{aligned}
 & \therefore \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \sim N(0, 1), \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k \\
 & \therefore \bar{X} \sim N(\mu, \sigma^2/n) \Leftrightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \\
 & \frac{s^2(n-1)}{\sigma^2} = \frac{(n-1) * \frac{1}{n-1} \sum_{k=1}^n (x_i - \bar{x})^2}{\sigma^2} \\
 & = \frac{\sum_{k=1}^n [(x_i - \mu) - (\bar{x} - \mu)]^2}{\sigma^2} \\
 & = \sum_{k=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 - n \left(\frac{\bar{x} - \mu}{\sigma} \right)^2 \\
 & = \sum_{k=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 - \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)^2 \\
 & \therefore \sum_{k=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \sim \chi^2(n), \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)^2 \sim \chi^2(1) \\
 & \therefore \frac{S^2(n-1)}{\sigma^2} \sim \chi^2(n-1) \\
 & t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \\
 & = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} / \frac{S}{\sigma} \\
 & = \frac{\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S^2(n-1)}{\sigma^2} / (n-1)}} \\
 & \therefore \frac{s^2(n-1)}{\sigma^2} \sim \chi^2(n-1) \\
 & \therefore t = \frac{N(0, 1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}} \\
 & \therefore t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t(n-1)
 \end{aligned}$$

对于两独立分布下样本

1. $\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$
 2. $\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$
- 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 时
1. $\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$
 2. $S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$

$$\begin{aligned}
 & \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1), \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1) \\
 & \therefore \text{有 } F \text{ 分布定义} \\
 & \frac{(n_1-1)S_1^2}{(n_1-1)\sigma_1^2} \bigg/ \frac{(n_2-1)S_2^2}{(n_2-1)\sigma_2^2} = \frac{\chi^2(n_1-1)}{n_1-1} \bigg/ \frac{\chi^2(n_2-1)}{n_2-1} \sim F(n_1-1, n_2-2) \\
 & \therefore \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-2) \\
 & \bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \rightarrow \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)
 \end{aligned}$$

对于方差相等的分布，定义U,V

$$U = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

$$\therefore \frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2(n_1 - 1), \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_2 - 1)$$

$$V = \frac{(n_1 - 1)S_1^2}{\sigma^2} + \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$$

$$\frac{U}{\sqrt{V/(n_1 + n_2 - 2)}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_\omega \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

$$S_\omega^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

参数估计

点估计

点估计是在处理一个估计问题，即有总体X的分布形式已知，但有一个或多个参数未知，我们可以通过总体X的一个样本（包含一个或多个个体）来估计总体未知参数的值。

点估计问题的一般提法为：设总体\$X\$的分布函数\$F(x;\theta)\$的形式已知，\$\theta\$为待估参数，\$X_k\$为\$X\$的一个样本，\$x_k\$为相应样本值，点估计问题在于构造一个适当统计量\$\hat{\theta}(X_1, X_2, \dots, X_n)\$，用其观察值作为未知参数的近似值。有称\$\hat{\theta}\$为\$\theta\$的估计量，观察值确定的估计量为估计值，统称估计

矩估计

设\$X\$为随机变量，服从分布\$F(x;\theta_1, \theta_2, \dots, \theta_k)\$，其中\$(\theta_1, \theta_2, \dots, \theta_k)\$为待估参数，有\$X_1, X_2, \dots, X_n\$为来自\$X\$的样本，我们假定总体\$X\$的前\$k\$阶矩

$$\mu_l = E(X^l) = \int_{-\infty}^{\infty} x^l f(x; \theta_1, \theta_2, \dots, \theta_k) dx$$

$$\mu_l = E(X^l) = \sum_{x \in R_X} x^l p(x; \theta_1, \theta_2, \dots, \theta_k)$$

由于已知样本矩\$A_l = \frac{1}{n} \sum_{i=1}^n X_i^l\$依概率收敛于相应总体矩\$\mu_l\$，我们使用样本矩作为相应总体矩的估计量，表现为联立解方程\$\hat{\theta}_i = \theta_i(A_1, A_2, \dots, A_k) \rightarrow \theta_i(\mu_1, \mu_2, \dots, \mu_n)\$

我们称其为矩估计量，对应观察值为矩估计值

最大似然估计

先将条件限制在单参数求解上，构建似然函数

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n \{p(x_i; \theta) | f(x_i; \theta)\}$$

即一个通过现有概率分布，既定样本值，对待定参数\$\theta\$进行估计的函数。我们使用最大估计法，在\$\theta\$的取值范围\$\Theta\$内通过最大化似然函数来获取最大的参数值\$\theta\$作为估计量，有

$$L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta)$$

称估计为最大似然估计。

对最大似然估计的求解方法有令方程

$$\frac{d}{d\theta} \ln L(\theta) = 0$$

即取对数似然方程为0时，L导数为0，L取极大值

区间估计

区间估计可以理解为加上了置信区间的点估计，给出一个区间包含了参数\$\theta\$的可能性即置信水平

置信区间

对于分布函数 $F(x; \theta)$ ，的未知参数，给定真值 α ，确定两个统计量 $\bar{\theta}, \underline{\theta}$ ，要求其满足

$$P\{\underline{\theta} < \theta < \bar{\theta}\} \geq 1 - \alpha$$

则称随机区间是 θ 置信水平为 $1 - \alpha$ 的置信区间，而 $\bar{\theta}$ 和 $\underline{\theta}$ 分别为置信水平为 $1 - \alpha$ 的双侧置信区间的置信下限和置信上限

对于标准正态分布，通常有通过上下 α 分位点来确定置信区间，即对于

$$P\left\{\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}\right\} = 1 - \alpha$$

有置信区间为

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$$

正态总体均值与方差的区间估计

对于单个总体

给定置信水平 $1 - \alpha$ ， \bar{X}, S^2 为样本的均值和方差，考虑均值 μ 和方差 σ^2 的置信区间

均值 μ

对于 **已知** σ^2 ，有枢轴量 $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ 的置信区间 $\left(\bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$

对于 **未知** σ^2 ，我们以 S 为 σ 的无偏估计进行替代，有

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

取 $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ 作为枢轴量，通过 t 分布的 α 分位点构建 μ 的置信区间

$$\left(\bar{X} \pm \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1)\right)$$

方差 σ^2

有 μ 未知，由 σ 的无偏估计 S 构建

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

取 $\frac{(n-1)S^2}{\sigma^2}$ 构建枢轴量，通过卡方分布的 α 分位点构建方差的置信区间

$$\left(\frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)}, \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)}\right)$$

对于多个总体

给定置信水平 $1 - \alpha$ ，给出两不同总体中样本的均值和方差，考虑均值间和方差间的关系

均值差 $\mu_1 - \mu_2$

对于 **已知** σ ，有样本均值为总体均值的无偏估计，有

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

即

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

取左侧函数为枢轴量，根据标准正态分布的 α 分位点构建置信区间

$$\left(\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

对于**未知** σ ，用无偏估计S代替

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_{\omega} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

取枢轴量，继续有依据t分布 α 分位点构建置信区间

$$\left(\bar{X} - \bar{Y} \pm t_{\alpha/2}(n_1 + n_2 - 2) S_{\omega} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

有 $S_{\omega}^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

方差比 σ_1^2/σ_2^2

对于总体均值未知，根据样本方差估计总体方差，即使用F分布

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

取枢轴量，根据F分布的 α 分位点构建置信区间

$$\left(\frac{S_1^2}{S_2^2} \frac{1}{F_{\alpha/2}(n_1 - 1, n_2 - 1)}, \frac{S_1^2}{S_2^2} \frac{1}{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)} \right)$$

特别的，对于方差比关系，若置信区间包含1，有方差无显著性差异

待估参数	其它参数	枢轴量分布	置信区间
μ	σ^2 已知	$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$	$\left(\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right)$
μ	σ^2 未知	$t=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}$	$\left(\bar{X}-t_{\alpha/2}\frac{S}{\sqrt{n}},\bar{X}+t_{\alpha/2}\frac{S}{\sqrt{n}} \right)$
σ^2	μ 未知	$\chi^2=\frac{(n-1)S^2}{\sigma^2}\sim \chi^2_{(n-1)}$	$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}},\frac{(n-1)S^2}{\chi^2_{\alpha/2}} \right)$
$\mu_1-\mu_2$	σ_1^2, σ_2^2 已知	$Z=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$\left(\bar{X}-\bar{Y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\bar{X}-\bar{Y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \right)$
$\mu_1-\mu_2$	$\sigma_1^2=\sigma_2^2=\sigma^2$ 未知	$t=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_{\omega}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{(n_1+n_2-2)}$ $S_{\omega}^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$	$\left(\bar{X}-\bar{Y}-t_{\alpha/2}\frac{S_{\omega}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}},\bar{X}-\bar{Y}+t_{\alpha/2}\frac{S_{\omega}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \right)$
σ_1^2/σ_2^2	μ_1, μ_2 未知	$F=\frac{S_1^2/(n_1-1)}{S_2^2/(n_2-1)}\sim F_{(n_1-1,n_2-1)}$	$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}},\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}} \right)$

单侧置信区间

对于部分统计的构建，我们只关注统计量 θ 的下限或者上限，这将带来构建单侧置信区间的需求。对于给定值 α 和样本 X ，确定统计量 \bar{X} ，若对于任意 $\theta \in \Theta$ 满足

则称随机区间 $(-\infty, \bar{X})$ 是 θ 置信水平为 $1-\alpha$ 的单侧置信区间，统计量称为置信区间上限，类似的，可以构造单侧置信区间下限

考虑正态总体 X ，若均值 μ ，方差 σ^2 均未知，构建

有单侧置信区间

假设检验

假设检验是对估计和分布的综合运用，用于检验推断是否能符合现实情况，以及用于判定我们是否愿意接受一个结论。对于解决这个问题，我们给出一组定义

原假设

表示为 $H_0:\mu=\mu_0$ ，表示假设样本的期望 μ 符合 μ_0 ，

备择假设

表示为 $H_1: \mu \neq \mu_0$ ，表示假设样本的期望 μ 不符合 μ_0

显著性水平

显著性水平 α ，用于判定假设与事实是否存在显著性差异的标准，与 α 分位点的概念联系

检验统计量

构建类似于枢轴量的检验统计量，通过对构建统计量进行检验来判定假设是否符合事实

拒绝域

当检验统计量取到某区域中的值时，选择拒绝原假设，将这样的区域定义为拒绝域，拒绝域的选择与显著性水平具有特定联系

拒绝域的构建在于防止错误决策发生，我们通常对两种主要的错误进行考虑，并通过显著性水平对犯错误的概率进行约束

1. $P\{ \text{当 } H_0 \text{ 为真时, 拒绝 } H_0 \} \leq \alpha$
2. $P\{ \text{当 } H_0 \text{ 为非真, 接受 } H_0 \} \leq \alpha$

在实际问题中，两类错误的概率总是此消彼长，在实际运用中，我们主要考虑第一类错误的概率，并为 α 取一个不至于太小的值以确保第二类错误的概率不至于飙升

显著性检验

对第一类错误的概率加以控制的检验被称作显著性检验

显著性检验存在方向性，有双边检验与单边检验等操作，需要考虑实际问题进行检验

检验问题的求解

首先，考虑通过样本均值的观察值 \bar{x} 对 μ 进行估计，因为要求 H_0 为真，所以需要保证观察均值与 μ_0 的误差不至于太大，即 $|\bar{x} - \mu_0| < k$ ，这里的 k 是与显著性相关的参数，当符合条件时，我们选择接受假设 H

（对于方差检验具有另行的逻辑）

为了方便构造统计量，我们对公式进行一定的加工，得到检验统计量
$$\frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

通常，考虑边界值即拒绝域的临界点，检验问题被构造为这样的形式

构建检验量 $Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$ ，通过 α 分位点的定义可以得出 k 对应 Z 检验量有 $k = z_{\alpha/2}$

即，当检验量 z 绝对值的取值大于 $z_{\alpha/2}$ 时，有 H_0 被拒绝，而小于时，有接受

有拒绝域 $|z| \geq z_{\alpha/2}$

再以构造单边检验为例给出另一种构造形式

同样是得到该检验问题的拒绝域为 $\bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{\alpha}$

不难发现，构造拒绝域其实跟构造置信区间是互逆的，可以直接借用区间估计的公式

接下来给出不同假设问题的检验

检验实例

Z检验

σ 已知，关于 μ 的检验

构造检验统计量 $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$

t检验

σ 未知，关于 μ 的检验

构造检验统计量 $t = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} \sim t(n-1)$

Z检验（多总体）

σ 已知，关于多个 μ 的检验

构造检验统计量 $Z = \frac{(\bar{X} - \bar{Y}) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

t检验（多总体）

σ未知（相等），关于多个μ的检验

有记 $H_0:\mu_1-\mu_2=\delta$

构造检验统计量 $t=\frac{(\bar{X}-\bar{Y}-\delta)}{S_{\omega}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$
 $S_{\omega}^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

t检验（成对数据）

用于衡量差异，采用逐对比较法

有记 $H_0:\mu_D=0$ 表示基本无差异

构造检验统计量 $t=\frac{(\bar{D}-0)}{S_D/\sqrt{n}}$

卡方检验

μ未知，关于σ的检验

有记 $H_0:\sigma^2=\sigma_0^2$ 表示方差符合 σ_0

构造检验统计量 $\chi^2=\frac{(n-1)S^2}{\sigma_0^2}\sim \chi^2(n-1)$

F检验

μ未知，关于多个σ的检验

构造检验统计量 $F=\frac{S^2_1/S^2_2}{\sigma^2_1/\sigma^2_2}\sim F(n_1-1,n_2-1)$

对于验证两主题的方差齐性，构造单边检验。有记 $H_0:\sigma_1^2\leq\sigma_2^2$ ，即限制σ之间的关系

为了确保样本方差的观察值之间的偏差不至于过大，给出拒绝域形式

通过第一类错误构建检验

即有 $k=F_{\alpha}(n_1-1,n_2-1)$

一张检验表格

条件	原假设 H_0	检验统计量	备择假设 H_1	拒绝域
σ^2 已知	$\mu\leq\mu_0\vee\mu\geq\mu_0\vee\mu=\mu_0$	$Z=\frac{(\bar{X}-\mu_0)}{(\sigma/\sqrt{n})}$	$\mu>\mu_0\vee\mu<\mu_0\vee\mu\neq\mu_0$	$Z\geq z_{\alpha}\vee Z\leq -z_{\alpha}$
σ^2 未知	$\mu\leq\mu_0\vee\mu\geq\mu_0\vee\mu=\mu_0$	$T=\frac{(\bar{X}-\mu_0)}{(S/\sqrt{n})}$	$\mu>\mu_0\vee\mu<\mu_0\vee\mu\neq\mu_0$	$T\geq t_{\alpha}(n-1)\vee T\leq -t_{\alpha}(n-1)$
σ_1^2,σ_2^2 已知	$\mu_1-\mu_2\leq\delta\vee\mu_1-\mu_2\geq\delta\vee\mu_1-\mu_2=\delta$	$Z=\frac{(\bar{X}-\bar{Y}-\delta)}{(\sqrt{\frac{(\sigma_1^2)}{(n_1-1)}+\frac{(\sigma_2^2)}{(n_2-1)}})}$	$\mu_1-\mu_2>\delta\vee\mu_1-\mu_2<\delta\vee\mu_1-\mu_2\neq\delta$	$Z\geq z_{\alpha}\vee Z\leq -z_{\alpha}$
$\sigma_1^2=\sigma_2^2=\sigma^2$ 已知	$\mu_1-\mu_2\leq\delta\vee\mu_1-\mu_2\geq\delta\vee\mu_1-\mu_2=\delta$	$T=\frac{(\bar{X}-\bar{Y}-\delta)}{(S_{\omega}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}})}$ $S_{\omega}^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$	$\mu_1-\mu_2>\delta\vee\mu_1-\mu_2<\delta\vee\mu_1-\mu_2\neq\delta$	$T\geq t_{\alpha}(n_1+n_2-1)\vee T\leq -t_{\alpha}(n_1+n_2-1)$
μ 未知	$\sigma^2\leq\sigma_0^2\vee\sigma^2\geq\sigma_0^2\vee\sigma^2=\sigma_0^2$	$\chi^2=\frac{(n-1)S^2}{(\sigma_0^2)}$	$\sigma^2>\sigma_0^2\vee\sigma^2<\sigma_0^2\vee\sigma^2\neq\sigma_0^2$	$\chi^2\geq\chi^2_{1-\alpha}(n-1)\vee\chi^2\leq\chi^2_{1-\alpha}(n-1)$
μ_1,μ_2 已知	$\sigma_1^2\leq\sigma_2^2\vee\sigma_1^2\geq\sigma_2^2\vee\sigma_1^2=\sigma_2^2$	$F=\frac{S_1^2}{S_2^2}$	$\sigma_1^2>\sigma_2^2\vee\sigma_1^2<\sigma_2^2\vee\sigma_1^2\neq\sigma_2^2$	$F\geq F_{1-\alpha}(n_1-1,n_2-1)\vee F\leq F_{\alpha}(n_1-1,n_2-1)\vee F\geq F_{1-\alpha}(n_1-1,n_2-1)$
成对数据	$\mu_D\leq 0\vee\mu_D\geq 0\vee\mu_D=0$	$T=\frac{(\bar{D}-0)}{(S_D/\sqrt{n})}$	$\mu_D>0\vee\mu_D<0\vee\mu_D\neq 0$	$T\geq t_{\alpha}(n-1)\vee T\leq -t_{\alpha}(n-1)$