第三章 假设检验

第一节 问题的提出

假设检验为统计学的一个大类，先结合实际问题需要提出一个假设（Hypothesis，不是Assumption）H_0，称为零假设，之后根据数据，按照一定的统计方法，判断能否认为H_0成立（或否定H_0）。由于数据有随机性，可能判断正确，也可能判断错误。

例1. 某人声称能够看出不透明盒子内硬币的正反面（特异功能），做试验20次，他猜对19次，能否认为他确有本领？若猜对16次、12次？问题为，判断H_0: $p = \frac{1}{2}$。

例2. 用仪器间接测量高温5次，得数据1250、1265、1245、1260、1275，此时已知温度真值为1277，问可否认为仪器存在系统偏差？问题为：假设测量值为X，判断H_0: $EX = 1277$。

第一节 问题的提出

例3. 针织品漂白时，需要考察温度对强力的影响。分别在两种温度下做了8个试验，数据为：

70摄氏度：20.5、18.8、19.8、20.9、21.5、19.5、21.0、21.2；

80摄氏度：17.7、20.3、20.0、18.8、19.0、20.1、20.2、19.1。

问两种温度下强力有无差异？如果根据以往经验可认为数据分别服从正态分布，则问题化为判断“$\mu_1 = \mu_2$”和“$\sigma_1^2 = \sigma_2^2$”。

例4. 研究抽烟与慢性支气管炎有无关系，数据为：

	有慢性支气管炎	无慢性支气管炎
抽烟	n_{11}	n_{12}
不抽烟	n_{21}	n_{22}

记$X = I\{抽烟\}$，$Y = I\{有慢性支气管炎\}$，问题变为：判断随机变量X、Y是否相互独立。

● 不独立不等同于存在因果关系。

第一节 问题的提出

例5. 由数据$X_1, \cdots, X_n$画出的直方图，猜测X服从正态分布，如何判断?
问题化为：能否认为X的分布函数为给定的$F_0(\cdot)$?

例6. 商人A在港口接到商人B的一批产品，如何确定产品质量符合要求?
验收方案如何确定才能保障双方的基本利益? （有时涉及市场管理，如商人A是购买奶粉的顾客。）

零假设（Null hypothesis）一般记为H_0，其反面称备择假设（Alternative），记为H_1。若参数空间仍为Θ，则H_0可写作H_0: $\theta \in \Theta_0$，其中$\Theta_0 \subset \Theta$是Θ的真子集，而H_1: $\theta \in \Theta \setminus \Theta_0$或: $\theta \notin \Theta_0$

我们不能观测到θ，只能观测到数据$(X_1, \cdots, X_n)$，它对应于样本空间S中的一个点。一个根据样本来给出检验结果的方法，数学上可描述为S到$[0, 1]$区间的一个（可测）函数$\varphi(X)$，$X \in S$，满足$0 \leq \varphi(X) \leq 1$。一般情况下，$\varphi(X)$仅取0或1。

第一节 问题的提出

当观测到数据$x=(x_1, \cdots, x_n)$时，计算$\varphi(x)$的值，以概率$\varphi(x)$拒绝H_0（即接受H_1），以概率$1-\varphi(x)$接受H_0（即拒绝H_1）。

当$\varphi(x)$仅取0或1时，记$S_1=\{x: \varphi(x)=0\}=\varphi^{-1}\{0\}$，$S_2=\{x: \varphi(x)=1\}=\varphi^{-1}\{1\}$。易知$S=S_1\cup S_2$，$S_1$与$S_2$不相交。

称S_1为接受域，S_2为否定域。一个检验法由其否定域唯一决定。

今后我主要用检验法$\varphi(x)$，教材主要用否定域。

检验法$\varphi(x)$可以简单扩展为随机化检验法：$\varphi(x)$可以取$[0, 1]$区间内任何值，之后以概率$\varphi(x)$否定H_0（即不否定H_1），以概率$1-\varphi(x)$不否定H_0（即否定H_1）。

随机化检验法理论上很有意义，实用中常受质疑：同样的数据，结论却可能不同。

第一节 问题的提出

● 假设H_0是否成立，及是否否定H_0，共$2 \times 2 = 4$种可能：

① H_0真，由检验法φ，否定H_0；
② H_0真，由检验法φ，不否定H_0；
③ H_0假，由检验法φ，否定H_0；
④ H_0假，由检验法φ，不否定H_0。

上面情形中②、③是好的，我们做了正确的判断，①、④是不好的，我们做了错误的判断。

称情形①为第I类错误（以真为假），情形④为第II类错误（以假为真）。

我们不可能完全杜绝两类错误，但“好的”检验法应尽量减少两类错误发生的概率。

第一节 问题的提出

一般地，当你控制第I类错误较小时，第II类错误就会较大，反之亦然。

如何比较两个检验法？如何找寻最优的检验法？

历史形成的方法是：给定检验水平$0 < \alpha < 1$（一般习惯上取0.01、0.05、0.10等），对所有满足第I类错误概率$\leq \alpha$的检验法，找寻其第II类错误概率（一致）最小的（若存在）。

定义1. 称$\beta_\varphi(\theta) =: E_\theta \varphi(X) = P_\theta((X_1, \cdots, X_n) \in S_2)$（第二个等号仅对非随机检验法）为检验法$\varphi$的功效函数；称$L_\varphi(\theta) = 1 - \beta_\varphi(\theta)$为$\varphi$的操作特性函数（OC函数）。

当$\theta \in \Theta_0$时，第I类错误的概率恰为$\beta_\varphi(\theta)$；当$\theta \notin \Theta_0$时，第II类错误的概率为$1 - \beta_\varphi(\theta)$。

第一节 问题的提出

定义2. 称φ是水平为α的检验法，若$sup\left\{\beta_{\varphi}(\theta) : \theta \in \Theta_0\right\} \leq \alpha$；称$\varphi$是水平为$\alpha$的一致最优检验（一致最大功效检验，UMP），如果其检验水平为α，且对于任一检验水平为α的检验法ψ，有

$$\beta_{\varphi}(\theta) \geq \beta_{\psi}(\theta) \qquad \forall \theta \in \Theta \setminus \Theta_0$$

● UMP检验最好，有时存在，但要求太高，许多时候不存在（当$\Theta \setminus \Theta_0$包含不止一点时，两个检验法按照此标准往往不可比）。

定义3. 称检验法φ是水平为α的无偏检验，若$\forall \theta \in \Theta \setminus \Theta_0$，有

$$\beta_{\varphi}(\theta) \geq \alpha$$

● 无偏是较“合理”的要求：当H_0不成立时，拒绝它的概率应不小于α（H_0成立时拒绝它的概率$\leq \alpha$）。

第一节 问题的提出

定义4. 称φ是水平为α的一致最优（一致最大功效）无偏检验，如果它是检验水平为α的无偏检验，且对于任一检验水平为α的无偏检验ψ，有

$$\boldsymbol{\beta}_{\varphi}(\boldsymbol{\theta}) \geq \boldsymbol{\beta}_{\psi}(\boldsymbol{\theta}) \qquad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus \boldsymbol{\Theta}_0$$

简记为UMPU检验。

UMPU检验在许多常见情形也存在，且往往是一个好的检验。

我们最想找UMP，如不行，找UMPU，如还不行，……

● 还有一致最优不变检验（UMPI）等，它经常就是UMPU（好的思想常导致一样的结果）。

第二节 N-P引理及似然比检验法

首先考虑参数空间最简单的情形：$\Theta=\{\theta_1, \theta_2\}$，$\Theta_0=\{\theta_1\}$。检验问题为：

$$H_0: \theta=\theta_1 \quad \leftrightarrow \quad H_1: \theta=\theta_2$$

不失一般性，设密度函数$f(x, \theta_1)$与$f(x, \theta_2)$不恒等，样本为$x=(x_1, \cdots, x_n)$，则似然函数为

$$L(x, \theta)=L(x_1, \cdots, x_n, \theta)=\prod_{i=1}^{n} f(x_i, \theta)$$

称$\lambda(x)=L(x, \theta_2)/L(x, \theta_1)$为似然比，它是样本的函数，是统计量。

● 直观上看（最大似然），当$\lambda(x)$太大时，应否定H_0，多大算“太大”？

---------------------end 20240307

● 似然比给出想法，检验水平α确定“太大”的界限。

第二节 N-P引理及似然比检验法

首先考虑参数空间最简单的情形：$\Theta=\{\theta_1,\theta_2\}$，$\Theta_0=\{\theta_1\}$。检验问题为：

$$H_0:\ \theta=\theta_1 \quad \leftrightarrow \quad H_1:\theta=\theta_2$$

不失一般性，设密度函数$f(x,\theta_1)$与$f(x,\theta_2)$不恒等，样本为$x=(x_1,\cdots,x_n)$，则似然函数为

$$L(x,\theta)=L(x_1,\cdots,x_n,\theta)=\prod_{i=1}^{n}f(x_i,\theta)$$

称$\lambda(x)=L(x,\theta_2)/L(x,\theta_1)$为似然比，它是样本的函数，是统计量。

- 直观上看（最大似然），当$\lambda(x)$太大时，应否定H_0，多大算“太大”？
- 似然比给出想法，检验水平α确定“太大”的界限。

第二节 N-P引理及似然比检验法

定理2.1. （Neyman-Pearson，简称N-P引理）给定$0<\alpha<1$，设

$$W_0=\{x:\ L(x,\theta_2)>\lambda_0 L(x,\theta_1)\}=\{x:\ \lambda(x)>\lambda_0\}$$

$$\varphi_0(x)=\begin{cases}1 & \text{当}\lambda(x)>\lambda_0\\ 0 & \text{否则}\end{cases}$$

其中λ_0满足$\int_{W_0}L(x,\theta_1)dx=E_{\theta_1}\varphi_0(x)=\beta_{\varphi_0}(\theta_1)=\alpha$（隐含假设此$\lambda_0$存在），则对任何水平不超过$\alpha$的检验$\varphi$，有

$$\beta_{\varphi}(\theta_2)\le\beta_{\varphi_0}(\theta_2)$$

定理说明，如此定义的检验法φ_0（如果其存在，称其为似然比检验法），是H_0的UMP检验，即W_0为UMP否定域。

下面我主要用φ_0，教材主要用W_0。

许多情况下，φ_0是存在的，但也有反例，如：

第二节 N-P引理及似然比检验法

例1. 设只取一个样本，X的分布为：

$$f(x,\theta_1)=\begin{cases}\dfrac{1}{3} & \text{当 } x\in[0,1]\\[2ex] \dfrac{2}{3} & \text{当 } x\in(1,2]\end{cases}$$

$$f(x,\theta_2)=\begin{cases}\dfrac{2}{3} & \text{当 } x\in[0,1]\\[2ex] \dfrac{1}{3} & \text{当 } x\in(1,2]\end{cases}$$

则似然比

$$\lambda(x)=\begin{cases}2 & \text{当 } x\in[0,1]\\[2ex] \dfrac{1}{2} & \text{当 } x\in(1,2]\end{cases}$$

此时，利用N-P引理中φ_0的定义（$\lambda>\lambda_0$），$\beta_{\varphi_0}(\theta_1)$只可能取0（当$\lambda_0\geq 2$, $\{\lambda>\lambda_0\}$为空集），$\frac{1}{3}$ （当$\lambda_0\in[\frac{1}{2},2)$）和1 （当$\lambda_0<\frac{1}{2}$，$\{\lambda>\lambda_0\}$为全空间）。

第二节 N-P引理及似然比检验法

因此，仅当检验水平$\alpha = {}^{1}\!/_{3}$时，N-P引理中的φ_0存在，否则不存在。

● 当允许使用随机化检验法（即允许φ_0取[0, 1]内任何值）时，即使$\alpha \neq {}^{1}\!/_{3}$，UMP检验也存在，有相应的N-P引理。

例2. 某人声称能够看出不透明盒子内硬币的正反面，做试验10次，对$H_0: p = \frac{1}{2} \leftrightarrow H_1: p = \frac{3}{4}$，试求检验水平为$\alpha$=0.05的UMP检验。

解：易知$\lambda(x) > \lambda_0$的充分必要条件为$X = \sum_{i=1}^{10} X_i > C$，常数$\lambda_0$、$C$均待定。若$H_0$成立，则$P(X \geq 9) \approx 0.0107 < \alpha$，$P(X \geq 8) \approx 0.05469 > \alpha$，故不存在非随机的UMP检验。若取

$$\varphi_0(x) = \begin{cases} 1 & \text{当 } x \geq 9 \\ 0.893 & \text{当 } x = 8 \\ 0 & \text{当 } x \leq 7 \end{cases}$$

则$\beta_{\varphi_0}\left(\frac{1}{2}\right) = E_{\frac{1}{2}}\varphi_0(X) = 1 \cdot P(X \geq 9) + 0.893 \cdot P(X = 8) + 0 \cdot P(X \leq 7) = 0.0107 + 0.893 \cdot (0.5469 - 0.0107) \approx 0.05$，即$\varphi_0$是水平为0.05的随机化的UMP检验。

第二节 N-P引理及似然比检验法

N-P引理的证明：设φ是任一水平$\leq \alpha$的检验（可以是随机化的）。

如果$\lambda(x) > \lambda_0$，则有$\varphi_0(x) = 1$，故而$\varphi_0(x) - \varphi(x) \geq 0$；

如果$\lambda(x) \leq \lambda_0$，则有$\varphi_0(x) = 0$，故而$\varphi_0(x) - \varphi(x) \leq 0$。因此总有

$$L(x,\ \theta_2)[\varphi_0(x) - \varphi(x)] \geq \lambda_0 L(x,\ \theta_1)[\varphi_0(x) - \varphi(x)] \qquad (1)$$

因此

$$\beta_{\varphi_0}(\theta_2) - \beta_{\varphi}(\theta_2) = E_{\theta_2}[\varphi_0(X) - \varphi(X)]$$

$$= \int_S [\varphi_0(x) - \varphi(x)]\, L(x,\ \theta_2) dx$$

$$\geq \int_S \lambda_0[\varphi_0(x) - \varphi(x)] L(x,\ \theta_1)\, dx \qquad \because (1)$$

$$= \lambda_0 E_{\theta_1}[\varphi_0(X) - \varphi(X)] = \lambda_0\left[\beta_{\varphi_0}(\theta_1) - \beta_{\varphi}(\theta_1)\right] \geq \lambda_0(\alpha - \alpha) = 0$$

证毕。

第二节 N-P引理及似然比检验法

定理2.2. 若$\{x\colon\ f(x,\ \theta_1)>0\}=\{x\colon\ f(x,\ \theta_2)>0\}$，$\lambda(x)$在$\theta=\theta_1$下的分布函数连续，则N-P引理中的$\varphi_0$存在，且（在a.s.意义下）是唯一的UMP检验。

证明：因为$\lambda(x)$在$\theta=\theta_1$下的分布函数连续，故存在（不一定唯一）λ_0，使得

$$P_{\theta_1}(\lambda(x)>\lambda_0)=1-P_{\theta_1}(\lambda(x)\le\lambda_0)=\alpha$$

记相应的检验法为φ_0，则$\beta_{\varphi_0}(\theta_1)=\alpha$。

记$D_0=\{x\colon\ \lambda(x)=\lambda_0\}\subset R^1$，由于分布函数连续，故而

$$P_{\theta_1}(D_0)=0=\int_{D_0}L(x,\ \theta_1)\,dx$$

推出$\mu(D_0)=0$。若存在水平$\le\alpha$的检验法φ，与φ_0不a.s.相等，即$\mu(D)=\mu\{x\colon\varphi_0(x)\ne\varphi(x)\}>0$，则在非零测集$D\setminus D_0$上，（1）式严格成立，从而$\beta_{\varphi_0}(\theta_2)>\beta_{\varphi}(\theta_2)$，即$\varphi$不是UMP检验。

第二节 N-P引理及似然比检验法

定理2.3. N-P引理中的φ_0是无偏的，即

$$\beta_{\varphi_0}(\theta_2) \geq \beta_{\varphi_0}(\theta_1) = \alpha$$

证明：（可以自己看书，我们的证明不同。实际上，任何UMP检验都是无偏的）

取随机化检验法$\psi_0(x) \equiv \alpha$，则$\beta_{\psi_0}(\theta) = \alpha$（$\forall\theta$），即$\psi_0$是无偏的。而$\varphi_0$是UMP检验，故对任意的$\theta \in \Theta \setminus \Theta_0$，有

$$\beta_{\varphi_0}(\theta) \geq \beta_{\psi_0}(\theta) = \alpha$$

例3. 设$X \sim N(\mu, 1)$，$\mu \in \Theta = \{0, 2\}$，$H_0$: $\mu = 0 \leftrightarrow H_1$: $\mu = 2$，α =0.05，求上述问题的UMP检验。

解：$f(x, 0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$，$f(x, 2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-2)^2}{2}}$，故而

第二节 N-P引理及似然比检验法

$\frac{f(x,2)}{f(x,0)} = e^{2x-2}$，因此$\lambda(X) = e^{2n\bar{X}-2n}$。

显然，$\{x: f(x,\mu) > 0\} = R^1$与$\mu$无关，且$\lambda(X)$的分布函数连续，由N-P引理得，

$$\varphi_0(X) = \begin{cases} 1 & \text{当}\lambda(X) > \lambda_0 \\ 0 & \text{否则} \end{cases}$$

其中λ_0满足$\beta_{\varphi_0}(0) = E_0\varphi_0(X) = P(\lambda(X) > \lambda_0|\mu = 0) = \alpha$。

因为“$\lambda(X) > \lambda_0$” $\Leftrightarrow$ “$\bar{X} > \frac{log\lambda_0}{2n} + 1$”，记$C = \frac{log\lambda_0}{2n} + 1$，则

$$P(\bar{X} > C|\mu = 0) = \alpha \iff \int_{\sqrt{n}C}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 0.05$$

即（查表）$\sqrt{n}C \approx 1.65$，所以UMP检验为

$$\varphi_0(X) = \begin{cases} 1 & \text{当}\bar{X} > \frac{1.65}{\sqrt{n}} \\ 0 & \text{否则} \end{cases}$$

第二节 N-P引理及似然比检验法

假设检验的思路:

- 类似于"反证法",但由于随机性,并不能导出绝对的矛盾。

- 先假设H_0成立,看是否可以找出"矛盾",如例3中,当H_0成立时,$\bar{X}$不应太大,若太大,则认为H_0不成立,将其否定。

- 称否定域$W_0=\{X: \varphi_0(X)=1\}$为"小概率事件",其概率恰为第I类错误的概率,一般控制为α =0.05(或0.01、0.10)。

- 所有H_0成立时概率为α的事件均可定义为"小概率事件",但其发生越不合理,对应的检验法越好,即其第II类错误概率越小。

- 当"小概率事件"未发生时,有人不说肯定H_0,而说不否定H_0。

第三节 单参数情形的假设检验

设X的分布函数为$F(x,\ \theta)$，其中$\theta\in\Theta=(a,\ b)$，其中$-\infty\leq a<b\leq+\infty$。本节的参数空间为一维，包含四种基本类型：

1. H_0: $\theta\leq\theta_1\quad\leftrightarrow\quad H_1$: $\theta>\theta_1$;
2. H_0: $\theta\notin(\theta_1,\ \theta_2)\quad\leftrightarrow\quad H_1$: $\theta\in(\theta_1,\ \theta_2)$;
3. H_0: $\theta\in[\theta_1,\ \theta_2]\quad\leftrightarrow\quad H_1$: $\theta\notin[\theta_1,\ \theta_2]$;
4. H_0: $\theta=\theta_1\quad\leftrightarrow\quad H_1$: $\theta\neq\theta_1$。

其中$a<\theta_1<\theta_2<b$为已知常数（Θ_0均为闭集，惯例，数学叙述方便，无统计学意义）。

定义：设$\Theta=(a,\ b)$如前所述，X的密度函数（或概率函数）可表示为

$$f(x,\ \theta)=S(\theta)h(x)e^{Q(\theta)V(x)}$$

其中$S(\theta)>0$，$h(x)>0$，$Q(\theta)$是θ的严格增函数，则称X服从单参数指数型分布。

● 二项分布、指数分布、Poisson分布、一个参数已知的正态分布等皆为单参数指数型分布。请自行验证。

第三节 单参数情形的假设检验

● 记$t(X_1, \cdots, X_n) = t(X) = \sum_{i=1}^{n} V(X_i)$，易知它是$\theta$的充分统计量。

● 对于许多单参数指数型分布的具体问题，利用直观很容易得到“好”的检验法。我们需要做的是：① 检验的优良性；② 找到确定临界值的方法。

● 上节例3中，有人可能注意到，所求出的UMP检验与$\Theta = \{0, 2\}$中的“2”无关，即若问题改为$\Theta = \{0, \mu_1\}$，$\mu_1 > 0$，则UMP检验不变！

★ 第I类错误的概率α（由定义）仅与$\Theta_0 = \{0\}$有关，与μ_1的大小无关。

★第II类错误的概率$\beta = 1 - \beta_{\varphi_0}(\mu_1) = \cdots = \Phi(1.65 - \sqrt{n}\mu_1)$，故$\mu_1$越大，$\beta$越小；$\mu_1$越小，$\beta$越大，即越不易区分0与$\mu_1$。

第三节 单参数情形的假设检验

- 将问题变的稍复杂：μ_1可取大于0的任何值，则

$$H_0: \mu = 0 \quad \leftrightarrow \quad H_1: \mu > 0$$

其UMP检验仍为φ_0。

- 再度复杂化：H_0也可扩展，则

$$H_0: \mu \le 0 \quad \leftrightarrow \quad H_1: \mu > 0$$

的UMP仍应该为φ_0。此时，功效函数$\beta_{\varphi_0}(\mu) = 1 - \Phi(1.65 - \sqrt{n}\mu)$的图像为…，是$\mu$的严格单调增函数。

- 将上述结果一般化，得到下述定理。

第三节 单参数情形的假设检验

定理3.1. 设X的分布为单参数指数型，检验问题为：

$$H_0\colon \theta \le \theta_1 \quad \leftrightarrow \quad H_1\colon \theta > \theta_1$$

若检验水平为$\alpha \in (0,\ 1)$，且存在C满足$P_{\theta_1}\left(\sum_{i=1}^{n} V(X_i) > C\right) = \alpha$，则此问题的水平为$\alpha$的UMP检验为：

$$\varphi_0(X) = \begin{cases} 1 & \text{当} \sum_{i=1}^{n} V(X_i) > C \\ 0 & \qquad\text{否则} \end{cases}$$

证明：（思路：Θ_0与Θ_1均包含无穷个点，拆成简单情形，用N-P引理）

① φ_0的水平为α，即$\forall \theta' \le \theta_1$，$\beta_{\varphi_0}(\theta') \le \alpha$。

显然$\beta_{\varphi_0}(\theta_1) = P_{\theta_1}\left(\sum_{i=1}^{n} V(X_i) > C\right) = \alpha$。对$\forall \theta' < \theta_1$，考虑假设检验问题$H_0'\colon \theta = \theta' \quad \leftrightarrow \quad H_1'\colon \theta = \theta_1$，记$\alpha' = \beta_{\varphi_0}(\theta')$，此时似然比为

$$\lambda(X) = \frac{S^n(\theta_1)}{S^n(\theta')} e^{[Q(\theta_1) - Q(\theta')] \sum_{i=1}^{n} V(X_i)}$$

因为Q严格单调增，由N-P引理，问题的水平为α'的UMP检验恰为φ_0，而φ_0是无偏的，故$\alpha' = \beta_{\varphi_0}(\theta') \le \beta_{\varphi_0}(\theta_1) = \alpha$。

第三节 单参数情形的假设检验

② φ_0是原问题的UMP检验。

设φ是水平$\leq \alpha$的任一检验，即$\forall\theta \leq \theta_1$，$\beta_\varphi(\theta) \leq \alpha$。$\forall\tilde{\theta} > \theta_1$，对假设检验问题$\tilde{H}_0$: $\theta = \theta_1 \quad \leftrightarrow \quad \tilde{H}_1$: $\theta = \tilde{\theta}$，φ也是其水平$\leq \alpha$的一个检验。由N-P引理，此问题的UMP检验为φ_0，故有$\beta_{\varphi_0}(\tilde{\theta}) \geq \beta_\varphi(\tilde{\theta})$。证毕。

Q：第II类错误的概率可以有多大？应用中有无问题？

实际应用中最关心正态分布情形，具体问题为：

一、$X\sim N(\mu, \sigma_0^2)$（方差已知），检验问题为

$$H_0: \mu \leq \mu_0 \quad \leftrightarrow \quad H_1: \mu > \mu_0$$

此时，密度函数为$f(x, \mu) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{\mu^2}{2\sigma_0^2}} e^{-\frac{x^2}{2\sigma_0^2}} e^{\frac{\mu x}{\sigma_0^2}}$，故而$Q(\mu) = \frac{\mu}{\sigma_0^2}$是$\mu$的严格单调增函数，$V(x) = x$。由定理3.1，问题的UMP检验为：

第三节 单参数情形的假设检验

$$\varphi_0(x)=\begin{cases}1 & 若\sum_{i=1}^{n}X_i>C\\ 0 & 否则\end{cases}$$

而常数C由$\beta_{\varphi_0}(\mu_0)=P_{\mu_0}\left(\sum_{i=1}^{n}X_i>C\right)=\alpha$确定。

当$\mu=\mu_0$时，$\frac{1}{n}\sum_{i=1}^{n}X_i\sim N(\mu_0,\ \frac{\sigma_0^2}{n})$，故

$$P_{\mu_0}\left(\sum_{i=1}^{n}X_i>C\right)$$

$$=P_{\mu_0}\left(\frac{\bar{X}-\mu_0}{\sqrt{\sigma_0^2/n}}>\frac{Cn^{-\frac{1}{2}}-\sqrt{n}\mu_0}{\sigma_0}\right)$$

$$=1-\Phi\left(\frac{Cn^{-\frac{1}{2}}-\sqrt{n}\mu_0}{\sigma_0}\right)=\alpha,$$

第三节 单参数情形的假设检验

其中Φ为标准正态分布函数。

给定$\alpha > 0$，查表可得C。如$\alpha = 0.05$时， $\Phi(1.65) \approx 0.95$，即$C = n(\mu_0 + \frac{1.65}{\sqrt{n}}\sigma_0)$，故

$$\varphi_0(X) = \begin{cases} 1 & \text{当} \overline{X} > \mu_0 + \dfrac{1.65}{\sqrt{n}}\sigma_0 \\ 0 & \text{否则} \end{cases}$$

例1. 设砖的强度服从正态分布$N(\mu, 1.21)$，从中随机抽取6块，测得强度为：32.56，29.66，31.64，30.00，31.87，31.03，问能否认为平均强度μ超过了30（$\alpha = 0.05$）？

解： 记H_0: $\mu \le 30 \quad \leftrightarrow \quad H_1$: $\mu > 30$，此时$n = 6$，故而$\mu_0 + \frac{1.65}{\sqrt{n}}\sigma_0 = 30 + \frac{1.65}{\sqrt{6}}\sqrt{1.21} = 30.74$，因此，UMP检验为：

$$\varphi_0(X) = \begin{cases} 1 & \text{当}\overline{X} > 30.74 \\ 0 & \text{否则} \end{cases}$$

现$\overline{X} = 31.3 > 30.74$，故否定$H_0$，即认为超过30。

另解： 记$H_0{}'$: $\mu \ge 30 \quad \leftrightarrow \quad H_1{}'$: $\mu < 30$，类似可得，

$$\varphi_0'(X) = \begin{cases} 1 & \text{当}\overline{X} < 29.26 \\ 0 & \text{否则} \end{cases}$$

因为$\overline{X} = 31.3 > 29.26$，不否定$H_0$‘，即认为超过30。

第三节 单参数情形的假设检验

● 对应此例，两种检验方法得到的结论一致。

● 但**如果**$\overline{X}$落在$[\mu_0-\frac{1.65}{\sqrt{n}}\sigma_0，\mu_0+\frac{1.65}{\sqrt{n}}\sigma_0]=[29.26, 30.74]$中，则会出现$H_0$和$H_0{}'$均不被否定的情况。矛盾？实际工作中、考试时如何选择？

二、$X\sim N(\mu_0,\sigma^2)$（期望已知），检验问题为

$$H_0: \sigma^2\le\sigma_0^2 \leftrightarrow \ H_1: \sigma^2>\sigma_0^2$$

此时，$f(x,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}exp\left\{-\frac{1}{2\sigma^2}(x-\mu_0)^2\right\}$，$Q(\sigma^2)=-\frac{1}{2\sigma^2}$单调增，$V(x)=(x-\mu_0)^2$，由定理3.1，UMP检验为：

$$\varphi_0(X)=\begin{cases}1 & 当\sum_{i=1}^{n}(X_i-\mu_0)^2>C\\ 0 & 否则\end{cases}$$

其中常数C满足$P_{\sigma_0^2}\left(\sum_{i=1}^{n}(X_i-\mu_0)^2>C\right)=\alpha$。当$\sigma^2=\sigma_0^2$时，因为$\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i-\mu_0)^2\sim\chi^2(n)$，查表得分位点，之后可得常数$C$。

第三节 单参数情形的假设检验

定理3.2. 设X服从单参数指数型分布。对检验问题

$$H_0:\ \theta \notin (\theta_1,\ \theta_2) \quad \leftrightarrow \quad H_1: \theta \in (\theta_1,\ \theta_2)$$

若存在常数$C_1 < C_2$，满足$\beta_{\varphi_0}(\theta_i) = \alpha$（$i = 1,\ 2$），而

$$\varphi_0(X) = \begin{cases} 1 & \text{当} C_1 < \sum_{i=1}^{n} V(X_i) < C_2 \\ 0 & \text{否则} \end{cases}$$

则φ_0即是此问题的UMP检验。

证明：略。

符合直观。功效函数的示意草图为……

第三节 单参数情形的假设检验

定理3.3. 设X服从单参数指数型分布。对检验问题

$$H_0: \theta \in [\theta_1, \theta_2] \quad \leftrightarrow \quad H_1: \theta \notin [\theta_1, \theta_2]$$

若存在常数$C_1 < C_2$，满足$\beta_{\varphi_0}(\theta_i) = \alpha$（$i = 1, 2$），而

$$\varphi_0(X) = \begin{cases} 1 & 当\sum_{i=1}^{n} V(X_i) < C_1 或 \sum_{i=1}^{n} V(X_i) > C_2 \\ 0 & 否则 \end{cases}$$

则φ_0即是此问题的UMPU检验。

证明：略。

● 功效函数的示意草图为……

● 此时，UMP检验不存在！直观上看，功效函数在Θ_1上越大越好，但此时Θ_1分布在两侧，抬高一侧时，另一侧会很小，即第II类错误的概率会很大。

● 增加无偏性要求，则有最优解。

第三节 单参数情形的假设检验

定理3.4. 设X服从单参数指数型分布。分布密度为

$$f(x,\theta)=S(\theta)h(x)e^{Q(\theta)V(x)}$$

其中$S(\theta)>0$，$Q'(\theta)>0$。对检验问题

$$H_0:\theta=\theta_0 \quad \leftrightarrow \quad H_1:\theta\neq\theta_0$$

若存在常数$C_1<C_2$，使得① $\beta_{\varphi_0}(\theta_0)=\alpha$，

② $E_{\theta_0}\left(\varphi_0(X_1,\cdots,X_n)\sum_{i=1}^{n}V(X_i)\right)=\alpha E_{\theta_0}\left(\sum_{i=1}^{n}V(X_i)\right)$。而

$$\varphi_0(X)=\begin{cases}1 & \text{当}\sum_{i=1}^{n}V(X_i)<C_1\text{或}\sum_{i=1}^{n}V(X_i)>C_2\\ 0 & \text{否则}\end{cases}$$

则φ_0即是此问题的UMPU检验。

证明：略。

● 功效函数的示意草图为……

--------------end 20240314

第三节 单参数情形的假设检验

定理3.4. 设X服从单参数指数型分布。分布密度为

$$f(x,\theta)=S(\theta)h(x)e^{Q(\theta)V(x)}$$

其中$S(\theta)>0$，$Q'(\theta)>0$。对检验问题

$$H_0:\theta=\theta_0 \quad \leftrightarrow \quad H_1:\theta\neq\theta_0$$

若存在常数$C_1<C_2$，使得① $\beta_{\varphi_0}(\theta_0)=\alpha$，

② $E_{\theta_0}\left(\varphi_0(X_1,\cdots,X_n)\sum_{i=1}^{n}V(X_i)\right)=\alpha E_{\theta_0}\left(\sum_{i=1}^{n}V(X_i)\right)$。而

$$\varphi_0(X)=\begin{cases}1 & \text{当}\sum_{i=1}^{n}V(X_i)<C_1\text{或}\sum_{i=1}^{n}V(X_i)>C_2\\ 0 & \text{否则}\end{cases}$$

则φ_0即是此问题的UMPU检验。

证明：略。

- 功效函数的示意草图为……

第三节 单参数情形的假设检验

- 类似于定理3.3，此时，UMP检验也不存在（习题有实例）。

- 增加无偏性要求，则有最优解。

- 与定理3.3的理论区别为，如何确定C_1与C_2。

- 与定理3.3的应用区别为，后者更易被实用者接受。

- 使用定理4.4时，理论上，可能出现悖论：检验“秤”是否准确，当样本量趋于无穷时，概率为1地会否定H_0（秤不可能绝对准）。

第三节 单参数情形的假设检验

例2. 设$X \sim N(\mu, \sigma_0^2)$，$\mu_1 < \mu_2$，检验问题为：

$$H_0: \mu \notin (\mu_1, \mu_2) \quad \leftrightarrow \quad H_1: \mu \in (\mu_1, \mu_2)$$

试求UMP检验（$\alpha > 0$）。

解：此时$t(X) = \sum_{i=1}^{n} X_i = n\bar{X}$，由定理3.2，

$$\varphi_0(X) = \begin{cases} 1 & \text{当} C_1 < n\bar{X} < C_2 \\ 0 & \text{否则} \end{cases}$$

而常数$C_1 < C_2$满足

$$\begin{cases} \Phi\left(\dfrac{C_2 - n\mu_2}{\sqrt{n}\sigma_0}\right) - \Phi\left(\dfrac{C_1 - n\mu_2}{\sqrt{n}\sigma_0}\right) = \alpha \\ \Phi\left(\dfrac{C_2 - n\mu_1}{\sqrt{n}\sigma_0}\right) - \Phi\left(\dfrac{C_1 - n\mu_1}{\sqrt{n}\sigma_0}\right) = \alpha \end{cases}$$

利用计算机可得C_1和C_2的数值解。

第三节 单参数情形的假设检验

例3. 设$X \sim N(\mu, \sigma_0^2)$，$\mu_1 < \mu_2$，检验问题为：

$$H_0: \mu \in [\mu_1, \mu_2] \quad \leftrightarrow \quad H_1: \mu \notin [\mu_1, \mu_2]$$

试求UMPU检验（$\alpha > 0$）。

解：此时仍有$t(X) = n\bar{X}$，由定理3.3，

$$\varphi_0(X) = \begin{cases} 1 & \text{当} n\bar{X} < C_1 \text{或} n\bar{X} > C_2 \\ 0 & \text{否则} \end{cases}$$

而常数$C_1 < C_2$满足

$$\begin{cases} \Phi\left(\dfrac{C_1 - n\mu_2}{\sqrt{n}\sigma_0}\right) + 1 - \Phi\left(\dfrac{C_2 - n\mu_2}{\sqrt{n}\sigma_0}\right) = \alpha \\ \Phi\left(\dfrac{C_1 - n\mu_1}{\sqrt{n}\sigma_0}\right) + 1 - \Phi\left(\dfrac{C_2 - n\mu_1}{\sqrt{n}\sigma_0}\right) = \alpha \end{cases}$$

利用计算机可得C_1和C_2的数值解。

第三节 单参数情形的假设检验

例4. 设$X \sim N(\mu, \sigma_0^2)$，$\mu_0$已知，检验问题为:

$$H_0: \mu = \mu_0 \quad \leftrightarrow \quad H_1: \mu \neq \mu_0$$

试求UMPU检验（$\alpha > 0$）。

解：此时仍有$t(X) = n\bar{X}$，由定理3.4，

$$\varphi_0(X) = \begin{cases} 1 & \text{当} n\bar{X} < C_1 \text{或} n\bar{X} > C_2 \\ 0 & \text{否则} \end{cases}$$

其中常数$C_1 < C_2$满足

$$\begin{cases} P_{\mu_0}(n\bar{X} < C_1 \text{或} n\bar{X} > C_2) = \alpha & (1) \\ E_{\mu_0}\left(\varphi_0(X) \sum_{i=1}^{n} X_i\right) = \alpha E_{\mu_0}\left(\sum_{i=1}^{n} X_i\right) & (2) \end{cases}$$

直观上看，因为$t(X) = n\bar{X} \sim N(n\mu_0, n\sigma_0^2)$关于$n\mu_0$对称，故$\varphi_0$也应关于$n\mu_0$对称，即应存在$C > 0$，使得$C_1 = n\mu_0 - C$，$C_2 = n\mu_0 + C$。容易验证，若此常数$C$满足（1），则也满足（2）。故由（1）得，

第三节 单参数情形的假设检验

$$P_{\mu_0}(n\bar{X} < n\mu_0 - C \text{或} n\bar{X} > n\mu_0 + C)$$
$$= P_{\mu_0}\left(\left|\frac{n\bar{X} - n\mu_0}{\sqrt{n}\sigma_0}\right| > \frac{C}{\sqrt{n}\sigma_0}\right) = \cdots = 2\Phi\left(-\frac{C}{\sqrt{n}\sigma_0}\right) = \alpha$$

查表可得C，例如当$\alpha = 0.05$时，$-\frac{C}{\sqrt{n}\sigma_0} = -1.96$，故而

$$\varphi_0(X) = \begin{cases} 1 & \text{当} |\bar{X} - \mu_0| > \dfrac{1.96\sigma_0}{\sqrt{n}} \\ 0 & \text{否则} \end{cases}$$

● 上述关于对称性的方法可以扩展到非正态的更广泛情形，书上有证明（不难），有兴趣的请自学。

第三节 单参数情形的假设检验

● 对于例4的背景，若H_{01}: $\mu \le \mu_0$，则UMP检验为

$$\varphi_1(X) = \begin{cases} 1 & \text{当} \bar{X} > \mu_0 + \dfrac{1.65\sigma_0}{\sqrt{n}} \\ 0 & \text{否则} \end{cases}$$

若H_{02}: $\mu \ge \mu_0$，则UMP检验为

$$\varphi_2(X) = \begin{cases} 1 & \text{当} \bar{X} < \mu_0 - \dfrac{1.65\sigma_0}{\sqrt{n}} \\ 0 & \text{否则} \end{cases}$$

φ_1和φ_2对于H_0都是有偏的，故H_0的UMP不存在。

● φ_0、φ_1和φ_2分别对应于μ的置信水平为0.95的置信区间、置信下限及置信上限。例如：对于给定的数据，否定H_0: $\mu = \mu_0 \iff \mu_0 \notin [\bar{X} - \frac{1.96\sigma_0}{\sqrt{n}}, \bar{X} + \frac{1.96\sigma_0}{\sqrt{n}}]$；又如，否定$H_{01} \iff \mu_0 \notin \left[\bar{X} - \frac{1.65\sigma_0}{\sqrt{n}},\ +\infty\right]$。

第四节 广义似然比检验法

当遇到复杂假设检验问题时，统计学家往往先考虑广义似然比检验法，这是因为它经常有较好的性质和表现。但当参数空间Θ复杂（多维）时，功效函数也复杂，缺乏单调性。故广义似然比定义为：

定义4.1. 对给定的样本值$x=(x_1,\cdots,x_n)$，若似然函数为$L(x,\theta)$，则称

$$\lambda(x)=\frac{sup\{L(x,\theta):\ \ \theta\in\Theta\}}{sup\{L(x,\theta):\ \ \theta\in\Theta_0\}}$$

为广义似然比。

- 恒有$\lambda(x)\geq 1$（与N-P引理中的似然比不同）。
- 理论上需考虑$\lambda(x)$的可测性（统计应用中一般省略）。
- 计算上又常归结于（带约束的）最大似然估计。
- 当MLE落在Θ_0中时，$\lambda(x)=1$，直观上，$\lambda(x)$太大时，否定H_0。

第四节 广义似然比检验法

定义4.2. 给定检验水平$\alpha > 0$，若存在$\lambda_0 > 0$，使得下述φ_0满足$sup\{\beta_{\varphi_0}(\theta): \theta \in \Theta_0\} = \alpha$，则广义似然比检验法为：

$$\varphi_0(x) = \begin{cases} 1 & 当\lambda(x) > \lambda_0 \\ 0 & 否则 \end{cases}$$

● 若$t(x)$是θ的充分统计量，即$L(x, \theta) = g(t(x), \theta)h(x)$，则

$$\lambda(x) = \frac{sup\{g(t(x), \theta): \ \theta \in \Theta\}}{sup\{g(t(x), \theta): \ \theta \in \Theta_0\}} = l(t(x))$$

即广义似然比总可表示为数据充分统计量的函数。

● 若“$\lambda(x) > \lambda_0$”等价于“$k(x) \in B$”，其中$k(x)$为某统计量，则$\varphi_0(x)$可以用$k(x)$表示。

● 实用中，先寻找出广义似然比检验法，再讨论其优良性。

第四节 广义似然比检验法

定义4.3*. 称检验法φ是相合的，若$\forall\theta\in\Theta\setminus\Theta_0$，有

$$\lim_{n\to\infty}\beta_\varphi(\theta)=1$$

即给定$\theta\in\Theta\setminus\Theta_0$，第II类错误的概率趋于0。

● 许多情况下，广义似然比检验法是相合的。

下面分别考察一些正态分布的检验问题。

一、$X\sim N(\mu,\sigma_0^2)$，σ_0^2已知，检验问题为：$H_0:\mu=\mu_0\leftrightarrow\quad H_1:\mu\neq\mu_0$

水平为α，此时，$\Theta_0=\{\mu_0\}$，$\Theta=(-\infty,\ +\infty)$，故

$$sup\{L(x,\mu):\ \ \mu\in\Theta_0\}=L(x,\mu_0)$$

$$sup\{L(x,\mu):\ \ \mu\in\Theta\}=L(x,\overline{x})$$

所以$\lambda(x)=\cdots=exp\left\{-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^n\overline{x}^2-2\overline{x}\sum_{i=1}^n x_i-n\mu_0^2+2\mu_0\sum_{i=1}^n x_i\right]\right\}$

$$=exp\left\{\frac{n}{2\sigma_0^2}(\overline{x}-\mu_0)^2\right\}$$

第四节 广义似然比检验法

因此 $\varphi_0(x)=\begin{cases}1 & 当\lambda(x)>\lambda_0\\0 & 否则\end{cases}=\begin{cases}1 & 当|\overline{x}-\mu_0|>C\\0 & 否则\end{cases}$

与已有结果一致，是UMPU检验。（C 的确定与α有关，前面有）

二、 $X\sim N(\mu,\sigma^2)$，σ^2未知，检验问题为：H_0: $\mu=\mu_0 \leftrightarrow \quad H_1$: $\mu\neq\mu_0$

水平为α，此时，$\Theta_0=\{(\mu_0,\ \sigma^2):\ \mu=\mu_0,\ \sigma^2>0\}$，$\Theta=\{(\mu,\ \sigma^2):\ -\infty<\mu<+\infty,\ \sigma^2>0\}$，故

$$sup\{L(x,\ \theta):\ \ \theta\in\Theta_0\}=L(x,\ \mu_0,\ \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2)$$

$$=\cdots=\left[\frac{n}{2\pi\sum_{i=1}^{n}(x_i-\mu_0)^2}\right]^{\frac{n}{2}}e^{-\frac{n}{2}}$$

$$sup\{L(x,\ \theta):\ \ \theta\in\Theta\}=L(x,\overline{x},\ \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2)$$

$$=\cdots=\left[\frac{n}{2\pi\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^{\frac{n}{2}}e^{-\frac{n}{2}}$$

第四节 广义似然比检验法

故而

$$\lambda(x)=\left[\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^{\frac{n}{2}}=\left[\frac{(\sum_{i=1}^{n}x_i-\overline{x})^2+n(\overline{x}-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^{\frac{n}{2}}=\left(1+\frac{T^2}{n-1}\right)^{\frac{n}{2}}$$

其中$T=\frac{\sqrt{n(n-1)}(\overline{x}-\mu_0)}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}}=\frac{\overline{x}-\mu_0}{\sqrt{S^2/n}}$。因为$\lambda(x)$是$|T|$的严格单调增函数，故

$$\varphi_0(x)=\begin{cases}1 & \text{当}|T|>C\\ 0 & \text{否则}\end{cases}$$

而其中C由$\beta_{\varphi_0}(\theta_0)=E_{\mu_0\sigma}\varphi_0(X)=P_{\mu_0\sigma}(|T|>C)=\alpha$决定。由前面知识，我们知道当$\mu=\mu_0$时，$T\sim t(n-1)$，故查表可得$C$。

- 可以证明，φ_0是UMPU检验。

- 直观上合理……

- 与置信区间的关系：不否定$H_0 \Leftrightarrow \mu_0\in[\overline{x}-\lambda\sqrt{S^2/n},\overline{x}+\lambda\sqrt{S^2/n}]$。

第四节 广义似然比检验法

例1. 用仪器间接测量高温5次，设测量值服从正态分布，数据为：1250、1265、1245、1260和1275，此时利用精确测量方法得知温度真值为1277，问仪器有无系统偏差（$\alpha = 0.05$）？

解： H_0: $\mu = \mu_0 = 1277$，$n = 5$，查$t(4)$分布表，得$C = 2.776$，故

$$\varphi_0(x) = \begin{cases} 1 & 当|T| > 2.776 \\ 0 & 否则 \end{cases}$$

现$|T| = \frac{18}{5.339} > 3 > 2.776$，故否定$H_0$，认为存在系统偏差。

另解：若$\alpha' = 0.01$，则相应的$C' = 4.604$，由于现$|T| = \frac{18}{5.339} < C'$，不否定$H_0$，即认为不存在系统偏差。

Q：为何检验水平变小后，反而检验不出系统偏差?

第四节 广义似然比检验法

三、$X\sim N(\mu,\sigma^2)$，σ^2未知，检验问题为：H_0: $\mu\le\mu_0 \leftrightarrow \quad H_1$: $\mu>\mu_0$

此时，参照“二”的推导，有： $sup\{L(x,\theta):\ \theta\in\Theta\}$不变，

$$sup\{L(x,\theta):\ \theta\in\Theta_0\}=\begin{cases}L\left(x,\mu_0,\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2\right) & \text{当}\bar{x}>\mu_0\\ L\left(x,\bar{x},\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right) & \text{否则}\end{cases}$$

故

$$\lambda(x)=\begin{cases}\left(1+\dfrac{T^2}{n-1}\right)^{\frac{n}{2}} & \text{当}\bar{x}>\mu_0\\ 1 & \text{否则}\end{cases}$$

注意到当$\bar{x}>\mu_0$时，$\lambda(x)$是T的单调增函数，因此

$$\varphi_0(x)=\begin{cases}1 & \text{当}T>C_1\\ 0 & \text{否则}\end{cases}$$

其中C_1满足$\beta_{\varphi_0}(\mu_0,\sigma)=E_{\mu_0\sigma}\varphi_0(X)=P_{\mu_0\sigma}(T>C_1)=\alpha$。

第四节 广义似然比检验法

● 与直观符合：如果T较小，与H_0不矛盾；太大时应否定H_0。

● 如何验证$sup\{\beta_{\varphi_0}(\theta):\ \theta \in \Theta_0\} \le \alpha$? 即如何证明当$\mu < \mu_0$时，$\beta_{\varphi_0}(\theta) \le \alpha$?

● 对同样的背景、数据和检验水平，“二”与“三”的结论是否会出现矛盾?

四、 $X \sim N(\mu, \sigma^2)$，μ未知，检验问题为：H_0: $\sigma^2 = \sigma_0^2 \leftrightarrow \quad H_1$: $\sigma^2 \ne \sigma_0^2$

此时，$\Theta_0 = \{(\mu,\ \sigma^2):\ -\infty < \mu < +\infty, \sigma^2 = \sigma_0^2\}$，$\Theta = \{(\mu,\ \sigma^2):\ -\infty < \mu < +\infty,\ \sigma^2 > 0\}$，故

$$sup\{L(x,\ \theta):\ \theta \in \Theta_0\} = L\left(x, \bar{x}, \sigma_0^2\right)$$

$$sup\{L(x,\ \theta):\ \theta \in \Theta\} = L(x, \bar{x},\ \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2) = L(x, \bar{x}, \frac{n-1}{n}s^2)$$

因此

第四节 广义似然比检验法

$$\lambda(x) = \cdots = \left[\frac{n\sigma_0^2}{\sum_{i=1}^n (x_i - \overline{x})^2}\right]^{\frac{n}{2}} e^{-\frac{n}{2}} exp\left\{\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i - \overline{x})^2\right\}$$

令$u = \frac{1}{\sigma_0^2}\sum_{i=1}^n (x_i - \overline{x})^2$，则$\lambda(x) = f(u) = e^{-\frac{n}{2}} n^{\frac{n}{2}} u^{-\frac{n}{2}} e^{\frac{u}{2}}$，其草图为……。

所以，广义似然比检验为

$$\varphi_0(x) = \begin{cases} 1 & \text{当} u < C_1 \text{或} u > C_2 \\ 0 & \text{否则} \end{cases}$$

其中$C_1 < C_2$满足

$$\begin{cases} \int_0^{C_1} g_{n-1}(y)dy + \int_{C_2}^{\infty} g_{n-1}(y)dy = \alpha & (1) \\ C_1^{\frac{n}{2}} e^{-\frac{C_1}{2}} = C_2^{\frac{n}{2}} e^{-\frac{C_2}{2}} \qquad \left(\lambda(C_1) = \lambda(C_2)\right) & (2) \end{cases}$$

其中$g_{n-1}(y)$为$\chi^2(n-1)$的分布密度函数。

● 但此时φ_0一般地不是UMPU检验（广义似然比检验非UMPU)。

第四节 广义似然比检验法

● 可以证明，若将条件（2）改为

$$C_1^{\frac{n-1}{2}} e^{-\frac{C_1}{2}} = C_2^{\frac{n-1}{2}} e^{-\frac{C_2}{2}} \qquad (2^*)$$

则得到的φ_0^*为UMPU检验。(2^*)也等价于条件

$$\int_{C_1}^{C_2} g_{n+1}(y)dy = 1-\alpha$$

● 实际应用中，上述方法都较繁琐，因此，常简单地取$C_1 < C_2$满足

$$\int_0^{C_1} g_{n-1}(y)dy = \int_{C_2}^{+\infty} g_{n-1}(y)dy = \frac{\alpha}{2}$$

此时得到的检验法不是无偏的，但除非n很小，离无偏相去不远。

● 显然，不存在UMP检验。

第四节 广义似然比检验法

五、$X \sim N(\mu,\sigma^2)$，μ未知，检验问题为：H_0: $\sigma^2 \le \sigma_0^2 \leftrightarrow \quad H_1$: $\sigma^2 > \sigma_0^2$

类似于“四”，可以推导出

$$\lambda(x) = \begin{cases} f(u) = e^{-\frac{n}{2}} n^{\frac{n}{2}} u^{-\frac{n}{2}} e^{\frac{u}{2}} & \text{当} u > n \\ 1 & \text{否则} \end{cases}$$

故

$$\varphi_0(x) = \begin{cases} 1 & \text{当} u > C_2 \\ 0 & \text{否则} \end{cases}$$

其中C_2满足$\beta_{\varphi_0}(\mu\sigma_0^2) = P_{\mu\sigma_0^2}(U > C_2) = \int_{C_2}^{+\infty} g_{n-1}(y)dy = \alpha$。

直观上看，当$u = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(x_i - \overline{x})^2$太大时，否定$H_0$。此时$\varphi_0$是UMPU检验。

- “五”是单边的，“四”是双边的。

第四节 广义似然比检验法

六、两个正态总体： $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$， X与Y相互独立，样本为$x = (x_1, \cdots, x_{n_1})$，$y =$ $(y_1, \cdots, y_{n_2})$。问题可以有若干种。

1. H_0: $\sigma_1^2 = \sigma_2^2 \leftrightarrow \quad H_1$: $\sigma_1^2 \neq \sigma_2^2$。此时易知Θ维度为4，容易推出（较繁琐），

$$sup\{L(\theta): \quad \theta \in \Theta\} = \left(\frac{1}{\sqrt{2\pi}}\right)^{n_1+n_2}\left[\frac{n_1}{\sum(x_i - \overline{x})^2}\right]^{\frac{n_1}{2}}\left[\frac{n_2}{\sum\left(y_j - \overline{y}\right)^2}\right]^{\frac{n_2}{2}} e^{-\frac{n_1+n_2}{2}}$$

$$sup\{L(\theta): \quad \theta \in \Theta_0\} = \left(\frac{1}{\sqrt{2\pi}}\right)^{n_1+n_2}\left[\frac{n_1+n_2}{\sum(x_i - \overline{x})^2 + \sum\left(y_j - \overline{y}\right)^2}\right]^{\frac{n_1+n_2}{2}} e^{-\frac{n_1+n_2}{2}}$$

● μ_1、μ_2和$\sigma_1^2 = \sigma_2^2$、$\sigma_1^2 \neq \sigma_2^2$的最大值点分别为……

第四节 广义似然比检验法

故

$$\lambda(x,\ y) = \cdots$$

$$= \left[\frac{\sum(x_i-\overline{x})^2+\sum\left(y_j-\overline{y}\right)^2}{n_1+n_2}\right]^{\frac{n_1+n_2}{2}} \Bigg/ \left[\frac{\sum(x_i-\overline{x})^2}{n_1}\right]^{\frac{n_1}{2}}\left[\frac{\sum\left(y_j-\overline{y}\right)^2}{n_2}\right]^{\frac{n_2}{2}}$$

表示成$F = \dfrac{\sum(x_i-\overline{x})^2/(n_1-1)}{\sum\left(y_j-\overline{y}\right)^2/(n_2-1)} = \dfrac{s_1^2}{s_2^2}$的函数，可得

$$\lambda(x,\ y)$$

$$= \left[\frac{n_1-1}{(n_1+n_2)(n_2-1)}F+\frac{1}{n_1+n_2}\right]^{\frac{n_1+n_2}{2}} \Bigg/ \left(\frac{n_1-1}{n_1(n_2-1)}\right)^{\frac{n_1}{2}}\left(\frac{1}{n_2}\right)^{\frac{n_2}{2}}F^{\frac{n_1}{2}}$$

(草图……)

第四节 广义似然比检验法

因此

$$\varphi_0(x,\ y)=\begin{cases}1 & 当F<C_1或F>C_2\\ 0 & 否则\end{cases}$$

表示成F是因为当H_0成立，即$\sigma_1^2=\sigma_2^2$时，

$$F=\frac{\frac{1}{\sigma_1^2}\sum(x_i-\overline{x})^2\Big/(n_1-1)}{\frac{1}{\sigma_2^2}\sum\left(y_j-\overline{y}\right)^2\Big/(n_2-1)}$$

的分子、分母各为χ^2分布除以其自由度（与未知参数无关），故为已知分布，从而可以求出常数$C_1<C_2$的值。

● 注意F是σ_1^2与σ_2^2的无偏估计之比。

第四节 广义似然比检验法

定理4.1. 设$\xi \sim \chi^2(n_1)$， $\eta \sim \chi^2(n_2)$，ξ与η相互独立，则$\zeta = \dfrac{\xi/n_1}{\eta/n_2}$的分布密度为

$$f_{n_1 n_2}(u) = \begin{cases} \dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right)}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}\left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}} u^{\frac{n_1}{2}-1}\left(1+\dfrac{n_1}{n_2}u\right)^{-\frac{n_1+n_2}{2}} & 当u>0 \\ 0 & 否则 \end{cases}$$

证明：略（随机变量函数的分布）。

---------------- end 20240318

定义4.4. 称密度函数如上的随机变量X服从第一自由度为n_1，第二自由度为n_2的F分布，记为$X \sim F(n_1, n_2)$。

- 三个重要统计分布中的最后一个，主要用于方差的比较。

- 密度函数不用背，但应会画草图，为（草图……）。

。

第四节 广义似然比检验法

定理4.1. 设$\xi \sim \chi^2(n_1)$， $\eta \sim \chi^2(n_2)$，ξ与η相互独立，则$\zeta = \dfrac{\xi/n_1}{\eta/n_2}$的分布密度为

$$f_{n_1 n_2}(u) = \begin{cases} \dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right)}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}\left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}} u^{\frac{n_1}{2}-1}\left(1+\dfrac{n_1}{n_2}u\right)^{-\frac{n_1+n_2}{2}} & \text{当} u > 0 \\ 0 & \text{否则} \end{cases}$$

证明：略（随机变量函数的分布）。

定义4.4. 称密度函数如上的随机变量X服从第一自由度为n_1，第二自由度为n_2的F分布，记为$X \sim F(n_1, n_2)$。

- 三个重要统计分布中的最后一个，主要用于方差的比较。

- 密度函数不用背，但应会画草图……。

第四节 广义似然比检验法

● F分布的性质：

① 若$X \sim F(n_1, n_2)$，则$1/X \sim F(n_2, n_1)$。
② 若$T \sim t(n)$，则$T^2 \sim F(1, n)$。

● F分布的分位数已列成表格，上两条性质可帮助查表。

● 由此，当H_0成立时，上面得到的$F = \dfrac{\sum (x_i - \bar{x})^2 / (n_1 - 1)}{\sum (y_j - \bar{y})^2 / (n_2 - 1)}$服从第一自由度为$n_1 - 1$，第二自由度为$n_2 - 1$的$F$分布。

● 实用中常数$C_1 < C_2$的取法：

$$\int_0^{C_1} f_{n_1-1,n_2-1}(u)du = \int_{C_2}^{\infty} f_{n_1-1,n_2-1}(u)du = \frac{\alpha}{2}$$

第四节 广义似然比检验法

● 可以证明， 若$C_1 < C_2$满足

$$\int_0^{C_1} f_{n_1-1,n_2-1}(u)du + \int_{C_2}^{\infty} f_{n_1-1,n_2-1}(u)du = \alpha$$

$$\int_{\alpha_1}^{\alpha_2} B_{\frac{1}{2}(n_1+1),\frac{1}{2}(n_2-1)}(y)dy = 1-\alpha$$

时，相关的φ是UMPU检验。这里，$\alpha_i = \frac{n_1-1}{n_2-1}C_i \Big/ \left(\frac{n_1-1}{n_2-1}C_i + 1\right)$，$i = 1, 2$，$B_{r,s}(y)$是$\beta$分布的密度函数。因不便查表等较少使用。

例2（第一节中例3）. 针织品漂白中强度分析。强度为:

70°C：20.5、18.8、19.8、20.9、21.5、19.5、21.0、21.2;

80°C：17.7、20.3、20.0、18.8、19.0、20.1、20.2、19.1。

问能否认为两种温度下强度的方差相同（$\alpha = 0.05$）？

解：设H_0: $\sigma_1^2 = \sigma_2^2$，此时$n_1 = n_2 = 8$，查表（$df_1 = df_2 = 7$），得

第四节 广义似然比检验法

$C_2 = 4.99$。由性质1，$C_1 = 1/4.99$，根据数据计算得，$F = \cdots = 1.07$，因为$C_1 < F < C_2$，不否定H_0，认为两种温度下强度的方差相同。

2. H_0: $\sigma_1^2 \le \sigma_2^2 \leftrightarrow \quad H_1$: $\sigma_1^2 > \sigma_2^2$

类似于1的推导，可以得到

$$\varphi_0(x,\ y) = \begin{cases} 1 & \text{当} F > C_2 \\ 0 & \text{否则} \end{cases}$$

其中C_2满足$\int_{C_2}^{\infty} f_{n_1-1,n_2-1}(u)du = \alpha$，此时$\varphi_0$是UMPU检验，而$f_{n_1-1,n_2-1}(\cdot)$是$F(n_1-1,\ n_2-1)$的密度函数。

● 直观上看，$F = \frac{S_1^2}{S_2^2}$，如果太大，则可能$\frac{\sigma_1^2}{\sigma_2^2}$大，应否定$H_0$。

● 与情况1不同的是，若F太小，与H_0看似不构成矛盾。

第四节 广义似然比检验法

3. 当σ_1^2、 σ_2^2未知，但已知$\sigma_1^2=\sigma_2^2$时， H_0: $\mu_1=\mu_2 \leftrightarrow \quad H_1$: $\mu_1 \neq \mu_2$。此时，$\lambda(x,\ y)$可表示为$|T|$的单调增函数，故广义似然比检验为：

$$\varphi_0(x,\ y)=\begin{cases} 1 & 当|T|>C \\ 0 & 否则 \end{cases}$$

其中C满足，当$\mu_1=\mu_2$时，$P(|T|>C)=\alpha$，而

$$T=\frac{\bar{x}-\bar{y}}{\sqrt{\sum(x_i-\bar{x})^2+\sum\left(y_j-\bar{y}\right)^2}}\sqrt{\frac{n_1 n_2(n_1+n_2-2)}{n_1+n_2}}$$

第四节 广义似然比检验法

定理4.2. 设$X_1, \cdots, X_{n_1}$是来自$N(\mu_1, \sigma_1^2)$的简单随机样本，$Y_1, \cdots, Y_{n_2}$是来自$N(\mu_2, \sigma_2^2)$的简单随机样本，且相互独立，则当$\sigma_1^2 = \sigma_2^2$、$\mu_1 = \mu_2$时，上述T服从$n_1 + n_2 - 2$个自由度的t分布。

证明：仍记$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$，$S_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}\left(Y_j - \bar{Y}\right)^2$，则由条件及第二章定理3.3，$\bar{X}$、$\bar{Y}$、$S_1^2$和$S_2^2$均相互独立，且

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

$$\frac{1}{\sigma_1^2}\sum_{i=1}^{n_1}(X_i - \bar{X})^2 \sim \chi^2(n_1 - 1), \quad \frac{1}{\sigma_2^2}\sum_{j=1}^{n_2}\left(Y_j - \bar{Y}\right)^2 \sim \chi^2(n_2 - 1)$$

第四节 广义似然比检验法

所以，当$\sigma_1^2=\sigma_2^2$、$\mu_1=\mu_2$时，

$$\bar{X}-\bar{Y}\sim N\left(0,\ \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\sigma_1^2\right)$$

$$\frac{1}{\sigma_1^2}\left(\sum_{i=1}^{n_1}(X_i-\bar{X})^2+\sum_{j=1}^{n_2}\left(Y_j-\bar{Y}\right)^2\right)\sim\chi^2(n_1+n_2-2)$$

且相互独立，因此，

$$T=\frac{(\bar{X}-\bar{Y})\Big/\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}{\sqrt{\frac{1}{n_1+n_2-2}\left(\sum_{i=1}^{n_1}(X_i-\bar{X})^2+\sum_{j=1}^{n_2}\left(Y_j-\bar{Y}\right)^2\right)}}\sim t(n_1+n_2-2)$$

证毕。

第四节 广义似然比检验法

例3（例2继续）. 70°C与80°C下的强度有无差别（$\alpha = 0.05$）？

解：$n_1 + n_2 - 2 = 14$，查表得$C = 2.145$，由数据，$T = 2.161 > C$，故否定H_0，即认为强度不同。

4. 当$\sigma_1^2 = \sigma_2^2$时，H_0: $\mu_1 \le \mu_2 \leftrightarrow \quad H_1$: $\mu_1 > \mu_2$

此时，

$$\varphi_0(x, y) = \begin{cases} 1 & \text{当} T > C_2 \\ 0 & \text{否则} \end{cases}$$

其中C_2满足，当$\mu_1 = \mu_2$时，$P(T > C_2) = \alpha$，此时有$T \sim t(n_1 + n_2 - 2)$，故可以查表得到C_2。

第四节 广义似然比检验法

例4（例2、例3继续）. 能否认为，70°C下的强度小于80°C下的强度（$\alpha = 0.05$）?

解：$n_1 + n_2 - 2 = 14$，H_0: $\mu_1 \leq \mu_2$，查表得$C' = 1.761$，由数据，$T = 2.161 > C'$，故否定H_0，即认为70°C下的强度不小于80°C下的强度。

● φ_0是UMPU检验。

● 如何证明第I类错误的概率不超过α?

5. 当$\sigma_1^2 \neq \sigma_2^2$时，$H_0$: $\mu_1 = \mu_2 \leftrightarrow \quad H_1$: $\mu_1 \neq \mu_2$。
此时有

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

第四节 广义似然比检验法

用S_1^2、S_2^2替代σ_1^2、σ_2^2，得到统计量

$$T=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

且当$|T|$太大时否定H_0。但此时T的精确分布未知（σ_1^2、σ_2^2无法约分）！

可以证明，当n_1、n_2较大时， T近似服从$t(m)$，其中m是与下m^*最接近的整数：

$$m^*=\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2\Bigg/\left[\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2\right]$$

- 当m较大时， $t(m)$近似标准正态分布。

第四节 广义似然比检验法

例5. 父亲患心脏病是否引起子女胆固醇偏高？随机调查甲组（父亲有心脏病）100个2-14岁子女，其胆固醇平均值为207.3，（样本）标准差为35.6；乙组（父亲无心脏病）74个2-14岁子女，其胆固醇平均值为193.4，（样本）标准差为17.3。问是否有显著差别（$\alpha = 0.05$）？

解：此时$n_1 = 100$，$n_2 = 74$。

首先检验$H_0: \sigma_1^2 = \sigma_2^2$，因为$F = S_1^2/S_2^2 = \cdots = 4.23$，而对于$\alpha = 0.05$，（$df_1 = 99, df_2 = 73$），$C_1 = 0.6548$，$C_2 = 1.5491$，故否定$H_0$，即认为$\sigma_1^2 \neq \sigma_2^2$。

现检验H_0': $\mu_1 = \mu_2$。经计算，$m^* = 151.4$，故而$m = 151$（$n_1 + n_2 - 2 = 172$），临界值$\lambda = 1.980$。现

$$T = \cdots = 3.40$$

因此否定H_0'，即认为甲组胆固醇平均值与乙组显著不同。

●应慎用引起，不随意说“因果关系”：可能是共同的基因为原因。

第四节 广义似然比检验法

● 成对数据的比较：不仅$n_1 = n_2 = n$，且$(x_1, y_1), \cdots, (x_n, y_n)$是天然成对的，例如……。

● 此时，$(x_1, y_1), \cdots, (x_n, y_n)$间相互独立，但$x_i$与$y_i$不独立，其排序不可随意交换。

待检验的假设仍为：H_0: $\mu_1 = \mu_2$，即H_0: $\mu = \mu_1 - \mu_2 = 0$。

数据转化为$z_i = x_i - y_i$，若x_i来自于$N(\mu_1, \sigma_1^2)$，y_i来自于$N(\mu_2, \sigma_2^2)$，则$z_i \sim N(\mu, \sigma^2)$，其中$\sigma_1^2$、$\sigma_2^2$不必相等，$\sigma^2$仍为未知参数。

可利用一个总体的方法做检验。

常是理想的对比检验方法（可比性最强）。

● 如果无法实现，则考虑随机化方案，实现可比性，例如药品检测中的双盲实验……。

● 如果既无法成对比较，又无法安排随机化试验，则更复杂。例如，某中学作用（人大附中、奋斗中学）、读论语得诺贝尔奖…… 。

第四节 广义似然比检验法

对于非正态情形，有时也可利用广义似然比求检验法。如：

例6. 设$X\sim\frac{1}{\theta_1}e^{-\frac{x}{\theta_1}}$（$x>0$），$Y\sim\frac{1}{\theta_2}e^{-\frac{y}{\theta_2}}$（$y>0$），简单随机样本为：$X_1, \cdots, X_{n_1}$，$Y_1, \cdots, Y_{n_2}$，且相互独立，问题为：$H_0: \theta_1 \le \theta_2 \leftrightarrow H_1: \theta_1 > \theta_2$，求广义似然比检验。

解：似然函数为：

$$L(\theta_1, \theta_2) = \frac{1}{\theta_1^{n_1}}e^{-\frac{\sum X_i}{\theta_1}}\frac{1}{\theta_2^{n_2}}e^{-\frac{\sum Y_j}{\theta_2}}$$

记$T_1 = \sum X_i$，$T_2 = \sum Y_i$，则

$$sup\{L(x, \theta): \theta \in \Theta\} = \frac{{n_1}^{n_1}}{{T_1}^{n_1}}e^{-n_1}\frac{{n_2}^{n_2}}{{T_2}^{n_2}}e^{-n_2}$$

$$sup\{L(x, \theta): \theta \in \Theta_0\} = \begin{cases} sup\{L(x, \theta): \theta \in \Theta\} & \text{当}\ \frac{T_1}{n_1} \le \frac{T_2}{n_2} \\ \left(\frac{n_1+n_2}{T_1+T_2}\right)^{n_1+n_2} e^{-(n_1+n_2)} & \text{否则} \end{cases}$$

第四节 广义似然比检验法

故$\lambda(x,\ y)$可表示为T_1/T_2的单调增函数，因此广义似然比检验为：

$$\varphi_0(x,\ y)=\begin{cases}1 & \text{当 } T_1/T_2>C\\ 0 & \text{否则}\end{cases}$$

如何确定常数C?

因为$X_i \sim Exp(\frac{1}{\theta_1})$，易知$T_1=\sum X_i \sim \Gamma(n_1,\ {}^1/_{\theta_1})$，故 ${}^{2T_1}/_{\theta_1} \sim \Gamma(n_1,\ {}^1/_2)$，即$2n_1$个自由度的$\chi^2$分布。同理，${}^{2T_2}/_{\theta_2} \sim \chi^2(2n_2)$。

当$\theta_1=\theta_2$时，由定理4.1，

$$F=\frac{\frac{2T_1}{\theta_1}\Big/_{(2n_1)}}{\frac{2T_2}{\theta_2}\Big/_{(2n_2)}}=\frac{n_2}{n_1}\cdot\frac{T_1}{T_2}\sim F(2n_1,2n_2)$$

第四节 广义似然比检验法

可以证明，φ_0是UMPU检验，直观上也合理。

如何证明φ_0的水平为α？对于任意的$\theta_1 < \theta_2$，

$$\frac{n_2}{n_1} \cdot \frac{T_1}{T_2} = \frac{\theta_1}{\theta_2} \cdot F < F \sim F(2n_1, 2n_2)$$

故而$P(T_1/T_2 > C) = P\left(\frac{\theta_1}{\theta_2} \cdot F > \lambda_0\right) < P(F > \lambda_0) = \alpha$。

● 广义似然比检验不一定都好：书上P99 例4.8构造了一个例子（人造的），说明当$\theta \in \Theta_1$时，$\beta_\varphi(\theta) < \alpha$（不是无偏的），甚至不如随机化检验法$\psi \equiv \alpha$。但多数常用情况下，广义似然比检验是好的。

● 假设检验与置信区间的关系（对已有结果）： 若由数据，水平α下不否定H_0: $\theta = \theta_0$，则充分必要地，相应的$1-\alpha$水平的置信区间包含θ_0。这提供了一种构造置信区间的方法。置信（上）下限也同理。

第五节 临界值与p值

对一般的假设检验问题，统一给出p-value的严格定义较难。但p值的概念很重要，医学界很常用。我们通过两个简单例子进行说明。

例1. 设$X_1, \cdots, X_{10} \sim N(\mu, \sigma_0^2)$，$\sigma_0^2 = 10$，若$\overline{X} = 13.9$，求$H_0$: $\mu = 12$的p值。

解：记$\eta = \dfrac{\overline{X}-12}{\sqrt{\sigma_0^2/_n}} = 1.9$，查正态分布表得，$\Phi(1.90) = 0.9713$，故

$$p = 2 \times (1 - 0.9713) = 2 \times 0.0287 = 5.74\%$$

例2. 设$X_1, \cdots, X_{10} \sim N(\mu, \sigma_0^2)$， $\sigma_0^2 = 10$，若$\overline{X} = 13.9$，求H_0: $\mu \le 12$的p值。

解：记$\eta = \dfrac{\overline{X}-12}{\sqrt{\sigma_0^2/_n}} = 1.9$，查正态分布表得，$\Phi(1.90) = 0.9713$，故

$$p = 1 - 0.9713 = 0.0287 = 2.87\%$$

第五节 临界值与p值

- （不严格地说）p值为当H_0成立时，产生如观测数据同样“古怪”，或“更古怪”的数据的概率。

- 上两例中，“古怪”的意义随H_0变化。

- 实用中，p值在适中范围内表示正常，太小表示H_0应受（强烈）怀疑，但不简单回答“否定”或“不否定”。

- p值很小时，常用$p \le 0.01$、$p \le 0.001$等方式表达。

- Fisher对孟德尔数据进行检验的p值为十万分之四。

第六节* 比率的假设检验

略。

第七节 拟合优度检验

Goodness-of-fit是统计学的一个重要方向。数据是统计分析的原材料，但建立模型是统计分析的基础。用模型去拟合某数据，拟合的好不好?需要检验!

对于同样的数据，不同的研究者可能建立不同的模型，采用不同的统计方法，从而得到不同的结果。这是正常的。

常说的“设随机变量X服从**分布”，就是简单的“建模”。

---------------- end 20240321

统计（数学）模型合理性的来源：一是相关背景知识，如由中心极限定理，误差服从正态分布，又如由传统数据知，轴承寿命服从Weibull分布。二是通过拟合优度检验。

当数据所含信息较少时，往往难以判断哪个模型合适，但当信息较多时，可以用检验的方法排除一些模型，或对其给出负面评价。

第七节 拟合优度检验

Goodness-of-fit是统计学的一个重要方向。数据是统计分析的原材料，但建立模型是统计分析的基础。用模型去拟合某数据，拟合的好不好?需要检验!

对于同样的数据，不同的研究者可能建立不同的模型，采用不同的统计方法，从而得到不同的结果。这是正常的。

常说的“设随机变量X服从**分布”，就是简单的“建模”。

统计（数学）模型合理性的来源：一是相关背景知识，如由中心极限定理，误差服从正态分布，又如由传统数据知，轴承寿命服从**分布。二是通过拟合优度检验。

当数据所含信息较少时，往往难以判断哪个模型合适，但当信息较多时，可以用检验的方法排除一些模型，或对其给出负面评价。

第七节 拟合优度检验

一、χ^2检验法

设随机变量X的值域为$\{t_0, t_1, \cdots, t_m\}$，即$X$为取有限个值的离散型随机变量，假设为：$H_0$: X的分布为F_0，其中F_0为某已知分布。此时可以用m个已知的非负实数p_k表示：$P_0(X = t_k) = F_0(t_k) - F_0(t_k -) =: p_k$，（$k = 0, 1, \cdots, m$, $\sum_{k=0}^{m} p_k = 1$）。数据为$\sum_{k=0}^{m} \nu_k = n$，其中n为样本量，ν_k为取到t_k的频数。

令

$$V = \sum_{k=0}^{m} \left(\frac{\nu_k}{n} - p_k\right)^2 \frac{n}{p_k} = \sum_{k=0}^{m} \frac{(\nu_k - np_k)^2}{np_k}$$

可以证明，若H_0成立，则当$n \to \infty$时，V依分布收敛于m个自由度的χ^2分布，即当n较大时，V近似地服从$\chi^2(m)$。

● 证明思路：中心极限定理。

第七节 拟合优度检验

● 直观上看，若H_0成立，则V的值应较小，若“太大”，则应否定H_0。多大算“太大”？由渐近分布$\chi^2(m)$，可以找到近似的临界值。

例1. 某工厂近5年来共发生了63次事故，按照星期几分类，得：

星期	一	二	三	四	五	六
次数	9	10	11	8	13	12

问事故发生是否与星期几有关（$\alpha = 0.05$）？

解：“事故发生与星期几无关”$\Leftrightarrow$“事故每天发生等可能”。取X为某事故发生后的相应随机变量：“$X = k$”$\Leftrightarrow$“事故发生在星期k”，则

$$H_0: P(X = k) = \frac{1}{6} \qquad k = 1, \cdots, 6$$

经按公式计算，$V = \sum_{k=1}^{6} \frac{\left(\nu_k - {63}/{6}\right)^2}{{63}/{6}} = \cdots = 1.67$，查表（$df = 5$）得临界值$C = 11.07$，故不否定$H_0$，即认为事故发生与星期几无关。

● 注意，此时真实的第I类错误概率≈ 0.05。

第七节 拟合优度检验

- 当X为连续型随机变量时，自然的想法就是离散化。在实轴上适当地取m个点$t_1 < \cdots < t_m$，则实轴被分为$m+1$段：$(-\infty, t_1]$，$(t_1, t_2]$，…，(t_m, ∞)。

- 记离散化后的概率分布为$F_0(\cdot)$，记p_k为X落入第k段的（理论）概率，即$p_1 = F_0(t_1)$，$p_k = F_0(t_k) - F_0(t_{k-1})$，（$k = 2, \cdots, m$），$p_{m+1} = 1 - F_0(t_m)$。仍记$\nu_k$为$X$落入第$k$段的频数，则仍有

$$V = \sum_{k=1}^{m+1} \left(\frac{\nu_k}{n} - p_k\right)^2 \frac{n}{p_k} = \sum_{k=1}^{m+1} \frac{(\nu_k - np_k)^2}{np_k}$$

当n较大时，V近似地服从$\chi^2(m)$，故可以检验假设

$$H_0: F(\cdot) = F_0(\cdot)$$

- 这是与直方图对应的检验法：m及$t_1 < \cdots < t_m$的取法可完全参照直方图方法。

- 当X为离散型，但m较大，而样本量n不大时，应对$t_1, \cdots, t_m$适当合并，以使ν_k大小适当，如体育彩票（结果依赖于分析者）。

- 此时，实际上是将数据合理“分组”。

第七节 拟合优度检验

更一般地，我们只想检验样本是否来源于某种类型的分布（例如正态），即建模的可能假设是否合理。此时，并没有一个给定的$F_0(\cdot)$，故称为复杂假设。

一般形式为：

$$H_0: F(x) \in \{F_0(x,\ \theta_1,\ \cdots,\ \theta_k):\ (\theta_1,\ \cdots,\ \theta_k) \in \Theta_0\}$$

有可能$\Theta_0 = \Theta$，即，假设样本分布属于某个已知类型（如正态），但其参数未知。

解决思路为：

① 先在H_0成立的假设下进行参数估计（MLE），得到$\hat{\theta} = \left(\hat{\theta}_1,\ \cdots,\ \hat{\theta}_k\right) \in \Theta_0$，即寻找拟合最好的$F_0$；

② 对于简单假设$\widetilde{H}_0: F(x) = F_0\left(x, \hat{\theta}_1,\ \cdots,\ \hat{\theta}_k\right)$做检验，若对此拟合最好的假设都否定，则也否定原复杂假设H_0。

第七节 拟合优度检验

③ 自由度：$m-k-1$，其中m为组数，k为参数个数（可以证明）。

例2. 观测某放射性物质20秒内放射的粒子数，重复500次，得到数据为：
$\nu_0=14, \nu_1=35, \nu_2=70, \nu_3=105, \nu_4=102, \nu_5=81, \nu_6=52,$
$\nu_7=23, \nu_8=8, \nu_9=5$，而大于等于10个的合计$\nu_{10}=5$。问能否认为服从Poisson分布（概率论有推导）（$\alpha=0.05$）？

解：记

$$p_0(x,\ \lambda)=\frac{\lambda^x}{x!}e^{-\lambda} \qquad (x=0,\ 1,\ 2,\ \cdots)$$

设（复杂假设）

$$H_0: p(x)\in\{p_0(x,\ \lambda): \lambda>0\}$$

则由数据，MLE为$\hat{\lambda}=\frac{\sum_{i=1}^n x_i}{n}\approx\frac{1}{n}\sum_{k=0}^{10}k\nu_k=\cdots=3.87$。

第七节 拟合优度检验

经计算，可得20秒内放射的粒子数恰为k的（估计）概率$p_0, p_1, \cdots, p_9$，及大于等于10个的概率p_{10}。

之后检验简单假设$\widetilde{H}_0$: $p_0(k) = p_k \ (k = 0, \cdots, 10)$，得到

$$V = \cdots = 7.275 < 16.9$$

其中16.9为$\alpha = 0.05$水平下自由度$df = 11 - 1 - 1 = 9$时χ^2分布的临界值，因此不否定$\widetilde{H}_0$及H_0，即认为服从Poisson分布。

例3. 从一批滚珠中随机抽取了50个，测得直径为（书P124例7.3）：

15.0，15.8，15.2，15.1，15.9，14.7，14.8，15.5，15.6，15.3，
15.1，15.3，15.0，15.6，15.7，14.8，14.5，14.2，14.9，14.9，
15.2，15.0，15.3，15.6，15.1，14.9，14.2，14.6，15.8，15.2，
15.9，15.2，15.0，14.9，14.8，14.5，15.1，15.5，15.5，15.1，
15.1，15.0，15.3，14.7，14.5，15.5，15.0，14.7，14.6，14.2。

问能否认为滚珠直径服从正态分布（$\alpha = 0.05$）？

第七节 拟合优度检验

解：设直径X的分布函数为$F(x)$，则问题可表示为

$$H_0: F(x) \in \left\{\Phi\left(\frac{x-\mu}{\sigma}\right):\ \mu \in R^1,\ \sigma > 0\right\}$$

首先通过MLE法估计参数，得

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \cdots = 15.1$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 = \cdots = (0.4325)^2$$

因此

$$\widetilde{H}_0: X \sim N(15.1,\ 0.4325^2)$$

根据数据，取$a = 14.05$，$b = 16.15$，将$[a, b]$分为7段（参照直方图法），得$a < t_1 < \cdots < t_6 < b$。查正态分布表，可得$\hat{p}_k$的值为

第七节 拟合优度检验

$$\hat{p}_1 = 0.0414, \hat{p}_2 = 0.1077, \hat{p}_3 = 0.2154, \hat{p}_4 = 0.2710,$$
$$\hat{p}_5 = 0.2154, \hat{p}_6 = 0.1077, \hat{p}_7 = 0.0414$$

分别计算ν_k（数数），可算得

$$V = \sum_{k=1}^{7} \frac{(\nu_k - n\hat{p}_k)^2}{n\hat{p}_k} = \cdots = 1.7284$$

查χ^2分布表（ df = 7-1-2=4，$\alpha = 0.05$)，得临界值为$\lambda =$ 9.49。

因为$V < 9.49$，故不否定$\widetilde{H}_0$，从而也不否定H_0，即认为X服从正态分布。

● Fisher对孟德尔遗传实验数据的统计分析，用的也是χ^2检验法。但他并不是当统计量太大时否定H_0，而是结合背景，当统计量太小时否定H_0。

第七节 拟合优度检验

二、独立性检验

若随机变量X与Y皆为离散型，不妨设X的值域为$\{1, 2, \cdots, s\}$，Y的值域为$\{1, 2, \cdots, t\}$，(X, Y)的联合分布为$\left(p_{ij}\right)_{s\times t}$，其中$p_{ij}=P(X=i, Y=j)$。

记$p_i=P(X=i)=\sum_{j=1}^{t}p_{ij}$， $q_j=P(Y=j)=\sum_{i=1}^{s}p_{ij}$分别为$X$、$Y$的边缘分布，由概率论知识，$X$与$Y$相互独立的充要条件是：$p_{ij}=p_iq_j$（对任意的$i=1, 2, \cdots, s$，$j=1, 2, \cdots, t$）。

设观测数据为$\left(n_{ij}\right)_{s\times t}$，其中$n_{ij}$为“$X=i, Y=j$”的频数，则数据为二维列联表。记

$$n_{i\cdot}=\sum_{j=1}^{t}n_{ij} \quad \text{“}X=i\text{”的频数}$$

$$n_{\cdot j}=\sum_{i=1}^{s}n_{ij} \quad \text{“}Y=j\text{”的频数}$$

第七节 拟合优度检验

$$n=\sum_{i=1}^{s} n_{i\cdot}=\sum_{j=1}^{t} n_{\cdot j}=\sum n_{ij}\quad \text{总数}$$

$$H_0: p_{ij}=p_i q_j\ \ (i=1,2,\cdots,s,j=1,2,\cdots,t)$$

这也是复杂假设检验问题，具体方法与“一”类似：先通过MLE方法估计参数$\hat{p}_i={n_{i\cdot}}/{n}$，$\hat{q}_j={n_{\cdot j}}/{n}$，之后将其代入（此时全部数据被分成$s\cdot t$个组，我们共估计了$(s-1)+(t-1)$个参数）得

$$V=\sum_{i=1}^{s}\sum_{j=1}^{t}\frac{\left(n_{ij}-n\hat{p}_i\hat{q}_j\right)^2}{n\hat{p}_i\hat{q}_j}$$

可以证明，当H_0成立时，若n较大，则V近似地服从χ^2分布，自由度为：$st-1-(s-1)-(t-1)=(s-1)(t-1)$。

引理：对上述问题，有

$$V=n\left(\sum_{i=1}^{s}\sum_{j=1}^{t}\frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}}-1\right)$$

证明：略。

第七节 拟合优度检验

例3. 研究吸烟与患慢性支气管炎有无关系。调查了339人，结果如下表（书P127例7.4），问可否认为有相关关系？

	患病	不患病	$n_{i\cdot}$
吸烟	43	162	205
不吸烟	13	121	134
$n_{\cdot j}$	56	283	339

解：此为2×2列联表。由前面结果，$(s-1)(t-1)=1$，设

$$H_0: p_{ij}=p_i q_j \quad (i=1,\ 2,\ j=1,\ 2)$$

由引理可知，$V=\cdots=7.48$，查表，当$\alpha=0.05$时，$\lambda=3.84$；当$\alpha=0.01$时，$\lambda=6.63$，均$<V$，故否定H_0，及认为吸烟与患慢性支气管炎显著相关。

- 此种独立性检验法是"一"中χ^2检验扩展到二维的一个应用。
- 当X、Y为连续型时，可以先将其离散化，再行处理⋯⋯。

第七节 拟合优度检验

三、Kolmogorov检验法（Kolmogorov-Smirnov）

由第二章定理4.1（Glivenko-Cantelli定理），经验分布函数$F_n(\cdot)$一致收敛于$F(\cdot)$。欲检验H_0: $F(\cdot)=F_0(\cdot)$，自然想到比较$F_n(\cdot)$与$F_0(\cdot)$。当X是连续型随机变量时，可以使用Kolmogorov-Smirnov（简称K-S）检验法。因为上述函数均有界，故可使用极大模定义检验统计量

$$D_n = sup\{|F_n(x)-F_0(x)|:\ -\infty < x <+\infty\}$$

利用连续型随机变量结论“若$X \sim F(\cdot)$，则$F(X)\sim U(0,\ 1)$”，记$Y=F_0(X)$，则当H_0成立时， $Y\sim U(0,\ 1)$，故而（令$y=F_0(x)$）

$$\begin{aligned} D_n &= sup\{|F_n(x)-F_0(x)|:\ -\infty < x <+\infty\} \\ &= sup\left\{\left|F_n\left(F_0^{-1}(y)\right)-y\right|:\ -\infty < x <+\infty, y=F_0(x)\right\} \\ &= sup\left\{\left|\widetilde{F}_n(y)-y\right|:\ 0<y<1\right\} \end{aligned}$$

其中$\widetilde{F}_n(y)$为$Y_i=F_0(X_i)\sim U(0,\ 1),\ i=1,\ \cdots,\ n$的经验分布函数。

即当H_0成立时，D_n的分布与$F_0(\cdot)$无关！因此，可以制作表格，以便查找临界值。

第七节 拟合优度检验

- 当样本量n较小时，书P362列出了临界值表。

- 当样本量n较大时，Kolmogorov证明了如下结论：

定理7.1. 设X的分布函数为连续函数$F_0(x)$，则

$$D_n = sup\{|F_n(x) - F_0(x)|: \ -\infty < x < +\infty\}$$

满足$\lim\limits_{n\to\infty} P\left(\sqrt{n}D_n \le z\right) = Q(z)$（$\forall z \in R^1$），其中

$$Q(z) = \begin{cases} \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2z^2} & 当z > 0 \\ 0 & 否则 \end{cases}$$

证明：略。

第三章 假设检验

● 传统假设检验对两类错误的态度（不对等）：控制第I类错误，不超过给定的检验水平α，一般较小；不控制第II类错误，可能很大。多大?

● 当样本量较小时，往往有各种假设均不被否定的情形，如对同一组数据，H_0：$\mu=\mu_0$、H_0：$\mu=\mu_0+\varepsilon_0$、 H_0：$\mu\le\mu_0$、 H_0：$\mu\ge\mu_0$等均不被否定。有人不说“肯定”，而说不否定，表示“否定”证据不足，但也不足以“肯定”。

● 保护消费者：美国FDA药品疗效标准的转变，食品细菌超标（传统上不出事就行，现代较严），澳大利亚奶粉重量 …… 。从统计观点，是零假设形式的改变。

● 与法律关系：H_0的选取为被告（弱势）无罪：无罪推定、疑罪从无。宁可不判（第II类错误），不能出现冤案（第I类错误）。南京彭宇案……。当然法律（形式上）要求更高（$\alpha=0$）。但如何避免被指责“不作为”（第II类错误太大）？谁主张，谁举证，即原告承担增加样本量费用。

● 对于具体问题，到底应假设H_0：$\mu\le\mu_0$还是 H_0：$\mu\ge\mu_0$？实际工作/考试

许多时候，充分利用直观和概率统计知识，可以构造好的检验法。

第三章 假设检验

例4. 在美国北卡的Raleigh，气象学家记录了1942至1981年间共40年的年平均气温（华氏），数据如书P138表8.1所示。

年份	温度	年份	温度	年份	温度	年份	温度
1942	61.0	1952	60.3	1962	58.2	1972	58.5
43	60.6	53	61.3	63	57.5	73	59.8
44	59.8	54	60.1	64	58.9	74	59.7
45	60.3	55	59.6	65	59.1	75	59.6
46	60.4	56	59.9	66	58.7	76	58.7
47	59.1	57	60.1	67	59.1	77	58.6
48	59.8	58	57.4	68	58.4	78	59.1
49	61.4	59	59.9	69	56.9	79	58.6
50	59.9	60	57.9	70	57.8	80	59.3
51	60.1	61	58.7	71	59.0	81	58.3

来源：The Fascination of Statistics, Brook et al. Marcel Dekker 1986.

第三章 假设检验

从散点图看，似乎有温度降低的趋势。问：Raleigh是否在变冷?

解：考虑任意两个不同年份温度对比形成的“对儿”：若后面年份的温度比前面年份的高，称其为“协调对儿”或“+1”（例如1944年/1946年）；否则称其为“不协调对儿”或“-1”（例如1942年/1943年）。

设H_0:无趋势，取$M =$ “+1”个数$-$ “-1”个数。直观上形成的检验法为，若M太大，说明在变暖，否定H_0； 若M太小，说明在变冷，否定H_0；若M适中，则不否定H_0。

M多大算“太大”？或具体地，临界值是多少?缺M的分布，缺概率空间。

因为年份数$n = 40$，故n个不同的温度共有$n!$种不同排列方式。记$\Omega = \{\omega\colon \omega$为$n!$种排列之一$\}$，$\mathcal{F}$为$\Omega$的所有子集构成的集合，定义

$$P(\{\omega\}) = 1/n!$$

第三章 假设检验

即所有排列方式等可能，则$(\Omega, \mathcal{F}, P)$为概率空间。

所有的“对儿”共有C_n^2个，因此，给定n后，M的分布是固定，即已知的，故而可以得到临界值（可计算，有表可查）。

此例中，$M=-329$，观测到≤ -329的概率≤ 0.001，故否定H_0，即认为Raleigh在变冷。

● 此检验法称为Mann-trend test。此方法的一大优点在于，未对温度的分布做出任何假设，仅假设n个不同的温度随机排列。

● 还可以有其他方法，如下章的回归分析方法，但其隐含线性假设。

请有兴趣的同学自学第八节。

第四章 回归分析与线性模型

第一节 引言

● 统计学中非常重要的一个领域，研究两个或多个变量间的关系。

● 相对简单，研究的较充分。

● 实用性强，应用前景广阔。

● 不断得到各种类型的推广，新的研究活跃。

------------- end 20240328

● “回归”（Regression）一词最早由十九世纪英国科学家Galton（高尔顿）在研究遗传规律时使用。父子身高/智商?

● “线性模型”指用线性组合描述变量之间的近似关系。