第二节 序贯概率比检验法

定义1. 称序贯检验法$\Delta = (\tau^*, d^*)$为序贯概率比检验（SPRT），如果

$$\tau^* = inf\{n: n \geq 1, \lambda_n \notin (A, B)\}$$

$$d^* = \begin{cases} 1 & 若\lambda_{\tau^*} \geq B \\ 0 & 若\lambda_{\tau^*} \leq A \end{cases}$$

简记为$S(A, B)$。

由λ_n的定义，$\log(\lambda_n) = \sum_{i=1}^{n} log\left(\frac{f_2(X_i)}{f_1(X_i)}\right)$是一个随机游动，因此

$$\tau^* = inf\{n: n \geq 1, log(\lambda_n) \notin (logA, logB)\}$$

是停时，即$\Delta = (\tau^*, d^*)$确为序贯检验法。

定理1. 如果$\mu\{x: f_1(x) \neq f_2(x)\} > 0$，则

$$P_i(\tau^* < \infty) = 1 \qquad (i = 1, 2)$$

证明：略。

第二节 序贯概率比检验法

例1. 两点分布的检验，$X \sim B(1, p)$，

$$H_1: p = p_1 \quad \leftrightarrow H_2: p = p_2$$

其中$0 < p_1 < p_2 < 1$是已知的常数，求SPRT。

解：$f_1(x) = p_1^x(1-p_1)^{1-x}$，$f_2(x) = p_2^x(1-p_2)^{1-x}$，所以

$$\lambda_n = \prod_{i=1}^{n} \left(\frac{p_2}{p_1}\right)^{x_i} \left(\frac{1-p_2}{1-p_1}\right)^{1-x_i} = \left(\frac{p_2}{p_1}\right)^{S_n} \left(\frac{1-p_2}{1-p_1}\right)^{n-S_n}$$

其中$S_n = \sum_{i=1}^{n} x_i$。故而，

$$log\lambda_n = S_n log\frac{p_2}{p_1} + (n - S_n) log\left(\frac{1-p_2}{1-p_1}\right)$$

由$log\frac{p_2}{p_1} > 0$，$log\left(\frac{1-p_2}{1-p_1}\right) < 0$，易知

$$\lambda_n \geq B \quad \Leftrightarrow S_n \geq R_n =: cn + d_1$$

$$\lambda_n \leq A \quad \Leftrightarrow S_n \leq A_n =: cn + d_2$$

第二节 序贯概率比检验法

其中

$$c = -\frac{log(1-p_2) - log(1-p_1)}{logp_2 - logp_1 - log(1-p_2) + log(1-p_1)} \in (0,\ 1)$$

$$d_1 = \frac{logB}{logp_2 - logp_1 - log(1-p_2) + log(1-p_1)} > 0$$

$$d_2 = \frac{logA}{logp_2 - logp_1 - log(1-p_2) + log(1-p_1)} < 0$$

因此，$\tau^* = inf\{n: n \geq 1, S_n \notin (A_n, R_n)\}$。

图像显示（画图），中间是继续区域，左上方拒绝，右下方接受。

● 形式简单，实用。

第二节 序贯概率比检验法

例2. 正态分布情形。设$X \sim N(\theta, 1)$，

$$H_1: \theta = \theta_1 \quad \leftrightarrow H_2: \theta = \theta_2$$

其中$\theta_1 < \theta_2$是已知常数，求SPRT。

解：

$$\lambda_n = \prod_{i=1}^{n} \frac{f(x_i, \theta_2)}{f(x_i, \theta_1)} = \cdots = e^{(\theta_2 - \theta_1)S_n - \frac{n}{2}(\theta_2^2 - \theta_1^2)}$$

其中$S_n = \sum_{i=1}^{n} x_i$。因此，

$$log\lambda_n = (\theta_2 - \theta_1)S_n - \frac{n}{2}\left(\theta_2^2 - \theta_1^2\right)$$

令

$$A_n = \frac{\theta_1 + \theta_2}{2} \cdot n + \frac{logA}{\theta_2 - \theta_1}$$

$$R_n = \frac{\theta_1 + \theta_2}{2} \cdot n + \frac{logB}{\theta_2 - \theta_1}$$

第二节 序贯概率比检验法

则

$$\tau^* = inf\{n: n \geq 1, S_n \notin (A_n, R_n)\}$$

图像为（画图）……，与两点分布情形类似，但S_n不必取整数值。

一个序贯检验法将平面分成3个区域：接受、拒绝、继续区域，SPRT由两条平行线划分，其斜率与H_1、H_2有关，而其截距与A、B有关。

如何选取A、B？取决于第I、第II类错误。

记$\alpha = P_1(\tau^* < \infty, \lambda_{\tau^*} \geq B)$（当$H_1$成立时，停止并拒绝$H_1$，即以真为假），$\beta = P_2(\tau^* < \infty, \lambda_{\tau^*} \leq A)$（当$H_2$成立时，停止并接受$H_1$，即以假为真）。我们不加证明地给出下列结论。

定理2. $\alpha \leq \frac{1}{B}(1-\beta)$，$\beta \leq A(1-\alpha)$。

第二节 序贯概率比检验法

这是两个很好的不等式：无论f_1、f_2是什么分布，均可用A、B直接来控制α、β（其它部分约为1）。

实际应用中，可采用近似方案：给定α、β，取$A^* = \frac{\beta}{1-\alpha}$，$B^* = \frac{1-\beta}{\alpha}$，则一般地，$A^*$、$B^*$与目标$A$、$B$相差不大。对于序贯概率比检验法$S(A^*, B^*)$，记$\alpha^*$、$\beta^*$分别为其第I、第II类错误的概率，则有：

定理3. $\alpha^* \le \frac{\alpha}{1-\beta}$，$\beta^* \le \frac{\beta}{1-\alpha}$，$\alpha^* + \beta^* \le \alpha + \beta$。

证明：略（不难）。

● 采用近似方案，第I、第II类错误的概率与原定的α、β差距不大，且并不增大两类错误概率之和，早期很有意义。

第二节 序贯概率比检验法

定理4. 设$\widetilde{\Delta}=\left(\tilde{\tau},\,\widetilde{d}\right)$为任一序贯检验法，其两类错误的概率满足$\widetilde{\alpha}\le\alpha$，$\widetilde{\beta}\le\beta$，其中$\alpha$、$\beta$为$S(A,\,B)$的两类错误的概率，则有

$$E_i\tau^*\le E_i\tilde{\tau}\qquad\qquad i=1,\,2$$

● 说明其最优性：在控制两类错误概率的条件下，SPRT的平均样本量最小。

● 教材P286表2.1、表2.2表明，与固定样本量方法比较，许多情况下，大约可以节省50%的样本（μ_i为H_i成立时平均样本量/固定样本量）。

例3. 秘书问题：10人依次面试应聘，面试前i个人后，可知其排序，面试后需立即表示是否录用（仅1人），求使得找到最优者概率最大的策略。

例4. 设$X_1,\,\cdots,\,X_n,\cdots$独立同分布，$c>0$为每次观测费用，求$\tau$使得

$$E\left(max\{X_1,\,\cdots,\,X_\tau\}-c\tau\right)$$

最大。

第七章 统计决策与贝叶斯统计大意

统计决策与贝叶斯统计关系密切，考虑问题的角度、方法有一些类似之处。如都利用数据之外的信息。

可以独立发展，课程中也可独立讲授，但往往放在一起。

有贝叶斯决策理论。

第一节 统计决策问题概述

决策论研究当存在未知因素及随机性（不确定性）时，如何利用统计知识做出决策。

仍假设未知参数为$\boldsymbol{\theta} \in \boldsymbol{\Theta}$，数据仍为$X = (X_1, \cdots, X_n) \in \mathfrak{X}$。

任一可能的决策称为行动$a \in A$， A称为行动空间。

第一节 统计决策问题概述

统计决策（函数）：样本空间$\mathfrak{X}$到行动空间A的一个可测映射：$\delta = \delta(X_1, \cdots, X_n) \in A$，即当观测到数据$X_1, \cdots, X_n$时，采取行动$\delta \in A$。

例1. 在参数估计问题中，若用$\varphi(X_1, \cdots, X_n)$估计目标$\theta$，则$A = \Theta$，$\delta(X_1, \cdots, X_n) = \varphi(X_1, \cdots, X_n)$，所以估计问题包含于决策理论中。

例2. 在假设检验问题中， $A = \{a_0, a_1\} = \{0, 1\}$，其中$a_0 = 0$表示接受$H_0$，$a_1 = 1$表示否定$H_0$，所以检验问题包含于决策理论中（随机化检验情形，$A =$ $[0, 1]$）。

● 新理论应包含新内容。

损失函数$L(\theta, a)$：表示当参数真值为θ时，采取行动a造成的损失。一般要求$L(\theta, a) \geq -K > -\infty$，故不失一般性，常假设$L(\theta, a) \geq 0$。

第一节 统计决策问题概述

● 估计问题中，常用的损失函数有：

① 平方损失函数：$L_1(\theta,\ a)=(a-\theta)^2$；

② 绝对偏差损失函数：$L_2(\theta,\ a)=|a-\theta|$。

均表示估计越准确，损失越小。

● 检验问题中，常用的损失函数为：

$$L(\theta,\ a_0)=\begin{cases}0 & 当\theta\in\Theta_0\\ l_2 & 当\theta\notin\Theta_0\end{cases},\quad L(\theta,\ a_1)=\begin{cases}l_1 & 当\theta\in\Theta_0\\ 0 & 当\theta\notin\Theta_0\end{cases}$$

其中$l_1>0$，$l_2>0$为已知常数。即犯第一类错误的损失为l_1，犯第二类错误的损失为l_2，不犯错误的损失为0。

当$l_1=l_2=1$时，称为0-1损失。当$l_1\neq l_2$时，表示两类错误后果不同。

定义1. 称$R(\theta,\ \delta)=E_\theta L\big(\theta,\delta(X_1,\ \cdots,\ X_n)\big)$为决策$\delta$的风险函数（表示参数为$\theta$时，决策$\delta$的平均损失）。

第一节 统计决策问题概述

● 统计决策的最优解（最优决策）：如果决策δ^*使得对任意决策δ，有

$$R(\theta,\ \delta^*) \le R(\theta,\ \delta) \qquad \forall\theta \in \Theta$$

● 最优解要求太高，常常不存在。例如，估计问题中，若取$L_1(\theta,\ a) = (a-\theta)^2$，则最优解就是最小均方误差估计（一般不存在）。

定义2. 称决策δ是（可）容许的，若不存在另一决策δ'，使得$R(\theta,\ \delta') \le R(\theta,\ \delta)$（$\forall\theta \in \Theta$），且对某$\theta_0 \in \Theta$，不等式严格成立。

定义3. 称决策δ^*是minimax决策，若对一切决策δ，有

$$sup\{R(\theta,\ \delta^*), \theta \in \Theta\} \le sup\{R(\theta, \delta), \theta \in \Theta\}$$

● minimax准则：先对所有θ求风险函数的最大值，再对δ求最小。

● 因为minimax准则比较的是最糟糕的情形，必然是保守的（例如保研时，maxmini，看谁最低分的课程分数高，不是好标准）。

第一节 统计决策问题概述

参数空间Θ、样本空间$\mathfrak{X}$、行动空间A和损失函数L合称统计决策问题的四个要素。

例3.（教材P298，例1.1）检查设备零件，其可能状态为：θ_1（好）、θ_2（坏），可能采取的行动为：a_1（保留）、a_2（更换）或a_3（修理），损失函数为：

	a_1	a_2	a_3
θ_1	0	10	5
θ_2	12	1	6

样本为$X=0, 1$，分别表示零件温度发烫、正常，其分布为：

	$x=0$	$x=1$
θ_1	0.3	0.7
θ_2	0.6	0.4

问，采取何种决策使得损失尽量地小?

第一节 统计决策问题概述

此问题既非估计，又非检验，显示统计决策包含更一般的问题。

解法为穷举法（全离散，共$3^2=9$个可能决策）：

	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8	δ_9
$x=0$	a_1	a_1	a_1	a_2	a_2	a_2	a_3	a_3	a_3
$x=1$	a_1	a_2	a_3	a_1	a_2	a_3	a_1	a_2	a_3

由分布及损失函数，可以计算出风险函数

$$R(\theta,\delta)=E[L(\theta,\delta(X))]=\cdots\cdots$$

例如$R(\theta_1,\delta_2)=L(\theta_1,a_1)P(\delta_2(X)=a_1)+L(\theta_1,a_2)P(\delta_2(X)=a_2)+L(\theta_1,a_3)P(\delta_2(X)=a_3)=0\times 0.3+10\times 0.7+5\times 0=7$。所有结果列表如下：

第一节 统计决策问题概述

	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8	δ_9
$R(\theta_1,\delta)$	0	7	3.5	3	10	6.5	1.5	8.5	5
$R(\theta_2,\delta)$	12	7.6	9.6	5.4	1	3	8.4	4	6

按照定义2，δ_1、δ_4、δ_5、δ_6、δ_7是容许的，δ_2、δ_3、δ_8、δ_9是不容许的（如δ_2一致地不如δ_4）。

minimax决策为δ_4，其含义为：正常时保留，发烫时更换。

例4. 设$X\sim N(\theta,1)$，样本为一个观测值，$L(\theta,a)=(a-\theta)^2$，目的为估计参数θ。考察所有形如$\delta_c(X)=cX$的决策，则

第一节 统计决策问题概述

$$\begin{aligned} R(\theta,\delta_c) &= E_\theta L\big(\theta,\delta_c(X)\big) = E_\theta(\theta - cX)^2 \\ &= E_\theta[c(\theta - X) + (1-c)\theta]^2 \\ &= c^2 E_\theta(\theta - X)^2 + 2c(1-c)E_\theta(\theta - X) + (1-c)^2\theta^2 \\ &= c^2 + (1-c)^2\theta^2 \end{aligned}$$

当$c > 1$时，$R(\theta,\delta_c) = c^2 + (1-c)^2\theta^2 > 1 = R(\theta,\delta_1)$，因此$\delta_c$不是可容许的；

当$c < 0$时，$R(\theta,\delta_c) = c^2 + (1-c)^2\theta^2 > c^2 + (1+c)^2\theta^2 = R(\theta,\delta_{-c})$，因此$\delta_c$也不是可容许的；

当$0 \le c \le 1$时，不同的δ_c之间无法比较。可以证明，它们都是可容许的。

$R(\theta,\delta_1)$与$R(\theta,\delta_{0.5})$的图像为（……）

第一节 统计决策问题概述

因为

$$sup\{R(\theta,\delta_c),\theta\in\Theta\}=\begin{cases}1 & 当c=1\\ +\infty & 否则\end{cases}$$

故而（非严格证明），minimax决策为$\delta_1=X$。

● 从例4可以看出，可容许性是很弱的要求，仅排除一些最不合理的δ_c。而δ_1、$\delta_{0.5}$、甚至δ_0（即不管数据是什么，总用0估计θ）都是可容许的。

● 效用函数（Utility function）：其一般图像为……

● 在商业保险中的解释与作用……：X：不买保险时的收益（负值为损失），Y：买保险时的收益，

$$EX>EY$$

但

$$EU(X)<EU(Y)$$

第二节 什么是贝叶斯统计（简介）

● 以往比较不同统计方法时，常需比较参数θ的两个函数，难有“一致”性。是否可以比较两个函数的“平均”？

● 技术层面上，需要计算“加权（对参数θ）”平均，从而需要“权重”，即需要参数θ有“密度”。

Bayes学派（Bayesian）将未知参数θ看成随机变量（随机向量、随机元），从而有密度。

● 思想基础是Bayes公式（逆概率公式）。

● 称对立者为频率学派（Frequentist）。

● 发展较快，尤其近年来，因MCMC的发展。

● 见吴喜之，现代贝叶斯统计学（2000）。

第二节 什么是贝叶斯统计

● 一个例子（改自Savage (1961)）

① 某人说，他仅听几个小节，就能分出是海顿或莫扎特的作品，随机地播放了10段音乐（5:5），他全分对了；

② 某人说，她能够尝出咖啡中，是将牛奶倒入咖啡，还是将咖啡倒入牛奶，随机地倒了10杯（5:5），她全尝对了；

③ 某人说（喝醉后），他能预测你投掷硬币的正反面，随机地测10次（5:5），他全预测对了。

能否认为他（她）真具备这些能力?

严格按照传统统计学家观点，这是检验问题$H_0: p=\frac{1}{2}$，无论使用何种检验法，因为数据相同，结论也应相同。

常识告诉我们：①是真的；③是假的；②待定！我们使用了数据外的知识。

第二节 什么是贝叶斯统计

又如，中国队组队后，对某队的以往比赛成绩是3战皆胜。如何预测下一场的胜负？按频率学派，$\hat{p}=\bar{x}=1$，故预测下一场必胜。贝叶斯学派给出的点估计可以是$4/5=0.8$，而下一场中国队恰好输了。

再如，某型号火箭以往共发射3次，皆成功，按照统计量方法，成功率p的95%置信下限为$\sqrt[3]{1-0.95}\approx 0.368$，太低。而贝叶斯学派有“合理”方法提高。

还有对贝叶斯学派有利的例子，也有对其不利的例子。

- 涉及到哲学等，双方争论会长期存在（不同学科真理标准不同）。

- 实际应用中，各有利弊。

- 贝叶斯学派主要问题：能否/如何利用数据外知识。

第二节 什么是贝叶斯统计

● **基本问题：概率的定义。**

频率学派定义为，某随机事件在多次重复后的“稳定”值，之后有“极限”，适合于掷硬币、物理试验等，不适合于天气预报、石油勘探等；

柯尔莫哥洛夫（A. H. Kolmogorov）的定义为公理化体系，可列可加性……，未从语义上真正定义何为概率，大家都同意；

贝叶斯学派定义为，对事件可能性的主观信任度，尤其适合于难以重复的事件，如“北京明天下雨的概率”、“某处下面有石油的概率”等，但可能不可重复，可能“不客观”。

硬币正面$^1/_2$的概率，不再是硬币的属性，而是人大脑的主观产物。如果世界上有硬币，没有人，就没有概率。

第二节 什么是贝叶斯统计

对同一事件，不同的人可以有不同的主观（subjective）概率，如我掷硬币，对我，相关概率为1，对你……

对同一事件，同一个人也可以有不同的主观概率，随着你知识的累积、信息的更新……

博弈论的定义：你愿意付多高赌注，如巴西↔中国，你觉得1：4公平，则你对巴西赢的主观概率为0.8（$1\times0.8-4\times0.2=0$）。

所有概率都是条件概率，同一事件的概率不同是因为条件不同。

主观概率缺乏唯一性，一直是贝叶斯学派的一个大问题。

● 参数是未知的，故具有不确定性，因此就将其视为随机变量，故有（主观）概率分布。

其分布可以有多个，一般仅考虑两个：先验分布（得到数据前，prior）；后验分布（结合数据更新后，posterior）。------------------end 20240527

第二节 什么是贝叶斯统计

对同一事件，不同的人可以有不同的主观（subjective）概率，如我掷硬币，对我，相关概率为1，对你……

对同一事件，同一个人也可以有不同的主观概率，随着你知识的累积、信息的更新……

博弈论的定义：你愿意付多高赌注，如巴西↔中国，你觉得1：4公平，则你对巴西赢的主观概率为0.8（$1\times0.8-4\times0.2=0$）。

所有概率都是条件概率，同一事件的概率不同是因为条件不同。

主观概率缺乏唯一性，一直是贝叶斯学派的一个大问题。

● 参数是未知的，故具有不确定性，因此就将其视为随机变量，故有（主观）概率分布。

其分布可以有多个，一般仅考虑两个：先验分布（得到数据前，prior）；后验分布（结合数据更新后，posterior）。

第二节 什么是贝叶斯统计

例1. 已知$X \sim B(1, p)$，其中参数p未知，数据为$x_1 = x_2 = x_3 = 1$。因为p未知，但（主观）知道$0 \le p \le 1$，取其先验分布为$U[0, 1]$，即

$$\pi(p) = I_{[0, 1]}(p)$$

此时，似然函数为：

$$L(x_1, x_2, x_3|p) = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

它是（新解释）给定参数p时，数据x_1, x_2, x_3的条件分布密度。

因此，所有随机变量(x_1, x_2, x_3, p)的联合密度为：

$$f(x_1, x_2, x_3, p) = L(x_1, x_2, x_3|p) \cdot \pi(p)$$

故当给定数据时，p的条件分布密度为

$$\pi(p|x_1, \cdots, x_n) = \frac{L(x_1, \cdots, x_n|p) \cdot \pi(p)}{\int_0^1 L(x_1, \cdots, x_n|p) \cdot \pi(p)dp}$$

$$= \frac{p^{\sum x_i}(1-p)^{n-\sum x_i}}{\int_0^1 p^{\sum x_i}(1-p)^{n-\sum x_i}dp} = \frac{p^{\sum x_i}(1-p)^{n-\sum x_i}}{B_e(\sum x_i + 1, n - \sum x_i + 1)}$$

$$= \frac{\Gamma(n+2)}{\Gamma(\sum x_i + 1)\Gamma(n - \sum x_i + 1)} p^{\sum x_i}(1-p)^{n-\sum x_i}$$

第二节 什么是贝叶斯统计

其中B_e为Beta函数。

称$\pi(p|x_1, \cdots, x_n)$为参数p的后验分布密度。

贝叶斯统计的所有推断均来自后验分布。

对例1的数据，$\pi(p|x_1, x_2, x_3) = \frac{\Gamma(5)}{\Gamma(4)\Gamma(1)} p^3(1-p)^{3-3} = 4p^3$，其后验均值为

$$E_{posterior}p = \int_0^1 p \cdot 4p^3 dp = \cdots = {}^4/_5$$

这是对p的（一个）点估计，注意频率学派为1。

第二节 什么是贝叶斯统计

- 若数据为恰三局两胜，则$\sum x_i = 2$，

$$\pi(p|x_1, x_2, x_3) = \cdots = 12p^2(1-p)$$

$$E_{posterior}p = \int_0^1 12p^3(1-p)dp = 12B_e(4, 2) = 12 \cdot \frac{3!}{5!} = \frac{3}{5}$$

- 若背景为投掷一枚（不均匀）硬币，则随着投掷次数n增加，后验分布密度的图像会越来越高、窄，集中于“真值”。

数据x与参数θ的联合密度有两种表达式：

$$f(x, \theta) = L(x|\theta)\pi(\theta) = \pi(\theta|x)h(x)$$

其中$\pi(\theta)$、$h(x)$分别为θ与x的边缘密度，$L(x|\theta)$、$\pi(\theta|x)$分别为条件密度。

作为θ的函数，后验分布密度$\pi(\theta|x)$与先验密度$\pi(\theta)$和似然函数$L(x|\theta)$的乘积成正比，仅当需要求归一化因子时，才考虑$h(x)$。

第二节 什么是贝叶斯统计

对θ的一切推断都通过后验分布$\pi(\theta|x)$做出，如后验均值、后验方差、后验中位数、MAP（Mode）等。

在较一般的条件下，有

$$\pi(\theta|x)=\frac{L(x|\theta)\pi(\theta)}{\int_{\Theta} L(x|\theta)\pi(\theta)d\theta}$$

例2. 设$X\sim N(\theta,1)$，数据为$x_1,\cdots,x_n$，若取参数θ的先验分布为$\theta\sim N(0,\tau_0^2)$，求其后验分布。

解:

$$L(x|\theta)\pi(\theta)=\left(\frac{1}{\sqrt{2\pi}}\right)^{n+1}\frac{1}{\tau_0}exp\left\{-\frac{\theta^2}{2\tau_0^2}-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}$$

$$\propto exp\left\{-\frac{1}{2}\left(n+\frac{1}{\tau_0^2}\right)\left(\theta-\frac{\sum_{i=1}^{n}x_i}{n+1/\tau_0^2}\right)^2\right\}$$

第二节 什么是贝叶斯统计

故参数θ的后验分布为$N\left(\frac{\sum_{i=1}^{n} x_i}{n+1/\tau_0^2}, \frac{1}{n+1/\tau_0^2}\right)$。

● 此例可一般化。先验分布方差τ_0^2大、小分别说明……

● 随着n越来越大，后验方差趋于0，后验均值几乎就是$\bar{x}$，先验分布的作用越来越小。

频率学派有置信区间（不易解释），贝叶斯学派有Bayes置信区间（Bayesian credible interval），方法及解释很直观：

当$\pi(\theta|x)$形状较好（单峰）时，取区间$[\theta_L, \theta_U]$，满足

$$\int_{\theta_L}^{\theta_U} \pi(\theta|x)d\theta = 1 - \alpha$$

（且$\pi(\theta|x)$的值在区间内更大）。

第二节 什么是贝叶斯统计

实用中为方便起见，常取θ_L、θ_U满足

$$\int_{-\infty}^{\theta_L} \pi(\theta|x)d\theta = \int_{\theta_U}^{\infty} \pi(\theta|x)d\theta = \frac{\alpha}{2}$$

的简单方法。

对例2，参数θ的95%的贝叶斯置信区间为：

$$\left[\frac{\sum_{i=1}^{n} x_i}{n+1/\tau_0^2} - \frac{1.96}{\sqrt{n+1/\tau_0^2}}, \frac{\sum_{i=1}^{n} x_i}{n+1/\tau_0^2} + \frac{1.96}{\sqrt{n+1/\tau_0^2}}\right]$$

● n越大，区间越窄，下降速度仍为$1/\sqrt{n}$。对例1类似。

● 贝叶斯学派的一大优势是对置信区间简单易懂的解释：参数θ是随机变量，以$1-\alpha$的概率落在$[\theta_L, \theta_U]$内。而频率学派对置信区间的解释较为曲折。

第二节 什么是贝叶斯统计

● Bayesian如何做预测（如对例2中x_{n+1}的预测）？若参数θ已知，则直接利用x_{n+1}的分布$f(x_{n+1}|\theta)$。现θ的后验分布为$\pi(\theta|x)$，综合θ的所有情况，得到x_{n+1}的预测（预报）分布（predictive distribution）为

$$f(x_{n+1}|x_1,\cdots,x_n)=\int_\Theta f(x_{n+1}|\theta)\cdot\pi(\theta|x_1,\cdots,x_n)d\theta$$

如对例1，

$$P(x_4=1|x_1,x_2,x_3)=\int_0^1 p\cdot 4p^3\,dp=\frac{4}{5}$$

恰为p的贝叶斯点估计；对例2，经过较繁琐的推导，可以得到，

$$x_{n+1}|x_1,\cdots,x_n\sim N\left(\frac{\sum_{i=1}^n x_i}{n+1/\tau_0^2},\ \frac{n+1+1/\tau_0^2}{n+1/\tau_0^2}\right)$$

预报分布是已知$x_1,\cdots,x_n$时，x_{n+1}的条件密度，恰符合由$x_1,\cdots,x_n$预报x_{n+1}的直观解释。

第二节 什么是贝叶斯统计

● 贝叶斯学派对频率学派的又一优势：按频率学派理论$x_1, \cdots, x_n, x_{n+1}$相互独立同分布，故$f(x_{n+1}|x_1, \cdots, x_n) = f(x_{n+1})$，即不能利用$x_1, \cdots, x_n$预测$x_{n+1}$！

贝叶斯学派的解释为：当给定参数θ时，$x_1, \cdots, x_n, x_{n+1}$相互独立（条件独立性）。

由于θ是随机变量，故当需要比较θ的两个函数时，可以通过求期望（加权平均）化为对两个数的比较。故在决策论中有：

定义1. 设θ的先验分布为$\pi(\theta)$，δ是一个（统计）决策，记

$$\rho(\delta) = E_{\pi}R(\theta, \delta) = \int_{\Theta} R(\theta, \delta)\,\pi(\theta)d\theta$$

称为δ的平均风险。

● 每个决策对应于唯一的一个数值，故不同的决策可比较。

第二节 什么是贝叶斯统计

定义2. 若决策δ^*使得平均风险达到最小，即

$$\rho(\delta^*) = inf\{\rho(\delta), \forall\delta\}$$

则称δ^*为贝叶斯决策。

贝叶斯决策的思想：损失函数为$L(\theta, \delta(X))$，先对X求平均，得风险函数$R(\theta, \delta)$，再对θ求平均，得平均风险$\rho(\delta)$。δ^*是综合所有随机因素后，平均风险（损失）最小的决策。

对于本节例1、例2，若取损失函数为平方损失，则贝叶斯决策（贝叶斯估计）恰为其后验均值。

例3（第一节例3）如果取先验分布为$\pi(\theta_1) = 0.7$，$\pi(\theta_2) = 0.3$，则平均风险可计算，结果列表如下：

	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8	δ_9
$\rho(\delta)$	3.6	7.18	5.33	3.72	7.3	5.45	3.57	7.15	5.3

故而， δ_7是贝叶斯决策（发烫则修理，不发烫则保留）。

第二节 什么是贝叶斯统计

一般情况下，可能的决策太多，无法穷举，如何求贝叶斯决策?

$\rho(\delta)$是通过两次积分得到的，故交换积分次序得:

$$\rho(\delta)=\int_{\Theta}\left[\int_{\mathcal{X}}L\big(\theta,\,\delta(x)\big)p(x|\theta)dx\right]\cdot\pi(\theta)\,d\theta$$

$$=\int_{\mathcal{X}}\left[\int_{\Theta}L\big(\theta,\,\delta(x)\big)\pi(\theta|x)d\theta\right]\cdot h(x)dx$$

故而，当得到数据x后，只需在行动空间A中找a_x，使得

$$\int_{\Theta}L\big(\theta,\,\delta(x)\big)\pi(\theta|x)d\theta$$

达到最小即可（不再考虑对x 的积分），即$\delta^*(x)=a_x$。

● 贝叶斯学派的又一观点：决策只与后验分布有关，而后验分布$\pi(\theta|x)$只与观测到的数据x有关！即不必考虑x未取到的可能数据的情形。

第二节 什么是贝叶斯统计

例4. 设$X \sim N(\mu,\ 1)$，$\mu \in \Theta = \{0,\ 2\}$，$H_0$: $\mu = 0 \leftrightarrow H_1$: $\mu = 2$，损失函数为：

$$L(\mu,\ a_0) = \begin{cases} 0 & \text{当}\mu = 0 \\ 1 & \text{否则} \end{cases},\quad L(\mu,\ a_1) = \begin{cases} 1 & \text{当}\mu = 0 \\ 0 & \text{否则} \end{cases}$$

其中$l_1 = l_2 = 1$为已知常数（0-1损失），若μ的先验分布为$\pi(0) = 1 - \pi(2) = p_0 = 0.5$，数据为简单随机样本$x_1, \cdots, x_n$，求问题的贝叶斯检验（贝叶斯决策），及两类错误的概率。

解：由前面知识，

$$L(x|\mu) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2} \propto e^{-\frac{n}{2}(\bar{x} - \mu)^2}$$

所以，

$$\pi(0|x) = 0.5C_0 e^{-\frac{n}{2}\bar{x}^2}$$
$$\pi(2|x) = 0.5C_0 e^{-\frac{n}{2}(\bar{x} - 2)^2}$$

其中C_0为某归一化常数。因此，

第二节 什么是贝叶斯统计

$$\rho(\delta) = \int_{\Theta} L\big(\mu,\ \delta(x)\big)\pi(\mu|x)d\mu$$
$$= 0.5C_0 e^{-\frac{n}{2}\overline{x}^2} L\big(0,\ \delta(x)\big) + 0.5C_0 e^{-\frac{n}{2}(\overline{x}-2)^2} L\big(2,\ \delta(x)\big)$$
$$\propto I\{\text{否定}H_0\} + e^{2n\overline{x}-2n} I\{\text{不否定}H_0\}$$

故贝叶斯解为：

$$\varphi(x) = \begin{cases} 1 & \text{当}\overline{x} > 1 \\ 0 & \text{否则} \end{cases}$$

两类错误的概率为：

$$\alpha = P(\overline{x} > 1|\mu = 0) = \cdots = 1 - \Phi\big(\sqrt{n}\big)$$
$$\beta = P(\overline{x} \le 1|\mu = 2) = \cdots = 1 - \Phi\big(\sqrt{n}\big)$$

● **当$p_0 \neq 0.5$或两类错误的损失不同时，检验法不必以1为临界值，两类错误的概率亦可不同。**

第三节 关于先验分布

首先，一些人不接受“主观概率”的定义，因此也不接受“先验分布”、“后验分布”及其它贝叶斯统计的概念和方法。此争议会长期存在。概率论中，选山羊或选汽车的例子就用到了先验分布。

其次，贝叶斯统计（以往）存在计算上的困难：后验分布正比于似然函数与先验分布的乘积，但归一化因子、后验均值、后验方差等的计算均归结于（高维）积分。近年来快速发展并得到广泛应用的马尔科夫链随机模拟（Markov chain Monte Carlo，简记为MCMC）方法可较好地解决此问题。当然，仍可能有计算速度、收敛性等问题。

再次，在同意使用贝叶斯统计方法的前提下，如何构造、确定先验分布，是最易引起争议的问题！先验知识往往较模糊，不同的人先验知识也不同。

下面几条非并列关系。

第三节 关于先验分布

一、经验贝叶斯方法

由数据（历史数据或现时数据）构造先验分布。例如，先选定先验分布的类型，再根据数据得到（估计）其（超）参数。可以看作贝叶斯学派与频率学派的结合。优点是相对客观；不足是当无历史数据时，现时数据（可能）两次被使用。

二、“同等无知”

若参数θ仅取有限个点，则先验分布取为离散型均匀分布；若参数$\theta \in [a, b]$，则取先验分布为$U[a, b]$；若参数$\theta \in (-\infty, \infty)$，取$\pi(\theta) = c > 0$（广义先验分布？Improper prior），后验分布往往是proper distribution，例如第二节中的例2，后验分布为$N(\bar{x}, \frac{1}{n})$。

问题：此先验分布对参数变换不具不变性，例如，参数θ是圆盘半径，若设其服从均匀分布，则圆盘面积$\pi\theta^2$（亦可设为参数）不服从均匀分布。

第三节 关于先验分布

三、无信息先验分布等

无信息先验分布（Non-informative prior）

非主观先验分布（Non-subjective prior）

即希望先验分布不包含关于参数θ的任何信息。其想法容易理解、接受，但具体实施有不少问题。

Jeffreys （1961）提出使用θ的Fisher信息量的平方根来定义其先验分布，（当θ是多维参数时，可使用其Fisher信息矩阵行列式的平方根），并“论证”了它是无信息的，称为Jeffreys原则。

对于多种多样的统计模型，无信息先验分布中的“无信息”界定较复杂，且其结果往往也不唯一。但一般地，各种无信息先验分布确实对后验分布的影响很小。

第三节 关于先验分布

四、共轭分布族

如果先验分布$\pi(\theta)$的类型使得后验分布$\pi(\theta|x)$仍为此类型，则称先验分布$\pi(\theta)$是密度函数$f(x|\theta)$的共轭分布。

优点：推导、计算非常方便。不足：仍有未知超参数；有的共轭分布不足以描述各种先验知识。

定理1. 设$X_1, \cdots, X_n$是来自两点分布$B(1, p)$的简单随机样本，若取参数p的先验分布为$Be(\alpha, \beta)$，其中$\alpha > 0$、$\beta > 0$为超参数，则p的后验分布为$Be\left(\alpha + \sum_{i=1}^{n} X_i, \beta + n - \sum_{i=1}^{n} X_i\right)$。

证明：略（第二节例1的推广。由此，贝塔分布是两点分布的共轭分布）。

$Be(\alpha, \beta)$的图像可为……，$Be(1, 1)$即为$U[0, 1]$。

第三节 关于先验分布

定理2. 设$X_1, \cdots, X_n$是来自参数为λ的Poisson分布的简单随机样本，若取λ的先验分布为$\Gamma(\alpha, \beta)$，其中$\alpha > 0$、$\beta > 0$为超参数，则λ的后验分布为$\Gamma\left(\alpha + \sum_{i=1}^{n} X_i\,, \beta + n\right)$。

证明：

$$\pi(\lambda) = C\lambda^{\alpha-1}e^{-\beta\lambda}$$

$$L(X_1, \cdots, X_n|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{X_i}}{X_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} X_i}}{X_1! \cdots X_n!} e^{-n\lambda}$$

故

$$\pi(\lambda|X_1, \cdots, X_n) \propto \pi(\lambda)L(X_1, \cdots, X_n|\lambda) \propto C'\lambda^{\alpha+\sum_{i=1}^{n} X_i - 1} e^{-(\beta+n)\lambda}$$

证毕。（由此，伽马分布是Poisson分布的共轭分布）。

定理3. 设$X_1, \cdots, X_n$是来自指数分布$\lambda e^{-\lambda x}$的简单随机样本，若取参数λ的先验分布为$\Gamma(\alpha, \beta)$，其中$\alpha > 0$、$\beta > 0$为超参数，则λ的后验分布为$\Gamma\left(\alpha + n, \beta + \sum_{i=1}^{n} X_i\right)$。

第三节 关于先验分布

证明：

$$\pi(\lambda) = C\lambda^{\alpha-1}e^{-\beta\lambda}$$

$$L(X_1, \cdots, X_n|\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} X_i}$$

故

$$\pi(\lambda|X_1, \cdots, X_n) \propto \pi(\lambda)L(X_1, \cdots, X_n|\lambda) \propto C'\lambda^{\alpha+n-1}e^{-(\beta+\sum_{i=1}^{n} X_i)\lambda}$$

证毕。（由此，伽马分布是指数分布的共轭分布）

定理4. 设$X_1, \cdots, X_n$是来自正态分布$N(\mu, \sigma_0^2)$的简单随机样本，其中σ_0^2已知，若取参数μ的先验分布为$N(\mu_0, \tau_0^2)$，则μ的后验分布为$N(\mu^*, (\sigma^*)^2)$，其中

$$\mu^* = \frac{\mu_0\sigma_0^2 + n\tau_0^2\overline{X}}{\sigma_0^2 + n\tau_0^2}$$

$$(\sigma^*)^2 = \frac{\sigma_0^2\tau_0^2}{\sigma_0^2 + n\tau_0^2}$$

证明：略（第二节例2的推广）

第三节 关于先验分布

定义1. 称随机变量Y服从参数为α、β的逆（倒）Γ分布，记为$Y \sim IG(\alpha,\ \beta)$，若

$$1/Y \sim \Gamma(\alpha,\ \beta) \qquad (\alpha > 0, \beta > 0)\text{。}$$

定理5. 设$X_1, \cdots, X_n$是来自正态分布$N(\mu_0,\ \sigma^2)$的简单随机样本，其中μ_0已知，若取参数σ^2的先验分布为$IG(\alpha,\ \beta)$，则σ^2的后验分布为$IG\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu_0)^2\right)$。

证明：利用概率论知识，易知$IG(\alpha,\ \beta)$的密度函数为

$$g(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{-\alpha-1} e^{-\frac{\beta}{t}} \qquad (t > 0)$$

记$T = \sigma^2$，则其先验分布密度为

$$\pi(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{-\alpha-1} e^{-\frac{\beta}{t}}$$

第三节 关于先验分布

似然函数为

$$L(X_1,\ \cdots,\ X_n|t) = Ct^{-\frac{n}{2}}e^{-\frac{1}{2t}\sum_{i=1}^{n}(X_i-\mu_0)^2},$$

其中C（及后面C'）为常数。故而，

$$\pi(t|X_1,\ \cdots,\ X_n) \propto \pi(t)L(X_1,\ \cdots,\ X_n|t) \propto C't^{-\left(\alpha+\frac{n}{2}\right)-1}e^{-\frac{1}{t}[\beta+\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu_0)^2]}$$

证毕。

● 在利用贝叶斯方法处理μ、σ^2均未知的正态模型时，我们引入下面的定义。

定义2. 称二维随机向量(M,R)服从参数为$(\mu,\ \tau,\ \alpha,\ \beta)$的正态- Γ先验分布，若R的边缘分布是参数为$\alpha,\ \beta$的Γ分布，而在$R=r$下M的条件分布是$N(\mu,\ \frac{1}{\tau r})$。这时$(M,R)$的分布密度为

$$\varphi(m,\ r;\mu,\ \tau,\ \alpha,\ \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\sqrt{\tau}r^{\alpha-\frac{1}{2}}exp\left\{-\frac{\tau r}{2}(m-\mu)^2-\beta r\right\}$$

第三节 关于先验分布

定理6. 设$X_1, \cdots, X_n$是来自正态分布$N(M, \frac{1}{R})$的简单随机样本， M、 R未知。设(M, R)的先验分布是参数为$(\mu, \tau, \alpha, \beta)$的正态- Γ分布，则在$X_1 = x_1$， $\cdots$ ，$X_n = x_n$下， (M, R)的后验分布是参数为$(\mu^*, \tau^*, \alpha^*, \beta^*)$的正态- Γ分布，其中

$$\mu^* = \frac{\tau\mu + n\overline{x}}{\tau + n}, \quad \tau^* = \tau + n, \quad \alpha^* = \alpha + \frac{n}{2},$$

$$\beta^* = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{\tau n(\overline{x} - \mu)^2}{2(n-1)} \quad \left(\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

证明：$X_1, \cdots, X_n$的联合密度为

$$f_n(x_1, \cdots, x_n) = \left(\frac{r}{2\pi}\right)^{\frac{n}{2}} exp\left\{-\frac{r}{2}\sum_{i=1}^{n}(x_i - m)^2\right\}$$

(M, R)的先验分布密度由定义2所示，因此知(M, R)的后验分布密度

$$\pi(m, r|x_1, \cdots, x_n) \propto r^{\alpha+\frac{n-1}{2}} \exp\left\{-\frac{r}{2}\left[\sum_{i=1}^{n}(x_i - m)^2 + \tau(m-\mu)^2\right] - \beta r\right\}$$

第三节 关于先验分布

由于

$$\sum_{i=1}^{n}(x_i-m)^2+\tau(m-\mu)^2=\sum_{i=1}^{n}(x_i-\overline{x})^2+n(\overline{x}-m)^2+\tau(m-\mu)^2$$

$$=\sum_{i=1}^{n}(x_i-\overline{x})^2+(\tau+n)\left(m-\frac{n\overline{x}+\tau\mu}{\tau+n}\right)^2+\frac{n\tau(\overline{x}-\mu)^2}{\tau+n}$$

因此

$$\pi(m,r|x_1,\cdots,x_n)\propto r^{\alpha+\frac{n-1}{2}}\exp\left\{-\frac{\tau+n}{2}r(m-\mu^*)^2-\beta r\right\}$$

$$=\varphi(m,r;\mu^*,\tau^*,\alpha^*,\beta^*),$$

即后验分布是参数为（μ^*、τ^*、α^*、β^*）的正态-Γ分布，证毕。

---------------------- end 20240530

定义3. 称随机变量X服从参数为$x_0>0$，$\alpha>0$的Pareto分布，若其密度函数为

$$h(x)=\begin{cases}\alpha x_0^{\alpha}x^{-\alpha-1} & 当x\ge x_0\\ 0 & 否则\end{cases}$$

第三节 关于先验分布

由于

$$\sum_{i=1}^{n}(x_i - m)^2 + \tau(m-\mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - m)^2 + \tau(m-\mu)^2$$

$$= \sum_{i=1}^{n}(x_i - \bar{x})^2 + (\tau + n)\left(m - \frac{n\bar{x} + \tau\mu}{\tau + n}\right)^2 + \frac{n\tau(\bar{x}-\mu)^2}{\tau + n}$$

因此

$$\pi(m, r|x_1, \cdots, x_n) \propto r^{\alpha + \frac{n-1}{2}} \exp\left\{-\frac{\tau + n}{2} r(m - \mu^*)^2 - \beta r\right\}$$

$$= \varphi(m, r; \mu^*, \tau^*, \alpha^*, \beta^*),$$

即后验分布是参数为（μ^*、τ^*、α^*、β^*）的正态-Γ分布，证毕。

定义3. 称随机变量X服从参数为$x_0 > 0$，$\alpha > 0$的Pareto分布，若其密度函数为

$$h(x) = \begin{cases} \alpha x_0^{\alpha} x^{-\alpha-1} & \text{当} x \geq x_0 \\ 0 & \text{否则} \end{cases}$$

第三节 关于先验分布

定理6. 设$X_1, \cdots, X_n$是来自均匀分布$U[0, \theta]$的简单随机样本，若取参数θ的先验分布为超参数θ_0，α的Pareto分布，则θ的后验分布是超参数为θ_0'，$\alpha + n$的Pareto分布，其中$\theta_0' = max\{\theta_0, X_1, \cdots, X_n\}$。

● 超参数α的大小表示对θ靠近θ_0（右侧）主观信念的强弱：越大越强，越小先验密度越平。

证明：θ的先验密度为

$$\pi(\theta) = \frac{\alpha\theta_0^\alpha}{\theta^{\alpha+1}} I_{[\theta_0, \infty)}(\theta)$$

似然函数为

$$L(X_1, \cdots, X_n|\theta) = \frac{1}{\theta^n}\prod_{i=1}^{n} I_{[0, \theta]}(X_i) = \frac{1}{\theta^n} I_{[0, \theta]}(maxX_i)$$

$$= \frac{1}{\theta^n} I_{[maxX_i, \infty)}(\theta)$$

第三节 关于先验分布

故而，

$$\pi(\theta|X_1, \cdots, X_n) \propto \pi(\theta)L(X_1, \cdots, X_n|\theta) \propto \frac{1}{\theta^{\alpha+n+1}} I_{[\theta_0', \infty)}(\theta)$$

证毕。

定义4. 称(Y_1, Y_2)服从参数为(r_1, r_2, α)的二维Pareto分布($r_1 < r_2$，且$\alpha > 0$)，若其分布密度函数是

$$g(y_1, y_2; r_1, r_2, \alpha) = \frac{\alpha(\alpha+1)(r_2 - r_1)^{\alpha}}{(y_2 - y_1)^{\alpha+2}} I_{(-\infty, r_1]}(y_1) I_{[r_2, \infty)}(y_2).$$

定理3.8 设$X_1, \cdots, X_n$是来自$[\theta_1, \theta_2]$上均匀分布的样本，$\theta_1 < \theta_2$未知。设(θ_1, θ_2)的先验分布是参数为r_1, r_2, α的二维Pareto分布，则在$X_1 = x_1, \cdots, X_n = x_n$下，$(\theta_1, \theta_2)$的后验分布是参数为$r_1^*, r_2^*, \alpha^*$的二维Pareto分布，其中$r_1^* = \min\{r_1, X_1, \cdots, X_n\}$，$r_2^* = \max\{r_2, X_1, \cdots, X_n\}$，$\alpha^* = \alpha + n$。

第三节 关于先验分布

证明： $X_1, \cdots, X_n$**的联合密度为**

$$f(x_1, \cdots, x_n; \theta_1, \theta_2) = \prod_{i=1}^{n} \frac{1}{\theta_2 - \theta_1} I_{[\theta_1, \theta_2]}(x_i)$$

$$= \frac{1}{(\theta_2 - \theta_1)^n} I_{[\theta_1, \infty)}(\min x_i) \cdot I_{(-\infty, \theta_2]}(\max x_i)$$

(θ_1, θ_2)的先验密度由定义4给出，因此(θ_1, θ_2)的后验密度为

$$\xi(\theta_1, \theta_2 | x_1, \cdots, x_n)$$

$$= \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{(\theta_2 - \theta_1)^{\alpha+n+2}} I_{(-\infty, r_1]}(\theta_1) I_{[r_2, \infty)}(\theta_2) I_{[\theta_1, \infty)}\left(\min x_i\right) I_{(-\infty, \theta_2]}\left(\max x_i\right)$$

$$= \frac{C}{(\theta_2 - \theta_1)^{\alpha+n+2}} I_{(-\infty, r_1^*]}(\theta_1) I_{[r_2^*, \infty)}(\theta_2) = g(\theta_1, \theta_2, r_1^*, r_2^*, \alpha^*),$$

因此后验分布是参数为(r_1^*, r_2^*, α^*)的Pareto**分布，证毕。**

教材中给出了较一般情形下利用充分统计量寻找共轭分布的方法，有兴趣的可查阅。

第三节 关于先验分布

四、层次贝叶斯学派（Hierarchical Bayesian）

前面已有涉及：设参数θ服从某种带有超参数α的分布，α待定，而超参数α又服从某个或某种分布（若某种，还可分层）。

例如，$X_1, \cdots, X_n|p\sim B(1, p)$，$p|\alpha, \beta\sim Be(\alpha, \beta)$，而$\alpha\sim Exp(\lambda_1)$，$\beta\sim Exp(\lambda_2)$。其图示为……

优点：先验分布不再是一个分布，而是一（大）类分布。利于描述多种先验知识；先验分布更“平”。适用于MCMC计算。

● 还有最大熵先验分布等，主要思想是构造尽可能不包含任何信息的先验分布。

第三节 关于先验分布

五、关于“主观”的讨论

自然科学的一般标准要求研究者应该是客观的。

一般初学者容易认为，贝叶斯学派主观，频率学派客观。

一些贝叶斯学派的研究者极力构造“客观”的先验分布。

统计分析中，主观性难以避免：模型的选择不同、优良性标准的不同、置信度、检验水平、分段数目、核函数的选择不同、假设的不同，到本章参数空间、损失函数等的定义不同，都不可避免地带有主观性。

较激烈的Bayesian说：他们将主观摆在桌面上，而频率学派将主观藏在桌面下，只拿出客观的东西来标榜自己。

第三节 关于先验分布

当数据量较大时，只要模型一样（似然函数相同），则频率学派和贝叶斯学派的结果很近似（后验=先验+似然）。

大多数情况下，应取较“平”的先验分布，并/或说明，先验分布的选取对分析结果影响不大。

如果先验分布的选取确实较显著地影响后验分布，则必须对先验分布选取的理由给出解释（如相关领域专家坚持）。

第四节 马氏链随机模拟

一、目的

近几十年统计学最主要的进展之一，极大促进了贝叶斯统计的发展，和各种复杂模型的应用。

贝叶斯统计的主要推断思路：后验分布正比于先验分布与似然函数的乘积。但求后验分布需要求积分，求后验均值等也同样。

当参数空间维数较高、或乘积函数形式较复杂时，积分可能无法进行。

马尔科夫链随机模拟（Markov chain Monte Carlo）可以解决此问题。需要大量的计算。

基本思路：利用随机模拟方法，通过分步抽样，构造一个适当的马氏链，得到近似的、相依的后验分布模拟样本，并通过此样本计算后验的特征，如均值、分位数等。

第四节 马氏链随机模拟

分步的作用在于将复杂问题简单化。

记$\pi(\theta)=\pi(\theta_1,\cdots,\theta_n)$为先验分布与似然函数的乘积（数据已给定，不再显式表示，同时n不再表示样本量，而表示参数个数），$\theta=(\theta_1,\cdots,\theta_n)\in\Theta\subset R^n$为模型参数，模拟出的马氏链为$\left(\theta^{(1)},\cdots,\theta^{(m)},\cdots\right)$。

在一定条件下（马氏链是非周期的，不可约的），由遍历性有

$$\frac{1}{m}\sum_{k=1}^{m}f\left(\theta^{(k)}\right)\rightarrow E_{posterior}[f(\theta)]$$

其中$f(\theta)$为参数θ的n元实值函数（数学要求不高），故可以适当选择来计算如θ_1的均值、方差等。还有，当m较大时，$\theta^{(m)}$可近似地看成后验分布的一个样本。独立重复可得多个样本。

二、Gibbs sampler

第四节 马氏链随机模拟

记$\pi(\theta_i|\theta_{-i})$为给定$\theta$的所有其他分量$\theta_{-i}=(\theta_1,\cdots,\theta_{i-1},\theta_{i+1},\cdots,\theta_n)$时，$\theta_i$的条件分布密度。一般$\pi(\theta)$的形式可能很复杂（多元函数），但$\pi(\theta_i|\theta_{-i})$往往较简单（一元函数），可由$\pi(\theta)$的形式得到（或正比于）。

一般算法如下：

① 产生起始点$\theta^{(0)}=\left(\theta_1^{(0)},\cdots,\theta_n^{(0)}\right)$（如从$\theta$的先验分布模拟抽样）；

② 从$\pi\left(\theta_1|\theta_2^{(0)},\theta_3^{(0)},\cdots,\theta_n^{(0)}\right)$中模拟产生$\theta_1^{(1)}$；

③ 从$\pi\left(\theta_2|\theta_1^{(1)},\theta_3^{(0)},\cdots,\theta_n^{(0)}\right)$中模拟产生$\theta_2^{(1)}$；

④ 从$\pi\left(\theta_3|\theta_1^{(1)},\theta_2^{(1)},\theta_4^{(0)},\cdots,\theta_n^{(0)}\right)$中模拟产生$\theta_3^{(1)}$；

⑤ ……

⑥ 从$\pi\left(\theta_n|\theta_1^{(1)},\cdots,\theta_{n-1}^{(1)}\right)$中模拟产生$\theta_n^{(1)}$。

第四节 马氏链随机模拟

如此完成从$\theta^{(0)}$到$\theta^{(1)}$的更新。反复重复以上②至⑥，得到马氏链$\left(\boldsymbol{\theta}^{(1)}, \cdots, \boldsymbol{\theta}^{(m)}, \cdots\right)$。

此过程之所以是马氏的，是因为任何一个参数的更新，只依赖于其他参数的当前值。任一个参数一旦被更新，其原值就（可以被存储）不再影响程序运行。

此马氏链每一步的转移概率为：

$$K\left(\theta^{(m)}, \theta^{(m+1)}\right) = \prod_{i=1}^{n} \pi\left(\theta_i^{(m+1)} | \theta_1^{(m+1)}, \cdots, \theta_{i-1}^{(m+1)}, \theta_{i+1}^{(m)}, \cdots, \theta_n^{(m)}\right)$$

θ的各分量的次序可以充分考虑计算方便。

不必一次仅更新一个分量（有时，同时更新多个更快，或更方便）。

第四节 马氏链随机模拟

三、Metropolis-Hastings算法

当Gibbs sampler的某些步骤难以直接实现时（即无法直接产生某种分布的随机数），可采用Metropolis-Hastings算法。一般地叙述如下：

设$q(\theta, \theta')$为（任意的）转移概率函数，希望从$\theta^{(m)}$转移到$\theta^{(m+1)}$（可能仅变化一个分量），先由$q\left(\theta^{(m)}, \theta'\right)$产生一个候选$\theta'$，之后计算Hastings ratio $r(\theta^{(m)}, \theta')$，再后以概率$r(\theta^{(m)}, \theta')$接受$\theta^{(m+1)} = \theta'$，以概率$1 - r(\theta^{(m)}, \theta')$停留在$\theta^{(m+1)} = \theta^{(m)}$。

$$r\left(\theta^{(m)}, \theta'\right) = \begin{cases} \min\left\{\dfrac{\pi(\theta')q(\theta', \theta^{(m)})}{\pi\left(\theta^{(m)}\right)q(\theta^{(m)}, \theta')}, 1\right\} & \text{若}\pi\left(\theta^{(m)}\right)q\left(\theta^{(m)}, \theta'\right) > 0 \\ 1 & \text{若}\pi\left(\theta^{(m)}\right)q\left(\theta^{(m)}, \theta'\right) = 0 \end{cases}$$

如此产生的马氏链的转移概率为：

第四节 马氏链随机模拟

$$p\left(\theta^{(m)},\ \theta'\right)=\begin{cases}q\left(\theta^{(m)},\theta'\right)r\left(\theta^{(m)},\ \theta'\right) & 若\theta'\neq\theta^{(m)}\\ 1-\sum_{\theta}q\left(\theta^{(m)},\theta\ \right)r\left(\theta^{(m)},\ \theta\right) & 若\theta'=\theta^{(m)}\end{cases}$$

可以验证，

$$\pi\left(\theta^{(m)}\right)p\left(\theta^{(m)},\theta'\right)=\pi(\theta')p(\theta',\ \theta^{(m)})$$

这是马氏链以相应后验分布为平稳分布的一个重要条件。

实际应用中，$\theta^{(m)}\rightarrow\theta^{(m+1)}$常常仅更新一个分量，合理选择$q(\theta,\ \theta')$（如先验分布），可以使得$r\left(\theta^{(m)},\ \theta'\right)$的计算简单。

此算法包含Gibbs sampler：当取$q(\theta,\theta')=\pi(\theta_i|\theta_{-i})$时（$\theta'=\theta^{(m+1)}$），

$$\frac{\pi(\theta')q(\theta',\theta^{(m)})}{\pi\left(\theta^{(m)}\right)q(\theta^{(m)},\theta')}=1$$

即$r(\theta,\ \theta')\equiv 1$，总是接受候选。

第四节 马氏链随机模拟

● 常见方式：当可以从$\pi(\theta_i|\theta_{-i})$直接抽样时，用Gibbs sampler；否则，用Metropolis-Hastings算法。

四、估计方法

● 因为起始点$\theta^{(0)}$可能受主观因素（先验分布）影响，往往选择（大整数）M，当$m < M$时，$\theta^{(m)}$不被使用。

● $\left(\theta^{(1)}, \cdots, \theta^{(m)}, \cdots\right)$是马氏链，不是独立随机样本，故当其相关性较强时，会采取每隔若干次循环，记忆一个样本的方式。

● 抽样样本质量监控：某些参数图像、似然函数图像、……

● 后验均值、后验方差、后验中位数、后验Mode、贝叶斯置信区间等：直接由模拟样本得到。

第四节 马氏链随机模拟

- 独立产生若干个链，得到近似$i.i.d.$的样本。

- 主要成果：使得从前无法解决的问题得以解决；产生了大量新的模型；算法研究（收敛速度、模型选择等）有了极大进展；大数据研究方面有用武之地；大大推进了统计学的应用。

- 主要缺点：计算时间长，有时不收敛。

- 不完全数据的处理：将当前参数值$\boldsymbol{\theta}^{(m)}$当作真值，从缺失数据的条件分布中产生其模拟值；将此模拟值当作真数据，用全体数据及模型更新$\boldsymbol{\theta}^{(m)}$到$\boldsymbol{\theta}^{(m+1)}$。

- 无论哪种方法，总有“反例”说明其不好。

第四节 马氏链随机模拟

例1. 设$X_1, \cdots, X_n|p \sim B(1, p)$，$p|\alpha, \beta \sim Be(\alpha, \beta)$，$\alpha \sim Exp(\lambda_1)$，$\beta \sim Exp(\lambda_2)$，其中参数为$p$、$\alpha$和$\beta$，$\lambda_1$和$\lambda_2$为已知超参数，求MCMC算法。

解：层次贝叶斯图示为……。

参数θ的先验分布为

$$\pi(\theta) = Cp^{\alpha-1}(1-p)^{\beta-1}e^{-\lambda_1\alpha}e^{-\lambda_2\beta}$$

所以，

$$\pi(\theta|X) \propto p^{\alpha-1}(1-p)^{\beta-1}e^{-\lambda_1\alpha}e^{-\lambda_2\beta}p^{\sum X_i}(1-p)^{n-\sum X_i}$$
$$= p^{\alpha+\sum X_i-1}(1-p)^{\beta+n-\sum X_i-1}e^{-\lambda_1\alpha}e^{-\lambda_2\beta}$$

故而，

第四节 马氏链随机模拟

$$\pi(p|X,\alpha,\beta) \sim Be\left(\alpha+\sum X_i, \beta+n-\sum X_i\right)$$

$$\pi(\alpha|X,p,\beta) \propto e^{(\alpha+\sum X_i-1)log(p)-\lambda_1\alpha}$$
$$\sim Exp\left(\lambda_1-log(p)\right)$$

$$\pi(\beta|X,p,\alpha) \propto e^{(\beta+n-\sum X_i-1)log(1-p)-\lambda_2\beta}$$
$$\sim Exp\left(\lambda_2-log(1-p)\right)$$

因此，例1中的MCMC具体算法（之一）为：

第四节 马氏链随机模拟

① 产生起始点$\theta^{(0)} = \left(\alpha^{(0)}, \beta^{(0)}, p^{(0)}\right)$，其中$\alpha^{(0)} \sim Exp(\lambda_1)$，$\beta^{(0)} \sim Exp(\lambda_2)$， $p^{(0)} \sim Be(\alpha^{(0)}, \beta^{(0)})$；

② 从$Be(\alpha^{(0)} + \sum X_i , \beta^{(0)} + n - \sum X_i)$中模拟产生$p^{(1)}$；

③ 从$Exp\left(\lambda_1 - log\left(p^{(1)}\right)\right)$中模拟产生$\alpha^{(1)}$；

④ 从$Exp\left(\lambda_2 - log\left(1 - p^{(1)}\right)\right)$中模拟产生$\beta^{(1)}$；

⑤ 完成$\theta^{(0)} = \left(\alpha^{(0)}, \beta^{(0)}, p^{(0)}\right)$到$\theta^{(1)} = \left(\alpha^{(1)}, \beta^{(1)}, p^{(1)}\right)$的模拟。重复上述过程②至④，产生$\theta^{(m)}$……。

---------------------- end 20240601

第四节 马氏链随机模拟

① 产生起始点$\theta^{(0)}=\left(\alpha^{(0)},\beta^{(0)},p^{(0)}\right)$，其中$\alpha^{(0)}\sim Exp(\lambda_1)$，$\beta^{(0)}\sim Exp(\lambda_2)$， $p^{(0)}\sim Be(\alpha^{(0)},\beta^{(0)})$；

② 从$Be(\alpha^{(0)}+\sum X_i,\beta^{(0)}+n-\sum X_i)$中模拟产生$p^{(1)}$；

③ 从$Exp\left(\lambda_1-log\left(p^{(1)}\right)\right)$中模拟产生$\alpha^{(1)}$；

④ 从$Exp\left(\lambda_2-log\left(1-p^{(1)}\right)\right)$中模拟产生$\beta^{(1)}$；

⑤ 完成$\theta^{(0)}=\left(\alpha^{(0)},\beta^{(0)},p^{(0)}\right)$到$\theta^{(1)}=\left(\alpha^{(1)},\beta^{(1)},p^{(1)}\right)$的模拟。重复上述过程②至④，产生$\theta^{(m)}$……。

第四节 马氏链随机模拟

例2. 某个MCMC算法中的某个步骤，需要从$f(\theta) = C\left(\theta^2 + 2 + 4\big/(\theta^4 + 1)\right)$（$\theta \in [0, 1]$）中抽样，完成从$\theta^{(m)}$到$\theta^{(m+1)}$的更新，$C$是待定常数。如何完成?

解：不妨取转移函数$q(\theta^{(m)}, \theta')$为$U(0, 1)$（或某个Be分布），即独立于$\theta^{(m)}$产生$\theta^{(m+1)}$的候选θ'，则Hastings ratio $r(\theta^{(m)}, \theta')$的计算为：

$$r\left(\theta^{(m)}, \theta'\right) = \frac{\pi(\theta')q(\theta', \theta^{(m)})}{\pi\left(\theta^{(m)}\right)q(\theta^{(m)}, \theta')} = \frac{C\left({\theta'}^2 + 2 + 4\big/\left({\theta'}^4 + 1\right)\right) \cdot 1}{C\left({\theta^{(m)}}^2 + 2 + 4\Big/\left({\theta^{(m)}}^4 + 1\right)\right) \cdot 1}$$

$$= \frac{{\theta'}^2 + 2 + 4\big/\left({\theta'}^4 + 1\right)}{{\theta^{(m)}}^2 + 2 + 4\Big/\left({\theta^{(m)}}^4 + 1\right)}$$

当其≥ 1时，以概率1令$\theta^{(m+1)} = \theta'$；

当其< 1时，以概率$r\left(\theta^{(m)}, \theta'\right)$令$\theta^{(m+1)} = \theta'$，以概率$1 - r\left(\theta^{(m)}, \theta'\right)$令$\theta^{(m+1)} = \theta^{(m)}$。

有时， 部分参数以正概率不更新的情形，比Gibbs sampler有更快的收敛速度。

第八章 抽样调查概述

第一节 简介

● 统计学中重要、且很有实用价值的方向，研究如何方便、简捷、有效地得到有代表性的“好”样本。

● 常具有较大新闻价值、商业价值、社会价值等，也需要一定费用。

● “简单随机子样”可能很难、有时甚至不可能得到，例如如何抽取5000个代表所有北京人?

● 历史上，1936年，《文学摘要》发出一千多万问卷，预测兰登以57%比43%获胜，实际上罗斯福以62%比38%获胜，为何差距这么大?

① 问卷选取有偏，偏重富人;

② 回答有偏，仅1/4弱的人寄回问卷，一般多为其中中等收入者，即存在回答偏倚。

第一节 简介

● 大样本量可能无用，仅重复错误。

● 盖洛普仅用几千个样本，得到了显著改善的预测，但其预测罗斯福56%比例获胜，仍有6%偏差：获得全体选民的随机样本不容易。

● 面询抽样：常有主观偏差，找面容和善者。

● 入户调查：易抽到家庭主妇，退休人士。

● 比例抽样：比例未知或不准确。

● 流行病学调查：样本相对固定。

● 数据缺失：一些人对部分或全部问题不回答（MAR、MCAR、MNI）。

第一节 简介

问卷设计的一个（经典）例子

对敏感问题的调查中，若直接询问，被询问者常不配合，如调查吸毒率r。一个解决方法是，设是否吸毒与生日上、下半年相互独立，则概率分布如下表：

	上半年出生	下半年出生
吸毒	$r/2$	$r/2$
不吸	$(1-r)/2$	$(1-r)/2$

对路人的问题是：你是否生日在下半年，且不吸毒？问题不敏感。回答是或否的概率分别为$(1-r)/2$、$(1+r)/2$，若数据分别为n_1、n_2，则由估计方程

$$\frac{n_1}{n_2} = \frac{(1-r)/2}{(1+r)/2}$$

可得估计量$\hat{r} = (n_2 - n_1)/(n_2 + n_1)$。它是MLE。

第二节 单纯随机抽样

设总体为$\{u_1, \cdots, u_N\}$（人群有限），其中相应指标值为$\{Y_1, \cdots, Y_N\}$，我们希望估计$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$。

完全随机的抽样法为：

① 从N个单元中抽取一个，记为y_1；

② 从其余$N-1$个单元中抽取一个，记为y_2；

… …

此方法称为单纯随机抽样，取得的n个样本，称为单纯随机样本。

定理1. 设$y_1, \cdots, y_n$是从$\{Y_1, \cdots, Y_N\}$中得到的单纯随机样本，则$\overline{y} =$:$\frac{1}{n}\sum_{i=1}^{n} y_i$是$\overline{Y}$的无偏估计，其均方误差为$Var(\overline{y}) = \frac{1}{n}\left(1-\frac{n}{N}\right)S^2$，其中$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \overline{Y})^2$。

第二节 单纯随机抽样

证明：

记$D_i = \begin{cases} 1 & \text{若第}i\text{个单元进入样本} \\ 0 & \text{否则} \end{cases}$**，则**

$$E(D_i) = P(D_i = 1) = \frac{1 \cdot C_{N-1}^{n-1}}{C_N^n} = \frac{n}{N}$$

$$E\left(D_i D_j\right) = P\left(D_i = 1, D_j = 1\right) = \frac{1 \cdot 1 \cdot C_{N-2}^{n-2}}{C_N^n} = \frac{n(n-1)}{N(N-1)} \quad (i \neq j)$$

故

$$E(\bar{y}) = E\left(\frac{1}{n}\sum_{i=1}^{N} D_i Y_i\right) = \frac{1}{n}\sum_{i=1}^{N} Y_i E(D_i) = \frac{1}{N}\sum_{i=1}^{N} Y_i = \bar{Y}$$

$$E\left(\bar{y}^2\right) = \frac{1}{n^2} E\left[\left(\sum_{i=1}^{N} D_i Y_i\right)^2\right] = \frac{1}{n^2} E\left(\sum_{i=1}^{N} D_i Y_i^2 + 2\sum_{i<j} D_i D_j Y_i Y_j\right)$$

第二节 单纯随机抽样

$$=\frac{1}{n^2}\sum_{i=1}^{N}Y_i^2\frac{n}{N}+\frac{2}{n^2}\sum_{i<j}Y_iY_j\frac{n(n-1)}{N(N-1)}$$
$$=\frac{1}{nN}\sum_{i=1}^{N}Y_i^2+\frac{2(n-1)}{nN(N-1)}\sum_{i<j}Y_iY_j$$

所以

$$Var(\overline{y})=E\left(\overline{y}^2\right)-(E\overline{y})^2$$
$$=\frac{1}{nN}\sum_{i=1}^{N}Y_i^2+\frac{2(n-1)}{nN(N-1)}\sum_{i<j}Y_iY_j-\frac{1}{N^2}\sum_{i=1}^{N}Y_i^2-\frac{2}{N^2}\sum_{i<j}Y_iY_j\quad\left(=(E\overline{y})^2\right)$$
$$=\frac{N-n}{nN^2}\sum_{i=1}^{N}Y_i^2-2\frac{N-n}{nN^2(N-1)}\sum_{i<j}Y_iY_j$$
$$=\frac{N-n}{nN^2}\sum_{i=1}^{N}Y_i^2-\frac{N-n}{nN^2(N-1)}\left[\sum_{i=1}^{N}Y_i^2+2\sum_{i<j}Y_iY_j\right]+\frac{N-n}{nN^2(N-1)}\sum_{i=1}^{N}Y_i^2$$
$$=\frac{1}{n}\left(1-\frac{n}{N}\right)\frac{1}{N-1}\left[\sum_{i=1}^{N}Y_i^2-\frac{1}{N}\left(\sum_{i=1}^{N}Y_i\right)^2\right]=\frac{1}{n}\left(1-\frac{n}{N}\right)S^2$$

证毕。

第二节 单纯随机抽样

定理2. $v(\bar{y})=\frac{1}{n}\left(1-\frac{n}{N}\right)s^2$是$Var(\bar{y})$的无偏估计，其中$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$。

证明：由定理1，只需证$Es^2=S^2$，而

$$s^2=\frac{1}{n-1}\left[\sum_{i=1}^{n}y_i^2-n\bar{y}^2\right]=\frac{1}{n-1}\left[\sum_{i=1}^{N}Y_i^2D_i-n\bar{y}^2\right]$$

$$E\left(s^2\right)=\frac{1}{n-1}\left[\sum_{i=1}^{N}Y_i^2E(D_i)-nE\left(\bar{y}^2\right)\right]$$

而

$$E\left(\bar{y}^2\right)=var(\bar{y})+E(\bar{y})^2=\frac{1}{n}\left(1-\frac{n}{N}\right)S^2+\left(\frac{1}{N}\sum_{i=1}^{N}Y_i\right)^2$$

所以

$$E\left(s^2\right)$$
$$=\frac{1}{n-1}\left[\frac{n}{N}\sum_{i=1}^{N}Y_i^2-n\left(\frac{1}{n}\left(1-\frac{n}{N}\right)S^2+\left(\frac{1}{N}\sum_{i=1}^{N}Y_i\right)^2\right)\right]$$

第二节 单纯随机抽样

$$= \frac{1}{n-1}\left[\frac{n}{N}(N-1)S^2 - \left(1-\frac{n}{N}\right)S^2\right] = S^2$$

证毕。

由上面两个定理，$\bar{y}$是$\bar{Y}$一个很好的点估计，其方差可以用$v(\bar{y})$估计。

● 若想估计某年北京人均收入，只需抽取n个单纯随机样本即可（如果能做到），利用定理1，求平均。

● 若想得知上述估计的精度，利用定理2即可。

● 更有信息量的描述：给出分位点。

● 对于收入，一般地，中位数<人均收入。

关于期末考试

- 考试时间：2024年6月13日，周四，下午2：00-4：00

- 考试地点：二教203

- 考试要求：闭卷；严禁作弊；不得使用计算器等任何电子设备；手机按学校要求放好。

- 分数分布：多数为基础题，约20%有一定难度题，约20%稍难题。

- 答疑：6月12日下午2：00-4：00，6月13日上午9：00-11：00，335/智华楼。

- 期中考试及平时作业成绩：有特殊情况的提前向助教说明。

- 关于查卷：你有权利，但希望尊重别人；判分标准的解释权归改卷人。

总复习

第二章

似然函数，最大似然估计，矩估计，相合性，无偏估计，均方误差，一致最小方差无偏估计，充分统计量，完全统计量，指数分布族，C-R不等式，置信区间（置信上限、置信下限），χ^2分布，t分布，自由度，枢轴量方法，统计量方法*，经验分布函数，直方图法，核估计法*。

第三章

零假设，否定域，随机化检验法*，第I类错误，第II类错误，功效函数， UMP检验，无偏检验， UMPU检验， N-P引理，似然比检验法，单参数指数族的四种典型问题（四个定理），假设检验与置信区间的对应关系，广义似然比检验法，检验法的相合性*，一些常见（正态）问题的检验法， F分布及其性质，临界值与p值，拟合优度检验*（χ^2检验、独立性检验、 Kolmogorov检验）。

总复习

第四章

自（协）变量，因（响应）变量，一元线性回归，最小二乘估计，残差，平方和分解，相关系数，预测，控制；线性模型，投影方法，线性可估性，高斯-马尔科夫定理，带约束的线性模型的参数估计，线性模型的假设检验（定理4.1、定理4.3等），多元回归模型的估计、检验、预测等，回归方程的解释，自变量选择*，线性性检验与残差分析*，Logistic模型*，Probit模型*。

第五章 试验设计与方差分析

单因素、两因素全面试验的方差分析，可加模型，重复试验，平方和（方差）分解公式及其背景和统计意义，方差分析表，正交设计*。

总复习

第六章 序贯分析初步

随机游动*，停止时间*，序贯判决法则*，序贯概率比检验*。

第七章 统计决策与贝叶斯统计大意

决策论，损失函数，可容许性，minimax准则，效用函数*，主观概率，先验分布，后验分布，贝叶斯置信区间，预测分布*，贝叶斯决策，共轭分布族，马尔科夫链随机模拟*。

第八章 抽样调查概述

抽样调查的意义*，单纯随机抽样*。

感谢所有同学的参与！

感谢助教王啸辰、杨云帆的帮助！