教书育人、不忘初心,《数理统计》本科课程

— 课程思政 20 讲

田国梁 统计学教授

南方科技大学 • 统计与数据科学系

Email: tiangl@sustech.edu.cn

中国 • 广东 • 深圳

2024 年 03 月 18 日

提纲 (Outline) Part I

第 1 讲 **South**与**Southern**之区别

第 2 讲 用**14 年创新编写**《数理统计》英文教材

第 3 讲 **Bayes**如何译成中文名？英文名如何读？

第 4 讲 全概率公式 (**Law of Total Probability**) 和 Bayes 公式

第 5 讲 自然常数 (**Natural Constant**) $\mathrm{e} = 2.718282\cdots$ 的起源

第 6 讲 从矩母函数与密度函数的关系出发，深度理解**国王函数** e^x

第 7 讲 从对数似然函数出发，深度理解**王后函数** $\log(x)$

第 8 讲 **标准正态分布密度**和**蛇吞象公式**

第 9 讲 从函数的**一阶泰勒展开式**到**线段中/外**任何一点之数学表达式

第 10 讲 函数的一阶泰勒展开之**四种形式**

提纲 (Outline) Part II

- 第 11 讲 指数分布与几何分布的无记忆性
- 第 12 讲 二项分布的生存函数与贝塔分布的累积分布函数之恒等式
- 第 13 讲 深度理解中心极限定理
- 第 14 讲 二项分布的正态近似和泊松近似
- 第 15 讲 从矩估计量到Monte Carlo 积分
- 第 16 讲 从KL 散度的角度来理解极大似然估计之定义
- 第 17 讲 从Laplace提出问题到Gauss解决问题: 正态分布的发现过程
- 第 18 讲 度量点估计量好坏的指标: 均方误差
- 第 19 讲 克拉默–拉奥 (Cramér–Rao) 不等式
- 第 20 讲 建立参数的置信区间过程中的枢轴量(Pivotal Quantity)

第 7 讲 从对数似然函数出发，深度理解王后函数 $\log(x)$

南科夜色美如画，会议中心灯火明

7.1 似然函数与对数似然函数

1° 联合密度与似然函数

- 设 $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x;\boldsymbol{\theta})$ 且 $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ 是 $\mathbf{x} = (X_1, \ldots, X_n)^\top$ 的观察值, 则 $\mathbf{x}$ 的联合密度为 $f(\boldsymbol{x};\boldsymbol{\theta})$, $\boldsymbol{\theta}$ 的似然函数为

$$L(\boldsymbol{\theta}) = f(\boldsymbol{x};\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i;\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

2° 对数似然函数与极大似然估计

- 定义 $\ell(\boldsymbol{\theta}) \doteq \log\{L(\boldsymbol{\theta})\} = \sum_{i=1}^n \log\{f(x_i;\boldsymbol{\theta})\}$.
- 极大似然估计 $\boldsymbol{\theta}^*$ 定义为满足: $L(\boldsymbol{\theta}^*) = \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L(\boldsymbol{\theta})$, 可重写为

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell(\boldsymbol{\theta}).$$

— **优点**: (1) $\log(\cdot)$ 是单调增加函数; (2) $\log(\cdot)$ 变换将乘法变为加法。

— **启发**: 在地球上难做的事, 搬到月亮上去做, 就容易一些。

7.2 函数的位阶或分层

$\int G_1(x)\,\mathrm{d}x \triangleq G_2(x)$	$\int g(x)\,\mathrm{d}x \triangleq G_1(x)$	$g(x)$	$g'(x)$	$g''(x)$
外空 祖爷爷	月亮 爷爷	地球 父亲	近海 儿子	深海 孙子
e^x	e^x	$\mathrm{e}^x,\ (x\in\mathbb{R})$	e^x	e^x
$\int \log(x)\,\mathrm{d}x = x\log(x) - x$	$\log(x)$	$x^{-1},\ (x>0)$	$-x^{-2}$	$2x^{-3}$
$\frac{1}{12}x^4$	$\frac{1}{3}x^3$	$x^2,\ (x\in\mathbb{R})$	$2x$	2
$[\log(b)]^{-2}b^x$	$[\log(b)]^{-1}b^x$	$b^x,\ (x>0)$	$[\log(b)]\,b^x$	$[\log(b)]^2 b^x$

注：$\mathbb{R}\triangleq(-\infty,\infty)$，$b>0$，$b\neq 1$，$b\neq \mathrm{e}$

7.3 为什么 $\log(x)$ 被称为王后函数 (Queen Function)

3° 将 x^{-1} 当作是地球上的函数，它的后代无穷无尽，为公民函数

- x^{-1} 的 n 阶导数为 $(-1)^n n! x^{-n-1}$，属于公民函数类 $\{x^r\}$。

4° 从 x^{-1} $(x > 0)$ 的积分 $\Rightarrow$ $\log(\cdot)$ 是一个月亮上的函数

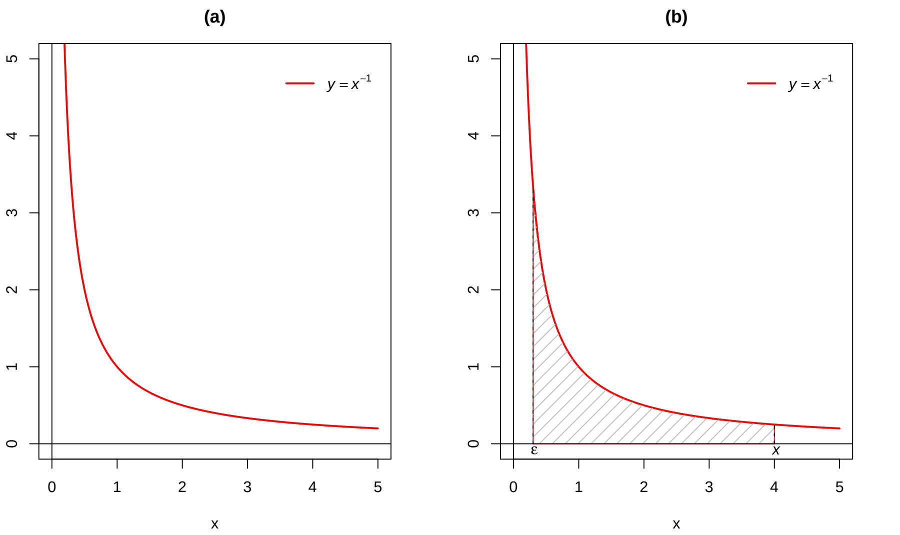
- $\int_0^{+\infty} x^{-1}\,\mathrm{d}x = +\infty$，即在月亮上 **不存在**。但是对任意小的 $\varepsilon > 0$，$\int_\varepsilon^x y^{-1}\,\mathrm{d}y < +\infty$，即在月亮上**存在，但不知道等于多少**。见图 1。
- 数学家将它记为：$\log(x) - \log(\varepsilon) = \int_{\varepsilon_0}^x y^{-1}\,\mathrm{d}y$，或

$$\log(x) = \int x^{-1}\,\mathrm{d}x. \tag{SZ7.1}$$

- 数学家经常将一个存在，但算不出来的积分，取一个名，例如：

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \mathrm{e}^{-x}\,\mathrm{d}x \quad \text{和} \quad B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x.$$

5• 图 7.1: (a) $\int_0^{+\infty} x^{-1}\,\mathrm{d}x = +\infty$; (b) $\int_\varepsilon^x y^{-1}\,\mathrm{d}y < +\infty$



$6^{\bullet}$ $\log(\cdot)$ 被称为王后函数的理由

- 从公式 (SZ7.1), 我们知道, 从 x^{-1} 变到 $\log(x)$, 函数名变了, 是一种突变。
- 西方女子出嫁后, 自己的姓要改为丈夫的姓。香港折中: **林郑月娥**
- 经过**若干代**的不懈努力, 终于从公民函数x^{-1}, 变为王后函数$\log(x)$。
- $\log(x)$ 是所有数学函数中的第二美的函数。
- $\log(x)$ 与 e^x 互为反/逆函数:

$$y = \mathrm{e}^x \Leftrightarrow x = \log(y).$$

7.4 所包含的思政元素

- **渐变与突变的统一**: 渐变与突变是事物发展的两种形式。渐变分为**量的渐变**和**质的渐变**。前者是指量的增加或减少，也就是**量变**；后者是指新质的逐渐积累和旧质的逐渐衰减，也是质变的一种形式。**突变**则是质变的另一种形式，是渐进过程的中断，是不经过任何过渡阶段从一种质态到另一质态的飞跃。
- **积小胜为大胜，积跬步至千里**: 经过努力奋斗与长期积累，汇聚事业发展的磅礴力量，终会成功。不得“三天打鱼，两天晒网”。
- **一种创造函数的方法**: 数学家经常将一个存在，但算不出来的积分，取一个名。
- **一种解决困难问题的思维方式**: 在地球上难做的事 (例如乘法和除法)，搬到月亮 (作对数变换) 上去做，就容易一些 (变为加法和减法)。