第 1 讲：非贝叶斯推断简介

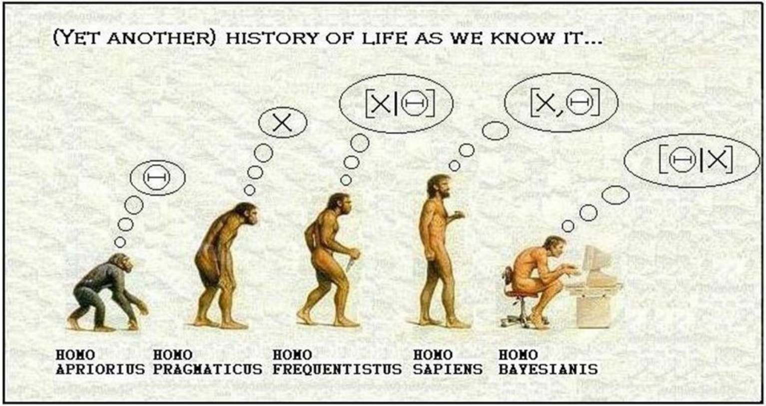
张伟平

部分资料来自于 M.P. Wiper 课件

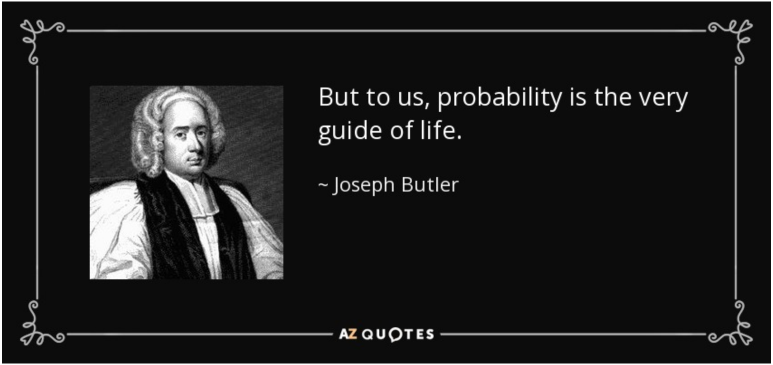
目录

1.1 概率及其解释 2
1.2 统计推断 18
1.2.1 古典推断 19
1.2.2 基于似然的推断 28
1.2.3 信仰推断及其问题 38

1.1 概率及其解释



介绍概率的客观 (Objective) 和主观 (Subjective) 解释, 统计推断的各种非贝叶斯处理及其问题


But to us, probability is the very guide of life.
~ Joseph Butler
AZ QUOTES

概率定义

Kolmogorov

起源于 Fermat 和 Pascal 研究赌博中的机会而发展起来的概率理论，是研究随机性的学科. Kolmogorov (1933) 奠定了严格的数学公理化基础。

The Kolmogorov axioms

记 Ω 为一个随机试验的样本空间, 则一个概率测度 P 为满足以下条件的函数:

1. 对任意事件 $A \in \Omega, P(A) \geq 0$
2. $P(\Omega) = 1$
3. $P\left(\cup_{j \in J} A_j\right) = \sum_{j \in J} P\left(A_j\right)$, 其中 $\{A_j : j \in J\}$ 为不相容事件的可数集.

概率定律都可以基于这个公理导出. 但是, 其仅仅为纯粹的数学定理, 并没有提供关于概率的一个实用解释.

概率的解释

广义上来讲, 概率可能有三种概念:

- 认识论的概念, 旨在衡量客观证据支持关系. 例如,“根据相关的地震和地质数据, 加州可能会在本十年内经历一次大地震”.
- 一种物理概念, 适用于世界上的各种系统, 与任何人的想法无关. 例如,“特定的镭原子可能会衰减 10,000 年”.
- 代理人的信任度（分级信念）的概念. 例如,“我不确定本周在堪培拉会下雨, 但可能会下雨”.

有多种概率的解释方式:

- 古典解释
- 逻辑解释
- 频率解释
- 倾向性解释
- 主观概率

Gillies (2000) 给了详细介绍.

■ 古典解释

Bernoulli

起源于 Jakob Bernoulli (1713) 的想法, Laplace(1814) 考虑了不充分推理原则 (principle of insufficient reason/principle of indifference), 提供了一种赋予认知和主观概率的方式.

The principle of insufficient reason

"If we are ignorant of the ways an event can occur (and therefore have no reason to believe that one way will occur preferentially compared to another), the event will occur equally likely in any way."

- 因此, 概率是感兴趣结果个数与所有可能结果个数之间的系数(古典概型).
- 这是一种非常局限的定义, 不能简单应用于无穷维和连续样本空间.

http://en.wikipedia.org/wiki/Principle_of_indifference

概率的逻辑解释

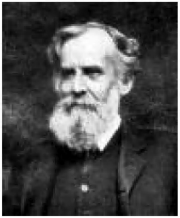
Keynes

Carnap

逻辑概率推广了古典概率的概念, 由 Keynes (1921) 和 Carnap (1950) 提出发展. 一个命题 H 在给定证据 E 下的概率被解释为 E 在逻辑上蕴含 H 的程度. 逻辑概率仅仅依赖两个命题之间的逻辑蕴含关系, 与人的信念程度无关. 例如, "恐龙灭绝 (H) 可能是由于陨星撞击地球 (E) 导致的" , 则 H 由 E 所支持的程度被称为是 H 给定 E 的逻辑概率.

逻辑概率通过使用形式语言 (formal language) 来构造, 具体参考阅读材料。Bernardo and Smith (1994) 指出, 概率的逻辑解释完全缺乏实施措施. 这种逻辑概率是被假设存在且依赖于定义它们的形式语言。

■ 频率解释



Venn

Von Mises

这种观点由 Venn (1876) 提出, von Mises (1919) 进行了阐述. 给定一个可重复的试验, 一个事件的概率是其发生的频率在实验次数趋于无穷时的极限.

这是概率的约束定义, 对于不可重复的试验就不能赋予事件的概率了.

■ 倾向性解释

Popper

Poper(1957) 提出发展了该理论. 概率是单个物体或者整个系统的一种本质的属性, 代表了物理系统具有某种倾向. 比如盐在水中会有溶解的倾向; 硬币被抛后有朝上或者朝下的倾向. 特别地, 长期倾向似乎与概率的频率定义一致, 尽管不清楚单次的倾向是什么, 或者它们是否遵循概率公理.

■ 主观概率解释

Ramsey

概率的主观解释是作为信念的强度，最早由 Ramsey (1926) 讨论. 根据这个定义，概率并不是关于物理系统的，而是关于物理系统和我们之间的关系.

实际上，大多数人是不理性的，因为他们自己的信仰强度不会满足概率公理，见 Kahneman et al (1982) . 因此，为了使主观概率的定义正式化，重要的是仅考虑理性主体，即其信念在逻辑上是一致的主体.

Consistent degrees of belief are probabilities

Cox(1946) 正式提出了一致性推理所需要的逻辑条件。他提出下面假定:

- 首先, 不同命题真实性的相对信念是可传递的, 即如果我们认为对三个事件 A, B, C, 有 $A \succeq B$ 和 $B \succeq C$, 那么我们必须相信 $A \succeq C$, 其中 $A \succeq B$ 表示 A 至少与 B 发生的可能性相同。这个假设意味着我们可以用数字表示信仰强度, 其值越高, 信念强度越大.

- 其次, 如果我们指定了我们相信事件 A 为真的信念强度, 那么我们也暗含指定了我们相信事件 A 为假的信念强度.

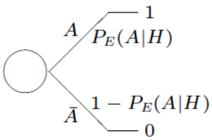
- 最后, 如果我们指定我们相信 A 是真的信念强度大小, 以及在给定 A 为真条件下, 我们相信假设 B 为真的信念强度大小, 那么就暗含了我们相信 A 和 B 都为真的信念强度.

给定这些公理, Cox 证明了这种逻辑一致性只有在用于表示信仰程度的数值是概率时候才成立.

定义主观概率

Cox 的方法是非构造性的, 他没有给出如何定义一个给定事件的概率. 定义概率的方法有多种, 大多是基于打赌的思想.

De Finetti (1937) 定义代理人 E 在信息 H 下对事件 A 的主观概率, $P_E(A|H)$, 为 E 在博彩中当且仅当 A 发生他才可能得到一个单位 (少量) 奖金情况下愿意购买一张彩票的最大付出 (公平价格).



期望收益为

$$P_E(A|H) \times 1 + (1 - P_E(A|H)) \times 0 = P_E(A|H)$$

一个荷兰赌是一系列下注，每个下注都可以被代理人所接受，但是实际情况表明所有下注一起会使得代理人肯定会输钱．可以证明，为了避免荷兰赌，则代理人的概率必须满足概率公理．

需要注意的是，De Finetti 提出的定义依赖于假定代理人对金钱的效用是线性的．Savage（1954）提供了定义概率和效用的更一般方法．O’Hagan(1988) 第 1 至第 3 章基于博彩和赔率定义主观概率。

概率的解释总结

A summary of some interpretations of probability [2]

	Classical	Frequentist	Subjective	Propensity
Main hypothesis	Principle of indifference	Frequency of occurrence	Degree of belief	Degree of causal connection
Conceptual basis	Hypothetical symmetry	Past data and reference class	Knowledge and intuition	Present state of system
Conceptual approach	Conjectural	Empirical	Subjective	Metaphysical
Single case possible	Yes	No	Yes	Yes
Precise	Yes	No	No	Yes
Problems	Ambiguity in principle of indifference	Circular definition	Reference class problem	Disputed concept

1.2 统计推断

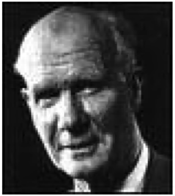
基于频率和主观概率, 很多统计推断方法已经被提出.

- 古典 (频率) 推断
- 基于似然的方法
- 信仰统计及相关方法
- 贝叶斯推断

Barnett (1999) 详细对比了不同方法.

1.2.1 古典推断

Neyman



Pearson

这种推断方法由 Neyman and Pearson (1933) 和 Fisher (1925) 的思想发展起来.

古典推断的特点

- 概率的频率解释
- 推断基于似然函数 $L(\theta|x) = f(x|\theta)$
- 给定 θ, 我们只能量化 (a priori)X 的不确定性
- 推断过程基于渐近性能:
  - 定义估计量 $t = t(X)$
  - $\theta = \theta_0$ 的合理性由似然 (密度) 来度量

$$L(\theta_0|x) \propto f(t(x)|\theta_0)$$

(假设 $t(\cdot)$ 是充分统计量)

  - 如果 t 不在 $f(t|\theta_0)$ 的尾部, 则 θ_0 是 θ 的一个合理值.

- 古典点估计基于选择有良好渐近性质 (无偏性, 方差最小, 效率等) 的估计量

- 有多种选择估计量的方法, 例如矩估计, 最大似然方法等

- 区间估计被定义为对指定的 α, 满足

$$P(l(X) < \theta < u(X)|\theta) = 1 - \alpha$$

的区间 $(l(X), u(X))$

- 假设检验则基于

$$P\left(t(X) > t(x)|\theta_0\right) < \alpha$$

在水平 α 下拒绝 θ_0

充分性原则

一个统计量 $t=t(x)$ 称为是 θ 的充分统计量，如果

$$f(x|t) \text{与 } \theta \text{ 无关}$$

Definition

Fisher(1922) 的充分性原则 (sufficiency principle) 如下:

如果一个充分统计量, t, 存在, 且对两个同样大小的样本 x_1, x_2 满足 $t(x_1)=t(x_2)$ 时, 则给定 x_1 和 x_2 下的结论是相同的.

所有推断标准方法都满足该原则.

因子分解定理

定理 1. 一个统计量 t 是 θ 的充分统计量当且仅当存在函数 g 和 h

使得

$$f(x|\theta) = g(t,\theta)h(x)$$

重复抽样原则

从样本 x 中得出的推断应该基于结论如何随着抽样样本变化而改变的分析基础上, 这可以通过对产生样本 x 的试验在完全相同条件下假设重复而得到.

这个原则直接来自于概率的频率定义且比较有争议性. 它意味着不确定性的度量仅仅是假设重复试验下的渐近频率, 从而, 不可能度量给定 x 下,θ 的后验不确定性.

古典推断的批评

- 整体批评:

  概率的频率定义限制了可以被合理分析的问题范围. 进一步, 古典推断理论像是由特定统计过程所组成的一本指南.

- 特定批评

  首先, 对于估计存在很多问题. 选择一个估计量的最优方法一般不存在, 一些常用的方法可能给出的是很差的估计量. 例如

  – 设 $X \sim Pois(\lambda)$, 则 $\phi = e^{-2\lambda}$ 的无偏估计存在且为 $\hat{\phi} = (-1)^X = \pm 1$, 然而 $0 < \phi \leq 1$.

  – 又如估计柯西分布位置参数 θ, 矩估计量不存在. 此外, 矩估计量没有最大似然估计的最优性.

  – 最大似然估计也不是完美的, 如果 $X \sim U(1, \theta)$, 则样本量为 1 时 θ 的最大似然估计为 $\hat{\theta} = X$, 其偏差 $EX - \theta =$

$(1-\theta)/2$. 当 θ 较大时偏差程度可以很大. 不管样本量多大, 最大似然估计都不是无偏估计.

– 又如 $Y \sim N(\theta, \sigma^2 I_n)$, 则 θ 的最大似然估计 $\hat{\theta} = y$(最小二乘估计). 但是当 $n \geq 3$ 时, $\hat{\theta}$ 是不可容许的, 因为 James-Stein 估计量 $(1-(n-2)\sigma^2/\|y\|^2)y$ 的均方误差更小.

– 更重要的是可解释性问题. 例如我们在一个试验下得到参数的 95% 置信区间值为 (1,3), 我们该如何解释?

– 古典推断在存在多余参数时候常出问题. 设 $Y_{ij} \sim N(\phi_i, \sigma^2), i = 1, \ldots, n; j = 1, 2$. 感兴趣参数为 σ^2, $\phi = (\phi_1, \ldots, \phi_n)$ 为多余参数. 似然函数为

$$L(\sigma^2, \phi|y) \propto \sigma^{-2n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_{i,1}-\phi_i)^2 + (y_{i,2}-\phi_i)^2\right)$$

估计 σ^2 的方法常常是 profile(剖面) 似然方法:

首先假设 σ^2 已知, 求 ϕ 的最大似然估计

$$
\begin{aligned}
L_P\left(\sigma^2|\mathbf{y}\right) &= \sup_{\phi} L\left(\sigma^2, \phi|\mathbf{y}\right) \\
&= \sigma^{-2n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\left(y_{i,1}-\bar{y}_i\right)^2+\left(y_{i,2}-\bar{y}_i\right)^2\right]\right) \\
&= \sigma^{-2n} \exp\left(-\frac{1}{4\sigma^2}\sum_{i=1}^{n}\left(y_{i,1}-y_{i,2}\right)^2\right)
\end{aligned}
$$

其次, 最大化 σ^2 的 profile 似然得到

$$
\hat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^{n}\left(y_{i,1}-y_{i,2}\right)^2
$$

但是

$$
E\left[\hat{\sigma}^2\right] = \frac{\sigma^2}{2}
$$

对任何 n, $\hat{\sigma}^2$ 都是不相合的.

– 不完全清楚如何预测. 例如 $X \sim f(\cdot|\theta)$. 一般来说, 给定样本, 使用 θ 的最大似然估计代替 θ 后得到估计的预测分布 $f(x|\hat{\theta})$. 但是, 这个过程明显低估了预测不确定性, 因此 θ 是未知的.

1.2.2　基于似然的推断

Barnard

这种方法完全建立在似然函数之上, 由 Barnard (1949) 和 Barnard et al (1962) 提出. 首先, 假设似然原则成立. Edwards (1992) 3.3 节定义似然原则如下:

似然原则

样本 X 的中关于 θ 的所有相关信息都包含在观测值 x 下的似然函数中. 进一步, 如果依赖于 θ 的两个不同样本 X_1 和 X_2 的似然函数满足

$$L(\theta|X_1) = cL(\theta|X_2), \forall\theta$$

则它们包含了关于 θ 的相同信息.

也就是说,

- 如果我们有两个试验, 即样本 X_i 和抽样分布 $f_i(\cdot|\theta)$ 下的试验 E_i, $i = 1, 2$. 参数空间相同. 那么如果 $f_1(X_1|\theta, E_1) \propto f_2(X_2|\theta, E_2)$, 则两个试验提供了相同的证据 (EV), 从而关于 θ

的推断是相同的. 即

$$L(\theta|E_2, X_2) = cL(\theta|E_1, X_1) \text{ for some } c \Leftrightarrow EV[E_1, X_1] = EV[E_2, X_2]$$

- 对任意两个值 θ_1 和 θ_2, 似然比 $L(\theta_1|x)/L(\theta_2|x)$ 为数据支持 θ_1 相对于 θ_2 证据的一种度量. 但是这在实际中有些场合很难应用, 特别是存在多元参数的场合. 例如, 我们怎样边际化似然函数?

停止规则原则

在一个试验序列中, 试验所提供的关于未知参数 θ 的证据不能依赖于停止规则.

- 停止规则在经典统计中常用来在结果充分好或者不好时候提前停止试验.
- 频率统计学家必须在试验前选择好停止准则, 然后严格按照停止规则进行 (否则检验将失去优良的频率性质). 比如要检验一个硬币是否是均匀的, 那么可以考虑多种停止规则, 例如抛一个硬币指定次; 抛一个硬币直至首次出现反面; 抛一个硬币直至第 r 次出现反面.
- 停止规则原则显然是似然原则下的结果, 因为似然函数独立于停止规则. 但是, 经典的假设检验并不满足停止规则原则. 例

如, 假设 $\theta = P(\text{ head })$. 我们要在 0.05 水平下检验 $H_0 : \theta = 1/2 \Leftrightarrow H_1 : \theta > 1/2$.

假设我们观察到 9 个正面和 3 个反面. 此信息对我们来说不足写出似然函数, 还需要知道抽样机制或停止规则. 假设我们固定抛掷次数为 12 次, 则似然函数为

$$L(\theta|x=9) = \begin{pmatrix} 12 \\ 9 \end{pmatrix} \theta^9 (1-\theta)^3$$

从而 p 值为

$$P\left(X \geq 9|H_0\right) = \sum_{x=9}^{12} \begin{pmatrix} 12 \\ x \end{pmatrix} (1/2)^x (1-1/2)^{12-x} = 0.073$$

因此我们不能拒绝原假设.

假设我们抛掷硬币直至出现第三次反面. 则由负二项分布知道

似然函数

$$L(\theta|x) = \begin{pmatrix} 11 \\ 9 \end{pmatrix} \theta^9 (1-\theta)^3$$

p 值为

$$P(X \geq 9|H_0) = \sum_{x=9}^{\infty} \begin{pmatrix} 3+x-1 \\ 2 \end{pmatrix} (1/2)^x (1/2)^3 = 0.0327$$

从而我们拒绝了原假设.

按照似然原理, 两个试验的结论应该是一致的. 但是经典检验方法的结论不一致. 原因在于为了进行经典假设检验, 我们必须指定样本空间或者停止规则. 两种试验下样本空间不同:

1. $\Omega = \{(u,d) : u+d = 12\}$
2. $\Omega = \{(u,d) : d = 3\}$

条件性原则

假设我们可以进行两个试验 E_1 和 E_2 来推断 θ, 选择哪个试验通过抛硬币决定. 那么对 θ 的推断应该仅依赖于选择的试验.

为了正式表示, 记试验为 $E_i = (\mathrm{x}_i, \theta, f_i)$, 表示在试验 E_i 下, $\mathrm{x}_i \sim f_i(\cdot|\theta)$. 现在定义复合试验 $E^\star = (\underbrace{(K, \mathbf{x}_K)}_{\mathbf{x}}, \boldsymbol{\theta}, \frac{1}{2} f_K(\mathbf{x}_K))$, 其中 K 服从 1 和 2 上的两点均匀分布. 条件性意味着

$$EV[E^\star, \mathbf{x}] = \begin{cases} EV[E_1, \mathbf{x}_1] & \text{if } K = 1 \text{ so } \mathbf{x} = (1, \mathbf{x}_1) \\ EV[E_2, \mathbf{x}_2] & \text{if } K = 2 \text{ so } \mathbf{x} = (2, \mathbf{x}_2) \end{cases}$$

Birnbaum (1962) 证明了下述定理

定理 2. *似然原则等于充分性原则加上条件性原则.*

我们仅证明充分性原则加上条件性原则可以导出似然原则.

证明. 记 E_1, E_2 分别为两个试验, 定义复合试验 E^* 同条件性原则里定义. 考虑 x_1^0 和 x_2^0 满足:

$$L\left(\boldsymbol{\theta}|E_1,\mathbf{x}_1^0\right)=cL\left(\boldsymbol{\theta}|E_2,\mathbf{x}_2^0\right)$$

对某个常数 c 成立. 似然原理意味着

$$EV(E_1,\mathrm{x}_1^0)=EV(E_2,\mathrm{x}_2^0). \tag{1.1}$$

对混合试验 E^*, 定义统计量 T

$$\mathrm{T}=\mathrm{t}(k,\mathrm{x}_k)=\begin{cases}(1,\mathrm{x}_1^0) & \text{if } k=2,\mathrm{x}_2=\mathrm{x}_2^0\\ (k,\mathrm{x}_k) & \text{otherwise}\end{cases}$$

则 T 对 $(1,\mathrm{x}_1^0)$ 和 $(2,\mathrm{x}_2^0)$ 取相同值. 那么统计量 T 是充分统计量. 事实上, 当 $\mathbf{t}\neq(1,\mathrm{x}_1^0)$ 时

$$P_\theta(\mathrm{x}^*=(k,\mathrm{x}_k)|T=t)=\begin{cases}1 & \text{if } \mathrm{t}=(k,\mathrm{x}_k)\\ 0 & \text{otherwise}\end{cases}$$

而当 $\mathbf{t}=(1,\mathrm{x}_1^0)$ 时

$$P_\theta(\mathrm{x}^*=(1,\mathrm{x}_1^0)|T=(1,\mathrm{x}_1^0))=\frac{\frac{1}{2}f_1\left(\mathbf{x}_1^0|\theta\right)}{\frac{1}{2}f_1\left(\mathbf{x}_1^0|\theta\right)+\frac{1}{2}f_2\left(\mathbf{x}_2^0|\theta\right)}=\frac{c}{1+c}$$

当 $t=(1,\mathrm{x}_1^0)$, 但是 $\mathrm{x}=(2,\mathrm{x}_2)$ 时

$$P_\theta(\mathrm{x}^*=(2,\mathrm{x}_2)|T=(1,\mathrm{x}_1^0))=\frac{\frac{1}{2}f_2\left(\mathrm{x}_2^0|\theta\right)}{\frac{1}{2}f_1\left(\mathrm{x}_1^0|\theta\right)+\frac{1}{2}f_2\left(\mathrm{x}_2^0|\theta\right)}=\frac{1}{1+c}$$

因此 $f(\mathrm{x}^*|t,\theta)=f(\mathrm{x}|\mathrm{t})$ 故 T 是充分统计量.

由充分性原则,$T((1,\mathrm{x}_1^0))=T((2,\mathrm{x}_2^0))$, 故

$$EV\left[E^\star,\left(1,\mathbf{x}_1^0\right)\right]=EV\left[E^\star,\left(2,\mathbf{x}_2^0\right)\right] \tag{1.2}$$

再由条件性原则得到

$$EV\left[E^\star,\mathbf{x}_1^0\right]=EV\left[E_1,\left(1,\mathbf{x}_1^0\right)\right]=EV\left[E^\star,\left(2,\mathbf{x}_2^0\right)\right]=EV\left[E_2,\mathbf{x}_2^0\right] \tag{1.3}$$

由此得证 (1.1), 这就是似然原则.

类似的, 可以证明似然原则 + 充分性原则 $\Rightarrow$ 条件性原则; 似然原则 + 条件性原则 $\Rightarrow$ 充分性原则. $\square$

详细证明可以参看 Birnbaum (1962).

1.2.3　信仰推断及其问题

Fisher

信仰推断的目的在于给出参数 θ 的不确定性的后验度量, 而无需先验度量. 这种推断方法由 Fisher(1930) 最早提出.

假设 $X|\mu,\sigma^2 \sim \mathcal{N}\left(\mu,\sigma^2\right)$. μ 为感兴趣参数. 由于 $T = \frac{(\bar{X}-\mu)\sqrt{n}}{S} \sim t_{n-1}$, 则对任何 $t, P(T>t) = p(t)$, 其中 $p(t)$ 已知. Fisher 的想法是

将其表示为

$$\begin{aligned} p(t) &= P(T > t) \\ &= P\left(\frac{(\bar{X}-\mu)\sqrt{n}}{S} > t\right) \\ &= P\left(\mu < \bar{X} - \frac{St}{\sqrt{n}}\right) \end{aligned}$$

然后定义 $p(t) = P\left(\mu < \bar{x} - \frac{st}{\sqrt{n}}\right)$ 为 μ 不超过 $\bar{x} - \frac{st}{\sqrt{n}}$ 的信仰概率 (fiducial probability).

信仰推断的问题在于

- 概率测度从样本空间转换到参数空间, 合理性是什么?
- 如果没有轴统计量存在的话, 怎么办?
- 如何把信仰推断方法应用到高维问题尚不清楚

很多时候, 信仰概率区间与特定无信息先验下的贝叶斯信任区间相同, 参见 Fraser(1968). 但是贝叶斯信任区间更加合理.

参考文献

[1] Barnard, G.A. (1949). Statistical Inference, Journal of the Royal Statistical Society. Series B, 11,115-149

[2] Barnard, G.A., Jenkins, G.M. and Winsten, C.B. (1962) Likelihood Inference and Time Series. Journal of the Royal Statistical Society, Series A, 125,321-372

[3] Bernardo, J.M. and Smith, A.F.M. (1994).

[4] Bayesian Theory. New York: Wiley.

[5] Barnett, V. (1999) . Comparative Statistical Inference (3rd ed.). Chichester: Wiley

[6] Bernoulli, J. (1713) . Ars conjectandi, opus posthumum. Accedit Tractatus de seriebus infinitis, et epistola gallicé scripta de ludo pilae reticularis. Basel: Thurneysen Brothers.

[7] Birnbaum, A. (1962). On the foundations of statistical inference (with discussion). Journal of the American Statistical Association, 57,269-306

[8] Carnap, R. (1950). Logical Foundations of Probability. University of Chicago Press.

[9] Cox, R. (1946). Probability, Frequency, and Reasonable Expectation, A m . Jour. Phys. 14,1-13

[10] Edwards, A.W.F. (1992). Likelihood (2nd ed.) Cambridge: University Press.

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

[12] Fisher, R.A. (1922). On the Mathematical Foundations Of Theoretical Statistics. Philosophical Transactions of the Royal Society of London. Series A, 222,309-68

[13] Fisher, R.A. (1925). Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.

[14] Fisher, R.A. (1930). Inverse probability. Proceedings of the Cambridge Philosophical Society, 26,528-535

[15] Fraser, D.A.S. (1968). The Structure of Inference. New York: Wiley.

[16] Gillies, D. (2000) . Philosophical theories of probability. Routledge.

[17] Hájek, A. (2003). Interpretations of Probability. In Stanford Encyclopedia of Philosophy.

[18] Kahneman, D., Slovic, P. and Tversky, A. (1982). Judgement under Uncertainty:Heuristics and Biases. Cambridge: University Press.

[19] Keynes, J.M. (1921). A Treatise on Probability. London: Macmillan.

[20] Kolmogorov, A. (1933) . Grundbegriffe der Wahrscheinlichkeitsrechnung. Berlin: Springer.

[21] Laplace, P.S. (1814). Théorie Analytique des probabilités. Paris: Courcier Imprimeur.

[22] Mises, R. von (1919). Wahrscheinlichkeit, Statistik und Wahrheit. Vienna, Springer.

[23] Neyman, J. and Pearson, E. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses, Philosophical Transactions of the Royal Society of London. Series A, 231,289-337

[24] O' Hagan, A. (1988). Probability: Methods and Measurement. London: Chapman and Hall.

[25] Pederson, J.G. (1978). Fiducial Inference. International Statistical Review, 46,147-170.

[26] Popper, K. (1957). The Propensity Interpretation of the Calculus of Probability and of the Quantum Theory. In Observation and Interpretation. Buttersworth Scientific Publications.

[27] Ramsey, F.P. (1926). Truth and Probability. In Foundations of Mathematics and other Essays. Routledge and Kegan.

[28] Savage, L.J. (1954). The Foundations of Statistics. New York: Wiley.

[29] Venn, J. (1876). The Logic of Chance: An Essay on the Foundations and Province of the Theory of Probability.