第 3 讲：先验分布的选取

张伟平

目录

本讲目的

- 学习**主观先验**分布方法和**客观贝叶斯**的不同先验选取方法
- 推荐阅读
  - Berger, J. (2006). The case for objective Bayesian analysis (with discussion). Bayesian Analysis, 1, 385–482.
  - Goldstein, M. (2006). Subjective Bayesian analysis: principles and practice (with discussion). Bayesian Analysis, 1, 403–420.

3.1 主观先验

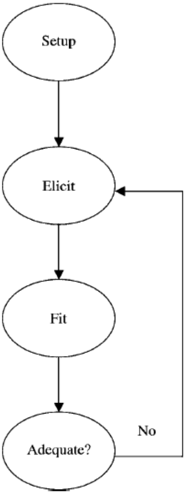
3.1.1 主观概率

主观概率是人们根据经验对事件发生机会的个人信念. 例如对一场比赛胜负的打赌, 对明天是否下雨的估计, 对股票市场行情的估计等都是主观概率. 当感兴趣未知参数是离散 (有限个) 取值时候, 常用**相对似然法**,**专家意见**和**利用历史资料**等方法确定先验概率.

专家意见方法 在很多实际问题中, 存在多个独立地, 具有重要信息的专家. 此时, 将这些专家信息转换为概率是非常重要的. 概率启发 (probability elicitation) 方法就是用来将专家意见转为概率的方法.

Garthwaite et al (2005, JASA) 将概率启发过程总结为以下四个阶段:

- 设立: 准备启发: 选择专家, 训练专家, 识别问题的哪些方面要启发, 等等
- **专家意见的启发**: 对目标问题的这些方面得出专家们分布的特定特征量 (数字特征). 这是整个过程的核心阶段, 也是心理学家和统计学家贡献至少一样多的阶段
- **分布拟合**: 对这些特征量拟合一个概率分布. 选取何种特征量用于概率启发常常受到诱导者试图进行拟合的分布形式选取的影响。在实际中此步骤和前一步骤界限模糊.
- **评估充足性**: 评估启发过程的充足性, 否则返回到第二步得出更多总结统计量.



专家训练的问题

专家多使用启发式方法来进行概率判断，这会产生偏差和不连贯性

- 动机偏差
- 认知偏差
  - 可得性偏差
  - 锚定与调整性偏差
  - 代表性偏差
  - control

专家训练的一个重要部分就是尽量使专家认识到这些偏差，从而消除它们

■ **可得性偏差**: 即由于最近的影响或者感情上、特征上的特殊印象，容易高估事情出现的可能性

例如, 下面事件对, 哪一种导致的年死亡更多?

- 胃癌和交通事故
- 结核病和火灾

死亡原因	选择	年死亡率/1000	新闻报刊报道数
胃癌	14%	95	46
交通事故	86%	1	137
结核病	23%	4	0
火灾	77%	5	0

人们对交通事故具有比胃癌更多的信息, 因此 “交通事故” 这个选项的可得性更大. 更多例子参阅 Russo and Shoemaker (1989) 和 Tversky and Kahneman(1973).

■ **锚定与调整性偏差**: 人们通常利用某个参照点和锚（Anchor）来降低模糊性，然后再通过一定的调整来得出最后的结论

例如 (Tversky y Kahneman 1974) 研究者希望能得出非洲国家在联合国百分比的估计. 一组专家被要求回答下述问题:

你认为这个百分比是高于还是低于 10%?

另一组专家被要求回答

你认为这个百分比是高于还是低于 65%?

两组专家中的每个人再给出该比例的一个点估计. 第一组的平均估计是 25%, 第二组的平均值为 45%. 显然, 一个随机不相关的锚点影响了两组人的估计.

■ **代表性偏差**: 在做决定时候是借鉴要判断事件本身或事件的同类事件以往的经验即以往出现的结果来进行

费德里克今年 35 岁, 聪明, 但不怎么有想象力, 有点无聊. 在大学里, 他在数学方面表现出很多天赋, 但是他不是很擅长艺术.

将下述关于费德里克的说法按照其概率 (最有可能 =1, 最不可能 =8) 进行排序:

1. 费德里克是一个医生, 玩纸牌是他的一个爱好
2. 他是一个建筑师
3. 他是一个会计师
4. 他弹爵士乐器
5. 他爱读体育报纸
6. 他喜欢登山
7. 他是会计师, 弹奏爵士乐器
8. 他是一名记者

大部分人认为 3 最有可能. 而且, 很多人认为 7 比 4 更有可能. 但这是不可能的, 因为对两个事件 A 和 B 满足

$$P(A \cap B) \leq \min\{P(A), P(B)\}$$

此问题说明了代表性启发式方法偏差以及基本比例谬论 (base rate fallacy)[1]。更多例子参见 Kahneman et al (1982)

基本比例谬论的例子. 一辆出租车晚上在达灵顿撞了一名行人. 达灵顿只有两个出租车公司. 第一家的出租车是绿色的, 第二家是蓝色的. 达灵顿的出租车中约有 85%是绿色的. 一位事故的目击者说出租车是蓝色的. 当在与事故当晚相同的气候条件下进行测试时, 目击者可以正确识别了大约 80%的测试车辆颜色. 你估计出租车是蓝色的概率是多少?

[1] In probability and statistics, base rate generally refers to the (base) class probabilities unconditioned on featural evidence, frequently also known as prior probabilities.

正确率是 80%, 但是如果用 A 表示出租车是蓝色的, a 为目击者说是蓝色的, 则由贝叶斯定理

$$\begin{aligned} P(A|a) &= \frac{P(a|A)P(A)}{P(a|A)P(A) + P(a|A)P(\bar{A})} \\ &= \frac{0.8 \times 0.15}{0.8 \times 0.15 + 0.2 \times 0.85} = 0.41 \end{aligned}$$

人们进行预测时候经常忽略基础比例.

概率启发

首先我们考虑得出一个专家 E 对事件 A 的概率 $P_E(A)$, 最简单的方法可能就是直接询问该专家的概率. 但是, 这种做法不允许该专家考虑他的所有概率, 同时如果他没有被训练的很好, 则将很难实施.

有两种基于**概率尺度**或**赌博机制**的两种替代方法.

- 最简单的概率尺度方法就是画一条线段, 最左边表示概率 0, 最右边表示概率 1, 然后让该专家在线段上标记一点表示他对该事件的概率. 这种做法不允许专家估计小概率或者大概率, 当我们要估计这类事件概率时, 最好使用机会比或者对数尺度. 此外, 在线段上包含特定的指导点值也是有帮助的, 尽管可能会导致锚点偏差.

- 另外一种方法就是考虑有一个大奖和一个非常小的奖的赌博. 例如我们想要得出一个医生对一位病人有肿瘤的概率判断. 我

们可以让医生选择玩下面的两个赌博游戏之一：依概率 p 获得度假, $1-p$ 获得巧克力; 有肿瘤时得到度假, 没有肿瘤时得到巧克力. 对于给定的 p, 我们可以检查医生偏好哪一种游戏. 变化 p 直至在某个点 $p = p_E$ 时医生对两个游戏没有偏好差异. 当然, 这种做法有伦理问题, 我们可以使用概率转盘来代替上述游戏. 这种方法只适用于二分事件概率问题, 也不适用与估计特别小或特别大的概率.

- 其他方法还有基于频率的方法 (Price 1998), 转换语言描述例如可能, 不大可能为数值概率的方法 (Witteman and Renooij 2003). Wallsten et al(1993) 做了一个比较全面的综述.

- 这些方法可以推广为得出专家的分布. 比较常见的是得出专家的概率分布的某些分位点而不是完整的分布. 这些分位点进一步可以使用设定的模型来拟合.

先验分布的启发

启发先验概率有很多种方式, 最坏的就是直接询问专家关于先验分布参数的值.

↑Example

假设我们感兴趣估计一个有偏硬币正面的概率, p, 的先验分布. 我们使用 (共轭) 先验分布为 Beta 分布, $p \sim Beta(\alpha, \beta)$. 希望得出 α, β 的值.

↓Example

一种可能方法是直接询问专家关于 p 的最有可能值, 专家的众数, p_E. 以及"p 位于区间 $(p_E - Kp_E, p + Kp_E)$" 这一说法的概率 r_E. 其中 K 为通过分析者依据信息而给定的固定值. 那么假设 p 有 Beta 分

布, 我们可以通过解以下方程得到 α, β 的值:

$$p_E = \frac{\alpha - 1}{\alpha + \beta - 2}$$

$$r_E = \int_{p_E - K p_E}^{p_E + K p_E} \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} dx$$

一般来说, 询问专家的可观测量更好一些. Chaloner and Duncan (1983) 提出让专家说出他估计的 n 重 Bernoulli 试验中成功次数 X 的众数 x_E, 以及下述 c_E 和 d_E:

$$c_E = \frac{P_E\left(X = x_E - 1\right)}{P_E\left(X = x_E\right)}, \quad d_E = \frac{P_E\left(X = x_E + 1\right)}{P_E\left(X = x_E\right)}$$

注意 X 的边际分布是 beta-binomial, 因此上述等式可以表示为

$$c_E = \frac{\left(n - x_E\right)\left(x_E + \alpha\right)}{\left(x_E + 1\right)\left(n - x_E + \beta - 1\right)}, \quad d_E = \frac{x_E\left(n - x_E + \beta\right)}{\left(n - x_E + 1\right)\left(n - \alpha + 1\right)}$$

从而可以解出 α, β. Hughes and Madden (2002) 讨论了其他的估计方法.

评估专家预报质量

常用的准则 (Lichtenstein et al 1982) 如下

- 诚实: 我们希望专家讲实话
- 一惯性: 他的预报应该满足概率法则.
- 一致性: 如果他没有新信息, 那么他的预测不变
- 校正: 如果专家说下雨的概率为 0.5, 那么 50% 的日子里是下雨的
- 信息性: 如果在一个地方一年大约有 50 天下雨, 那么一个专家说明天下雨的概率是 50/364, 则每天都不是很有信息.

■ **诚实与严格适当的得分规则** 假设我们希望得出一个专家对某个事件 A 发生的真实概率 p_E. 一种为鼓励专家是诚实的做法是给他一定的奖励 $R(A,p)$, 其依赖于 A 是否发生和他说的概率 p. 我们该如何定义 $R(A,p)$?

我们假设专家希望最大化他的期望收入. 如果 p_E 是他的真实概率, 则他说概率为 p 时的期待收入为

$$p_E R(1,p) + (1-p_E)\, R(0,p)$$

Definition

一个（严格）适当的得分规则（Savage 1971）是当（且仅当）$p = p_E$ 时他能最大化期望收入的规则 $R(A,p)$.

假设 $R(A,p) = 1 - |A - p|$, 则专家在说概率为 p 下的期待收入

为

$$\begin{aligned} E[R] &= p_E(1-|1-p|)+(1-p_E)(1-|0-p|) \\ &= p_Ep+(1-p_E)(1-p) \\ &= 1-p_E+(2p_E-1)p \end{aligned}$$

因此 $R(A,p)$ 不是一个恰当的得分规则, 如果 $p_E > (<) 0.5$ 则专家可以通过说 $p=1(0)$ 来最大化他的期望收入.

Brier(1950) 得分 $R(A,p)=1-(A-p)^2$. 则

$$\begin{aligned} E[R] &= p_E\left(1-(1-p)^2\right)+(1-p_E)\left(1-p^2\right) \\ &= 1-p_E+2pp_E-p^2 \\ &= 1-p_E+p_E^2-(p-p_E)^2 \end{aligned}$$

其最大值在 $p=p_E$ 处达到. 因此 R 为一个严格适当的得分规则.

还有其他的适当的得分规则, 也已经被推广到连续型随机变量. 例如 Winkler (1986), Buehler (1971) and Matheson and Winkler (1976).

假设专家 E 被要求给出变量 X 的一个点估计 e, 考虑如下得分规则

$$R(X,e)=\begin{cases} a(e-x) & \text{if } e<x \\ b(x-e) & \text{if } e>x \end{cases}$$

记 p_E 为专家对 X 的真实分布, 则

$$\begin{aligned} E[R(X,e)] &= \int R(x,e)p_E(x)dx \\ &= a\int_e^{\infty}(e-x)p_E(x)dx + b\int_{-\infty}^{e}(x-e)p_E(x)dx \\ &= ae\left(1-F_E(e)\right) - a\int_e^{\infty} xp_E(x)dx + \\ & b\int_{-\infty}^{e} xp_E(x)dx - beF_E(e) \end{aligned}$$

因此

$$
\begin{aligned}
\frac{dE[R(X,e)]}{de} =& a\left(1-F_E(e)\right)-aep_E(e)+aep_E(e)- \\
& bep_E(e)-bF_E(e)+bep_E(e) \\
0 =& a\left(1-F_E(\hat{e})\right)-bF_E(\hat{e}) \\
F_E(\hat{e}) =& \frac{a}{a+b}
\end{aligned}
$$

专家最大化他的期望收益, 如果他说的是 $b/(a+b)\times 100\%$ 分位数.

使用适当的得分规则以鼓励诚实看起来有些人为性, 但是它们或许可以被用来作为专家概率值的一个后验评估工具.

■ **专家质量的数值度量** 假设专家为一列 Bernoulli 事件 $X_1,\ldots,X_n$ 给出了概率 $\mathbf{p}_E=(P_{E_1},\ldots,p_{E_n})$. 给定数据 $\mathbf{x}$, 我们如何评价专家预测的质量?

考虑 Brier 得分 $R(X,p_E)=1-(X-p_E)^2$. 给定数据, 我们可以计算统计量

$$R(\mathbf{x},\mathbf{p}_E)=\frac{1}{n}\sum_{i=1}^{n}R(x_i,p_{E_i})=1-\sum_{i=1}^{n}(x_i-p_{E_i})^2$$

用来度量他预测质量的平均水平. 此度量可以分为校正度量和信息度量. 根据 Murphy(1973), 假设专家使用概率 p_j, n_j 次, 成功的次数是 f_j, 以及一个相对频率 $r_j=f_j/n_j, j=1,\ldots,k$. 则

$$\begin{aligned}R(\mathbf{x},\mathbf{p})&=1-\frac{1}{n}\sum_{j=1}^{k}\left(f_j(1-p_j)^2+(n_j-f_j)(0-p_j)^2\right)\\&=1-\frac{1}{n}\sum_{j=1}^{k}n_j\left(r_j(1-p_j)^2+(1-r_j)(0-p_j)^2\right)\end{aligned}$$

我们有下述定理

定理 1.

$$R(\mathbf{x},\mathbf{p}) = 1 - C(\mathbf{x},\mathbf{p}) - I(\mathbf{x},\mathbf{p})$$

其中

$$C(\mathbf{x},\mathbf{p}) = \frac{1}{n}\sum_{j=1}^{k} n_j \left(r_j - p_j\right)^2 \qquad \textit{is a measure of calibration}$$

$$I(\mathbf{x},\mathbf{p}) = \frac{1}{n}\sum_{j=1}^{k} n_j r_j \left(1 - r_j\right) \qquad \textit{is a measure of information}$$

证明.

$$
\begin{aligned}
R(\mathbf{x},\mathbf{p}) &= 1-\frac{1}{n}\sum_{j=1}^{k} n_j\left(r_j\left(1-p_j\right)^2+\left(1-r_j\right)\left(0-p_j\right)^2\right)\\
&= 1-\frac{1}{n}\sum_{j=1}^{k} n_j\left(r_j-2r_jp_j+p_j^2\right)\\
&= 1-\frac{1}{n}\sum_{j=1}^{k} n_j\left(r_j-r_j^2+r_j^2-2r_jp_j+p_j^2\right)\\
&= 1-\frac{1}{n}\sum_{j=1}^{k} n_j\left(r_j\left(1-r_j\right)+\left(r_j-p_j\right)^2\right)\\
&= 1-I(\mathbf{x},\mathbf{p})-C(\mathbf{x},\mathbf{p})
\end{aligned}
$$

$\square$

C 有以下性质

- $0 \leq C \leq 1$
- $C = 0$ 当且仅当 $r_j = p_j$, $j = 1, \ldots, k$
- 对一个校正较好的专家, 当 $n \to \infty$ 时, $C \to 0$
- 如果观测的相对频率 r_i 与专家所述概率 p_i 差异较大时候, C 值比较大.

I 有以下性质

- $0 \leq I \leq 0.25$
- 如果所有的 p_j, 相对频率 $r_j = 0$ 或 1, 则 $I = 0$
- 如果对所有 p_j, $I = 0.25$, 则 $r_j = 0.5$

任何严格适当的得分规则可以类似的被分为校正和信息度量. 参考 De Groot and Fienberg (1982).

Wiper (1987,1990) 给 12 个专家 50 个陈述进行研究. 专家被要求每个陈述是真还是假, 并给出这个其结论是正确的概率, 要求用 50% 到 99% 之间的数值表示. 我们假设专家给出自己判断正确的概率属于 $\mathbf{p} = \{0.53, 0.64, 0.75, 0.86, 0.97\}$. 下表总结了 3 位专家相关的绝对和相对频数.

E	p_i	0.53	0.64	0.75	0.86	0.97
	n_i	25	6	6	5	8
2	f_i	15	4	4	3	8
	r_i	0.6	0.67	0.67	0.6	1.0
	n_i	25	5	10	5	5
3	f_i	16	1	3	2	4
	r_i	0.64	0.2	0.3	0.4	0.8
	n_i	10	5	15	1	19
10	f_i	10	5	15	0	15
	r_i	0.6	0.4	0.3	1.0	0.79

则下表表示了每位专家的校正, 信息和 Brier 得分:

E	C	I	Brier
2	.0093	.1973	.7934
3	.0900	.2132	.6968
10	.1059	.2018	.6923

可以看出, 专家 2 校正的比较好, 但是比其他专家信息少. 观测到的频率 r_i 对不同的概率 p_j 作图称为专家校正曲线. 一个校正非常好的专家, 其校正曲线应该接近对角线.

Cooke 的方法

Cooke et al (1988) and Cooke (1991) 发展了基于经典 p 值的度量校正和信息的方法, 并且可以用于离散和连续变量的预测.

假设专家 j 使用了 n_j 次 $p_j, j = 1, \ldots, k$. 则可以使用 χ^2 检验来检验观测到的 $r_1, \ldots, r_k$ 是否可以从理论分布 $p_1, \ldots, p_k$ 生成, 即检验零假设 $H_0 : r \sim p \leftrightarrow H_1 : r \not\sim p$:

$$S = \sum_{j=1}^{k} n_j \frac{(r_j - p_j)^2}{r_j}$$

检验的 p 值为 $P_{\chi^2_k}(S > S_{obs}|H_0)$.

例如前例中使用 Cooke 的方法得到

E	2	3	10
p	.7	.00	.00

可以看出, 只有专家 2 接收原假设, 即校正的非常好.

Cooke(1991) 基于组合 p 值和一种信息度量提出了关于得分规则的一种理论. 当专家对连续变量进行预测时候, 则可以考虑使用 Kolmogorov-Smirnov test. 参考 Wiper et al. (1994).

3.2 利用边缘分布确定先验分布

3.2.1 边缘分布的定义及例

- 设 r.v. X 有概率函数 $f(x|\theta)$, θ 有先验分布 $\pi(\theta)$, 则随机变量 X 的边缘分布为

$$
\begin{aligned}
m(x) &= \int_{\Theta} f(x|\theta)dF^{\pi}(\theta) \\
&= \begin{cases} \int_{\Theta} f(x|\theta)\pi(\theta)d\theta, & \text{当 } \theta \text{ 为连续随机变量,} \\ \sum\limits_{i} f(x|\theta_i)\pi(\theta_i), & \text{当 } \theta \text{ 为离散随机变量.} \end{cases}
\end{aligned}
$$

- 当先验分布含有未知的超参数 λ, 如 $\pi(\theta) = \pi(\theta|\lambda)$, 则边缘分布 $m(x)$ 也依赖 λ, 此时可记 $m(x) = m(x|\lambda)$. 这种边缘分布在本节后面求后验分布时常用到.

混合分布及注释

- 为了说明边缘分布的统计意义，下面引入混合分布的概念．设随机变量

$$X \sim \begin{cases} F(x|\theta_1), & \pi \\ F(x|\theta_2), & 1-\pi \end{cases}$$

则 X 的混合分布函数为

$$F(x) = \pi\ F(x|\theta_1) + (1-\pi)F(x|\theta_2).$$

- 视 θ_1, θ_2 为 θ 的取值, 满足 $\pi(\theta_1) = \pi$, $\pi(\theta_2) = 1-\pi$. 则上式为边际分布的表示形式.

- 从混合分布的定义可见, 边缘分布 $m(x)$ 是混合分布的推广.

3.2.2 选择先验分布的 ML-II 方法

- 当观测到样本 $X = x$, 若有两个个先验分布 π_1 和 π_2, 使得

$$m(x|\pi_1) > m(x|\pi_2)$$

则认为先验取 π_1 时样本 $X = x$ 出现的似然性比先验取为 π_2 时的似然性更大，即认为样本 X 由分布 $m(x|\pi_1)$ 中产生．这一思想与经典统计方法中的极大似然原理相类似，这儿的 π 起似然函数中 θ 的作用, 这就引入了 ML-II 方法.

- 设 Γ 为所考虑的先验类,若存在 $\hat{\pi} \in \Gamma$, 有了样本 $\mathbf{x} = (x_1, \cdots, x_n)$ 后, 使得

$$m(\mathbf{x}|\hat{\pi}) = \sup_{\pi \in \Gamma} m(\mathbf{x}|\pi) = \sup_{\pi \in \Gamma} \prod_{i=1}^{n} m(x_i|\pi)$$

则称 $\hat{\pi}$ 为类型 II 中最大似然先验, 或简称 ML-II 先验.

选择先验分布的 ML-II 方法

- 若样本 $\mathbf{X} = (X_1, \cdots, X_n)$ 所涉及的先验密度函数的形式已知，未知的仅是其中的超参数，如 $\theta \sim N(\mu_\pi, \sigma_\pi^2)$, μ_π 和 σ_π^2 称为超参数，即先验密度的类 Γ 可表为

$$\Gamma = \left\{\pi(\theta|\lambda), \lambda \text{为超参数 (或超参数向量)}, \ \lambda \in \Lambda\right\},$$

此处 Λ 为 λ 的参数空间.

- 这时寻求 ML-II 先验较为简单. 只要求出这样的 $\hat{\lambda}$, 使得

$$m\big(\mathbf{x}|\hat{\lambda}\big) = \sup_{\lambda \in \Lambda} m\big(\mathbf{x}|\lambda\big) = \sup_{\lambda \in \Lambda} \prod_{i=1}^{n} m(x_i|\lambda)$$

即通过使似然函数极大化方法求出 $\hat{\lambda}$, 则先验分布 $\pi\big(\theta|\hat{\lambda}\big)$ 即为所确定的先验分布. 而 $\hat{\lambda}$ 称为 ML-II 超参数.

例 2.3.2(P37) 设 $X \sim N(\theta, 1)$, 参数 θ 的先验分布取为共轭先验 $N(\mu_\pi, \sigma_\pi^2)$, 其中 $\lambda = (\mu_\pi, \sigma_\pi^2)$ 未知. 设 $\mathbf{X} = (X_1, \cdots, X_n)$ 为从边缘分布 $m(x|\lambda)$ 中抽取的 iid 样本, 由样本算出样本均值 $\bar{X} = 10$, $S^2 = 3$. 试确定 θ 的先验分布.

↑Example

↓Example

解 由边缘分布 $X \sim N(\mu_\pi, \sigma^2 + \sigma_\pi^2)$ 知超参数的对数似然函数为

$$\ell(\mu_\pi, \sigma_\pi^2) \propto -\frac{n}{2} log(1 + \sigma_\pi^2) - \frac{nS^2}{2(1 + \sigma_\pi^2)} - \frac{n(\bar{x} - \mu_\pi)^2}{2(1 + \sigma_\pi^2)}$$

令导数为零, 求解得到

$$\hat{\mu}_\pi = \bar{x}, \quad \hat{\sigma}_\pi^2 = (S^2 - 1)I(S^2 > 1)$$

3.2.3 选择先验分布的矩方法

- 当先验分布 $\pi(\theta|\lambda)$ 的形式已知，但含有未知超参数 λ 时, 可利用先验分布的矩与边缘分布的矩之间的关系寻求超参数 λ 的估计量 $\hat{\lambda}$, 从而获得先验分布 $\pi(\theta|\hat{\lambda})$.

- 这个方法的思想如下:

  - 首先，将边缘分布一些阶的矩表成超参数的函数，得到一个方程组或方程;

  - 将方程组或方程中的边缘分布的矩用相应的样本矩代替, 得到以超参数为变量的方程或方程组, 解方程或解方程组获得超参数的估计量，从而确定先验分布.

例 2.3.3(P39) 设 $X \sim N(\theta, 1)$, 参数 θ 的先验分布取为共轭先验 $N(\mu_\pi, \sigma_\pi^2)$, 其中 $\lambda = (\mu_\pi, \sigma_\pi^2)$ 未知. 设 $\mathbf{X} = (X_1, \cdots, X_n)$ 为从边缘分布 $m(x|\lambda)$ 中抽取的 iid 样本, 由样本算出样本均值 $\bar{X} = 10$, $S^2 = 3$. 试确定 θ 的先验分布.

↑Example

↓Example

解 由边缘分布 $X \sim N(\mu_\pi, \sigma^2 + \sigma_\pi^2)$ 知边缘分布的期望和方差为

$$\mu_m = \mu_\pi$$
$$\sigma_m^2 = 1 + \sigma_\pi^2$$

使用样本值 $\bar{X}$ 和 S^2 分别代替 μ_m 和 σ_π^2, 并带入数值求解即得 $\mu_\pi = 10, \sigma_\pi^2 = 2$.

3.3 客观贝叶斯方法

有些时候我们希望使用不包含主观信息的先验分布, 因为

- 我们对手边的问题一无所知
- 我们希望是客观的 (objective)

在这些情形下, 我们应该选择无信息先验分布 (non-informative prior). 但是存在很多种选择, 哪一种更有用呢?

3.3.1 均匀先验

Bayes(1763) 和 Laplace(1812) 选择均匀先验作为一般的无信息先验, 其合理性的根据是不充分推理原理. 这在有限维问题中没有问题, 但是当参数空间是无界连续时候, 这种先验分布是不正常的 (不满足密度积分为 1). 但这个问题并不严重, 只要后验密度是正常密度即可. 此时称该先验为**广义先验分布**.

设随机变量 $X \sim f(x|\theta)$, $\theta \in \Theta$, 若 θ 的先验分布 $\pi(\theta)$ 满足条件: (i) $\pi(\theta) \geq 0$ 且 $\int_\Theta \pi(\theta)d\theta = \infty$, (ii) 后验密度 $\pi(\theta|x)$ 是正常密度函数, 则称 $\pi(\theta)$ 为广义先验密度. Definition

均匀先验的一个重要问题是其**缺乏变换不变性**. 例如设 $\pi(\theta) \propto 1$, 令 $\phi = 1/\theta$, 则意味着 ϕ 的先验分布 $\pi(\phi) \propto 1/\phi^2$, 这不是均匀分布了. 因此使用均匀分布表示无信息是不一致的: 如果我们对 θ 一无所知, 我们应该对 ϕ 也一无所知.

↑Example

(位置参数族) 设 $X \sim f(x-\theta), \theta \in R$ 为参数. 求 θ 的无信息先验.

↓Example

- 因为位置参数族在平移变换下不变, 即对 X 作平移变换得到 $Y = X + c$, 同时对 θ 也作平移变换得到 $\eta = \theta + c$. 显然 Y 的密度函数有形式 $f(y-\eta)$, η 仍为位置参数. 所以 (X, θ) 与 (Y, η) 的统计问题结构相同. 因此主张它们有相同的无信息先验是合理的.

- 理解这一点的另一方法: X 和 Y 的测量原点不同, 由于测量原点的选择是非常任意的, 所以无信息先验应当与这种选择无关. 如果无信息先验不依赖于原点的选择, 则它在等长区间内的先验概率应当一样. 即取 θ 的无信息先验密度 $\pi(\theta) \equiv 1$. 它是一

个广义先验密度. 可以证明当 θ 为位置参数时, 其无信息先验密度取为常数 c 或者 1.

例 2.4.2(P42) 设 $X_1, \ldots, X_n$ 为来自 $N(\theta, 1)$ 的样本, 若选用 θ 的无信息先验, 求 θ 的后验期望估计.

↑Example
↓Example

↑Example

(刻度参数分布族) 设总体 X 的密度函数有形式 $\sigma^{-1}\varphi(x/\sigma)$, 其中 $\sigma>0$ 为刻度参数, 参数空间为 $R^+=(0,\infty)$, 求 σ 的无信息先验分布.

↓Example

- 刻度参数族具有在刻度变换群下的不变性. 对 X 作变换 $Y=cX$, $c>0$, 同时对 σ 作相应的变换 $\eta=c\sigma$. 不难算出 Y 的密度仍为

$$\eta^{-1}\varphi(y/\eta).$$

可见 (X,σ) 和 (Y,η) 统计问题的结构相同, 故主张 σ 与 η 的无信息先验相同.

- 理解这一点的另一方法: X 和 Y 的度量单位不同, 先验分布应当不依赖于度量单位的选择, 则对任何 a, b, $0<a<b$,

$c > 0$, σ 落在 $[a, b]$ 内的先验概率, 应当等于 η 落在 $[ca, cb]$ 内的先验概率. 可以证明, 这只有在先验密度为 $1/\sigma$ 时才可能, 即取 σ 的无信息先验为 $\pi(\sigma) = 1/\sigma, \quad \sigma > 0.$

例 2.4.3(P44) 设 $X_1, \ldots, X_n$ 为来自 $Exp(1/\lambda)$ 的样本, 若选用 λ 的无信息先验, 求后验期望和方差.

↑Example
↓Example

3.3.2 Jeffreys 先验

Jeffreys

Jeffreys(1946) 介绍一个具有变换不变性质的先验分布.

设 $X \sim f(\cdot|\theta)$, θ 为未知参数. θ 的 Jeffreys 先验为

$$\pi(\theta) \propto \sqrt{det I(\theta)}$$

Definition

其中 $I(\theta)$ 为 Fisher 信息阵.

下述定理揭示了 Jeffreys 先验的变换不变性:

定理 2. 设 $\phi = h(\theta)$ 为一单调函数, $\pi(\theta) \propto \sqrt{det I(\theta)}$ 为 *Jeffreys* 先验, 则 ϕ 的 *Jeffreys* 先验分布为

$$\pi(\phi) \propto \sqrt{det I(\phi)}, \quad i.e., \quad [I(\phi)]^{1/2} = [I(\theta)]^{1/2} |\frac{d\theta}{d\phi}|$$

其中 $I(\phi)$ 为 X 的分布重参数化为 ϕ 后的 *Fisher* 信息量.

证明. 首先由链式求导法则，如果 $y = f(g(x))$, 则 $\frac{dy}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx}$, 进而

$$\frac{d^2y}{dx^2} = \frac{d^2f}{dg^2} \cdot \frac{dg}{dx} \cdot \frac{dg}{dx} + \frac{df}{dg} \cdot \frac{d^2g}{dx^2} = \frac{d^2f}{dg^2} \cdot \left(\frac{dg}{dx}\right)^2 + \frac{df}{dg} \cdot \frac{d^2g}{dx^2}$$

由 $\phi = h(\theta)$, 则 $\theta = h^{-1}(\phi)$. 故

$$\frac{d^2 \log f(x \mid \phi)}{d\phi^2} = \frac{d^2 \log f(x \mid \phi)}{d\theta^2} \cdot \left(\frac{d\theta}{d\phi}\right)^2 + \frac{d \log f(x \mid \theta)}{d\theta} \cdot \frac{d^2\theta}{d\phi^2}$$

因此

$$\begin{aligned} I(\phi) &= -E_{X|\phi}\left[\frac{d^2}{d\phi^2} \log f(X \mid \phi)\right] \\ &= -E_{X|\phi}\left[\frac{d^2 \log f(X \mid \phi)}{d\theta^2} \cdot \left(\frac{d\theta}{d\phi}\right)^2 + \frac{d \log f(X \mid \theta)}{d\theta} \cdot \frac{d^2\theta}{d\phi^2}\right] \end{aligned}$$

$$
\begin{aligned}
&= -E_{X|\phi}\left[\frac{d^2 \log f(X \mid \phi)}{d\theta^2}\right] \cdot \left(\frac{d\theta}{d\phi}\right)^2 - E_{X|\phi}\left[\frac{d \log f(X \mid \theta)}{d\theta}\right] \cdot \frac{d^2\theta}{d\phi^2} \\
&= -E_{X|\phi}\left[\frac{d^2 \log f(X \mid \phi)}{d\theta^2}\right] \cdot \left(\frac{d\theta}{d\phi}\right)^2 \\
&= I(\theta) \cdot \left(\frac{d\theta}{d\phi}\right)^2
\end{aligned}
$$

其中利用 $E\left[\frac{d}{d\theta}\log f(X|\theta)\right] = 0.$ $\square$

例 2.4.5(P46) 设 $X \sim Binom(n, \theta)$, 求 θ 的 Jeffreys 先验.

↑Example
↓Example

解 由于

$$
\begin{aligned}
\log f(X|\theta) =& c + X\log\theta + (n-X)\log(1-\theta) \\
\frac{d}{d\theta}\log f(X|\theta) =& \frac{X}{\theta} - \frac{(n-X)}{(1-\theta)} \\
\frac{d^2}{d\theta^2}\log f(X|\theta) =& -\frac{X}{\theta^2} - \frac{(n-X)}{(1-\theta)^2} \\
E\left[\frac{d^2}{d\theta^2}\log f(X|\theta)\right] =& -n\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) = -n\frac{1}{\theta(1-\theta)} \\
I(\theta) \propto& \frac{1}{\theta(1-\theta)}
\end{aligned}
$$

因此 Jeffreys 先验为 $\pi(\theta) \propto \sqrt{\frac{1}{\theta(1-\theta)}}$, 即 $\theta \sim Beta(1/2, 1/2)$. 这是一个正常先验.

设 $X \sim N(\mu, 1/\phi)$, ϕ 已知, 求 μ 的 Jeffreys 先验.

↑Example
↓Example

解因为

$$\begin{aligned}
\log f(X|\mu) &= c - \frac{\phi}{2}(X-\mu)^2 \\
\frac{d}{d\mu}\log f(X|\mu) &= \phi X - \phi\mu \\
\frac{d^2}{d\mu^2}\log f(X|\mu) &= -\phi
\end{aligned}$$

所以 $\pi(\mu) \propto 1$.

设 $X \sim N(\mu, 1/\phi)$, μ 已知, 求 ϕ 的 Jeffreys 先验.

↑Example
↓Example

解 因为

$$\log f(X|\phi) \propto \frac{1}{2}\log\phi - \phi\frac{(X-\mu)^2}{2}$$

$$\frac{d}{d\phi}\log f(X|\phi) = \frac{1}{2\phi} + \frac{(X-\mu)^2}{2}$$

$$\frac{d^2}{d\phi^2}\log f(X|\phi) = -\frac{1}{2\phi^2}$$

所以 $\pi(\phi) \propto 1/\phi$. 如果 σ 为标准偏差, 则 $\phi = 1/\sigma^2$, $\frac{d\phi}{d\sigma} = -\frac{2}{\sigma^3}$, 因此 σ 的 Jeffreys 先验为

$$p(\sigma) \propto \sigma^2\frac{2}{\sigma^3} \propto \frac{1}{\sigma}$$

多元参数下的 Jeffreys 先验

设 $X \sim MN(m, \theta)$ 为 k 维多项分布, 则求 $\theta = (\theta_1, \dots, \theta_k)$ 的 Jeffreys 先验.

↑Example
↓Example

解因为

$$\log f(X|\boldsymbol{\theta}) = c + \sum_{i=1}^{k} x_i \log \theta_i$$

$$\frac{d}{d\theta_i} \log f(X|\boldsymbol{\theta}) = \frac{X_i}{\theta_i}$$

$$\frac{d^2}{d\theta_i^2} \log f(X|\boldsymbol{\theta}) = -\frac{X_i}{\theta_i^2}$$

$$\frac{d^2}{d\theta_i d\theta_j} \log f(X|\boldsymbol{\theta}) = 0 \quad \text{for } i \neq j$$

$$E\left[\frac{d^2}{d\theta_i^2}\log f(X|\boldsymbol{\theta})\right] = -\frac{m}{\theta_i}$$

$$det I(\boldsymbol{\theta}) = \frac{m^k}{\prod_{i=1}^k \theta_i}$$

因此 Jeffreys 先验为 $\pi(\boldsymbol{\theta}) \propto \prod_{i=1}^k \frac{1}{\sqrt{\theta_i}}$ 即为 Dirichlet 分布 $\boldsymbol{\theta} \sim \mathcal{D}\left(\frac{1}{2},\ldots,\frac{1}{2}\right)$.

设 $X \sim N(\mu, 1/\phi)$, 求 μ, ϕ 的 Jeffreys 先验.

↑Example
↓Example

解

$$\frac{d^2}{d\mu^2}\log f(X|\mu,\phi) = -\phi$$

$$\frac{d^2}{d\phi^2}\log f(X|\mu,\phi) = -\frac{1}{2\phi^2} \qquad \text{以及}$$

$$\frac{d^2}{d\mu d\phi}\log f(X|\mu,\phi) = -(X-\mu)$$

$$E[\mathbf{J}] = -\begin{pmatrix} \phi & 0 \\ 0 & \frac{1}{2\phi^2} \end{pmatrix}$$

从而 Jeffreys 先验是 $\pi(\mu, \phi) \propto \frac{1}{\sqrt{\phi}}$.

3.3.3 最大熵先验

Jaynes

基本思想 (Jaynes 1968,1983) 是只有部分信息时候找一个最小的有信息先验分布.

设 θ 为一元离散变量，$p(\theta)$ 为其分布，则称

$$e(p)=-\sum_{i\in\Theta}p\left(\theta_i\right)\log p\left(\theta_i\right)$$

Definition

为其分布的熵 (**entropy**).

如果 $p\left(\theta=\theta_i\right)=1$ 对某个 $\theta_i\in\Theta$, 则, $e(p)=0$, 即为零不确定, 或者最小熵; 相反地, 如果 $p\left(\theta_i\right)=1/|\Theta|$, , 即均匀分布, 则

$$e(p)=-\sum_{i\in\Theta}\frac{1}{|\Theta|}\log\frac{1}{|\Theta|}=\log|\Theta|$$

即为最大熵.

最大熵 (maxent) 分布为最小信息分布.

在很多实际情况中, 我们可能只希望固定先验分布的某些特征, 例如分位数或者矩, 除此之外, 让先验分布尽可能的无信息.

假设我们对参数具有部分信息:

$$E\left[g_k(\theta)\right]=\sum_{i\in\Theta}p\left(\theta_i\right)g_k\left(\theta_i\right)=\mu_k$$

对 $k=1,\ldots,m$. 这包含了固定矩, 即 $g_1(\theta)=\theta$, 以及分位数

$$g_k(\theta)=I_{(-\infty,z_k]}\Rightarrow E\left[g_k(\theta)\right]=p\left(\theta\leq z_k\right)$$

给定这些限定, 下述定理给出最大熵先验的形式.

定理 3. 给定部分信息

$$E\left[g_k(\theta)\right] = \sum_{i\in\Theta} p\left(\theta_i\right) g_k\left(\theta_i\right) = \mu_k$$

对 $k = 1, \ldots, m$, 则最大熵先验为

$$p\left(\theta_i\right) = \frac{\exp\left(\sum_{k=1}^{m} \lambda_k g_k\left(\theta_i\right)\right)}{\sum_{j\in\Theta} \exp\left(\sum_{k=1}^{m} \lambda_k g_k\left(\theta_j\right)\right)}$$

其中 λ_k 可以从信息中确定.

证明参考 Jaynes (1968).

设 $X|N \sim Binom(N, 1/2)$. 假设我们知道 $N \geq 1$ 并且我们固定均值为 $E[N] = 10$. 求 N 的最大熵先验.

↑Example
↓Example

解 由

$$P(N = n) = \frac{\exp(\lambda_1 n)}{\sum_{j=1}^{\infty} \exp(\lambda_1 j)}$$
$$= \exp(\lambda_1 n) \frac{1 - \exp(\lambda_1)}{\exp(\lambda_1)}$$
$$= \left(1 - e^{\lambda_1}\right) \exp(\lambda_1 (n - 1))$$

因此, $N - 1$ 服从以 $1 - e^{\lambda_1}$ 为参数的几何分布, 从而

$$E[N] = 1 + \frac{e^{\lambda_1}}{1 - e^{\lambda_1}}$$

固定 $E[N] = 10$, 我们有 $e^{\lambda_1} = \frac{9}{10}$, 即 N 的最大熵先验是 $N - 1 \sim Ge(9/10)$

连续变量的最大熵先验

连续变量熵的定义

$$e(\pi) = -\int \pi(\theta)\log\pi(\theta)d\mu(\theta)$$

依赖于测度 μ, 因此前述方法推广到连续变量比较复杂. 一种可行的方法 (Jaynes,1968) 是定义

$$e(\pi) = -\int \pi(\theta)\log\frac{\pi(\theta)}{\pi_0(\theta)}d\theta$$

其中 $\pi_0(\theta)$ 为 θ 的 Jeffreys 先验. 则给定限制 $E\left[g_k(\theta)\right] = \mu_k, k = 1, ..., m$. 最大熵先验为

$$\pi(\theta) = \frac{\pi_0(\theta)\exp\left(\sum_{k=1}^m \lambda_k g_k(\theta)\right)}{\int \pi_0(\theta)\exp\left(\sum_{k=1}^m \lambda_k g_k(\theta)\right)d\theta}$$

其中 λ_k 满足 $E\left[g_k(\theta)\right] = \mu_k, k = 1, ..., m$.

但是不幸的是, 最大熵先验对一些情况有可能不存在. 例如如果 $X|\mu \sim N(\mu, 1)$, 固定先验均值 $E[\mu] = m$. 由前可知 Jeffreys 先验为 $\pi(\mu) \propto 1$. 因此最大熵分布为

$$\pi(\mu) = \frac{\exp(\lambda_1 \mu)}{\int_{-\infty}^{\infty} \exp(\lambda_1 \mu) \, d\mu}$$

此积分无解.

3.3.4 Reference 先验

Bernardo

这是一类最一般的方法, 由 Bernardo (1979) 提出发展. 它基于最大化一个专家提供的 θ 的期望信息.

(K-L 距离) 设有两个概率密度函数 $p(x)$ 和 $q(x)$, 它们的 Kullback-Leibler (K-L) 距离定义为

$$KL\big(p(x), q(x)\big) = E_p[\log(p/q)] = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx.$$

Definition

K-L 距离实际上可视为连续型随机变量的熵, 其越大表示先验信息越少.

考虑 θ 为一维情形. $\mathbf{X}$ 为一个大小为 n 的样本, 满足 $X|\theta \sim f(\cdot|\theta)$. 则给定先验分布 $\pi(\theta)$, θ 的期望信息定义为

$$I(\pi(\theta)) = E[KL(\pi(\theta|\mathbf{x}), \pi(\theta))] = \int f(\mathrm{x}) \int \pi(\theta|\mathrm{x}) \log \frac{\pi(\theta|\mathrm{x})}{\pi(\theta)} d\theta d\mathrm{x}$$

其中 $f(\mathrm{x}) = \int f(\mathrm{x}|\theta)\pi(\theta)d\theta$ 以及 $\pi(\theta|\mathrm{x}) = \frac{f(\mathrm{x}|\theta)\pi(\theta)}{f(\mathrm{x})}$.

令 $\mathscr{P}=\{\pi(\theta)>0:\int_\Theta \pi(\theta|\mathbf{x})d\theta<\infty\}$,此处 $\pi(\theta|\mathbf{x})$ 为 θ 的后验分布. 那么, *reference prior* 定义为在先验分布类 $\mathscr{P}$ 中, 能够在 $n\to\infty$ 时候最大化期望信息 $I(\pi(\theta))$ 的那个先验分布. 即若 $\pi^*(\theta)\in\mathscr{P}$, 且满足

$$\pi^*(\theta)=\arg\max_{\pi(\theta),n\to\infty}\left\{I_{\pi(\theta)}(\theta,\mathbf{x})\right\}$$

则 $\pi^*(\theta)$ 为 Reference 先验.

可以证明, 最大熵先验和 Jeffreys 先验是某些特定情形下的 reference 先验.

多元问题下的 Reference 先验

假设 θ 为感兴趣参数, λ 为一个多余参数. 则 Reference 先验可以通过如下两步得到:

- 固定 θ, 利用前述定义计算 reference 先验 $\pi(\lambda|\theta)$.
- 如果是正常分布, 则 λ 可以从 $\mathbf{X}$ 的密度中积分掉, 从而得到 θ 的 Reference 先验. 即

$$p(\mathbf{x}|\theta) = \int p(\mathbf{x}|\theta, \lambda)\pi(\lambda|\theta)d\lambda$$

基于 $p(\mathbf{x}|\theta)$, 利用前述方法得到 θ 的 Reference 先验. 如果不是正常分布, 则可以从极限角度进行此过程. 见 Bernardo(1979).

对多于两个参数的情形, 我们将感兴趣参数按降序排列, 重复利用上述方法.

参考文献

[1] Bayes, T.R. (1763). An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society of London, 53,370-418.

[2] Berger, J. (2006). The case for objective Bayesian analysis (with discussion). Bayesian Analysis, 1, 385482.

[3] Brier, G.W. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78,1-3

[4] Buehler, R. (1971). Measuring information and uncertainty. In Godambe, V. and Sprott, D. eds. Foundations of Statistical Inference. Toronto: Holt, Rinehart and Winston.

[5] Bernardo, J.M. (1979). Reference posterior distributions for Bayesian inference. Journal of the Royal Statistical Society Series B, 41,113-147

[6] Chaloner, K. and Duncan, G.T. (1983). Assessment of a beta distribution: PM elicitation. The Statistician, 32,174-180

[7] Cooke, R.M. (1991). Experts in Uncertainty. New York: Oxford University Press.

[8] Cooke, R.M.., Mendel, M. and Thijs, W. (1988). Calibration and information in expert resolution: a classical approach. Automatica, 24,87-94

[9] DeGroot, M. and Fienberg, S. (1982). Assessing probability assessors: calibration and refinement. Statistical Decision Theory and Related Topics, III, eds. Gupta, S. and Berger, J. New York: Academic Press.

[10] Finetti, R. de (1937). La Prévision: Ses Lois Logiques, Ses Sources Subjectives. Annales de llnstitut Henri Poincaré, 7,1-68.

[11] Fox, B.L. (1966). A Bayesian approach to reliability assessment. Memorandum RM-SO84NASA, The Rand Corporation, Santa Monica, USA.

[12] Garthwaite, P.H.., Kadane, J.B. and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. Journal of the American Statistical Association, 100, 680-701.

[13] Goldstein, M. (2006). Subjective Bayesian analysis: principles and practice (with discussion). Bayesian Analysis, 1, 403420.

[14] Hughes, G. and Madden, L.V. (2002). Some methods for eliciting expert knowledge of plant disease epidemics and their application in cluster sampling for disease incidence. Crop Protection, 21,203-215

[15] Kahneman, D. Slovic, P. and Tversky, A. (1982). Judgment under uncertainty: heuristics and biases. Cambridge: University Press.

[16] Jaynes, E.T. (1968). Prior probabilities. IEEE Trans. on Systems Science and Cybernetics, 4,227-241

[17] Jaynes, E.T. (1983). Papers on Probability, Statistics and Statistical Physics. Ed. R.D. Rosenkrantz. Dordrecht: Reidel.

[18] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society, A. 186,453-.461

[19] Laplace, P.S. (1812). Teoría analitica de las probabilidades. Lichtenstein, S.. Fischoff, B. and Phillips, L. (1982). Calibration of probabilities: the state of the art to 1980. In Kahneman, D.., Slovic, P. and Tversky, A., eds. Judgement under uncertainty: heuristics and biases. Cambridge: University Press.

[20] Kass, R.E. and Wassermann (1996). The selection of prior distributions by formal rules. Journal of the American Statistical Association, 91,1343-1370.

[21] Matheson, J. and Winkler, R.L. (1976). Scoring rules for continuous probability distributions. Management Science, 22,1087-1096

[22] Murphy, A. (1973). A new vector partition of the probability score. Journal of Applied Meteorology, 12,595-600

[23] Price, P.C. (1998). Effects of a relative-frequency elicitation question on likelihood judgment accuracy: The case of external correspondence. Organizational Behavior and Human Decision Processes, 76,277-297

[24] Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory. Boston: Harvard University Press.

[25] Russo, J.E. and Shoemaker, P.J.H. (1989). Decision traps. New York: Simon and Schuster. Savage, L.J. (1971). Elicitation of personal probabilities and expectations. Journal of the American Statistical Association, $66, 781 - 801$

[26] Stone, M. (1982), Review and analysis of some inconsistencies related to improper priors and finite additivity. Logic, Methodology, and Philosophy of Science VI: Proceedings of the Sixth International Congress, Hannover 1979,413-426, Amsterdam: North-Holland.

[27] Tversky, A. and Kahneman, D. (1973). On the psychology of prediction. Psychological Review, 80,237-51

[28] Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. Science, 185,1124-1130

[29] Tversky, A. and Kahneman, D. (1980). Causal Schemas in Judgment Under Uncertainty. In Fischbein, M. ed. Progress in Social Psychology, Hillsdale, NJ: Lawrence Erlbaum, 49-72

[30] Wallsten, T.S., Budescu, D.V., and Zwick, R. (1993). Comparing the Calibration and Coherence of Numerical and Verbal Probability Judgments. Management Science, 39,176-190

[31] Winkler, R.L. (1986). On good probability appraisers. In Goel, P. and Zellner, A. eds. Bayesian Inference and Decision Techniques. New York: Elsevier.

[32] Wiper, M.P. (1987). The expert problem. M.Sc. Dissertation. Department of Statistics, Manchester University.

[33] Wiper, M.P. (1990). Calibration and use of expert probability judgements. Ph.D. Thesis. Department of Computer Studies, Leeds University.

[34] Wiper, M.P., French, S. and Cooke, R. (1994). Hypothesis based calibration scores. The Statistician, 43,231-236

[35] Witteman, C. and Renooij, S. (2003). Evaluation of a verbalnumerical probability scale. International Journal of Approximate Reasoning, 33, 117-131.