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 05 月 06 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**) $e = 2.718282\cdots$ 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** $\log(x)$

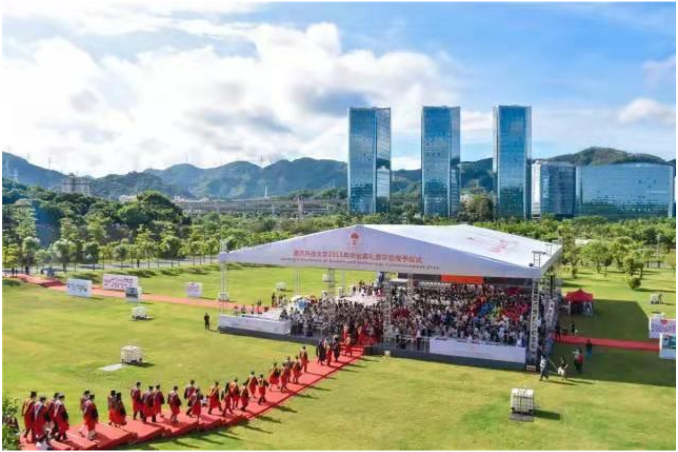
第 8 讲 **标准正态分布密度**和**蛇吞象公式**

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

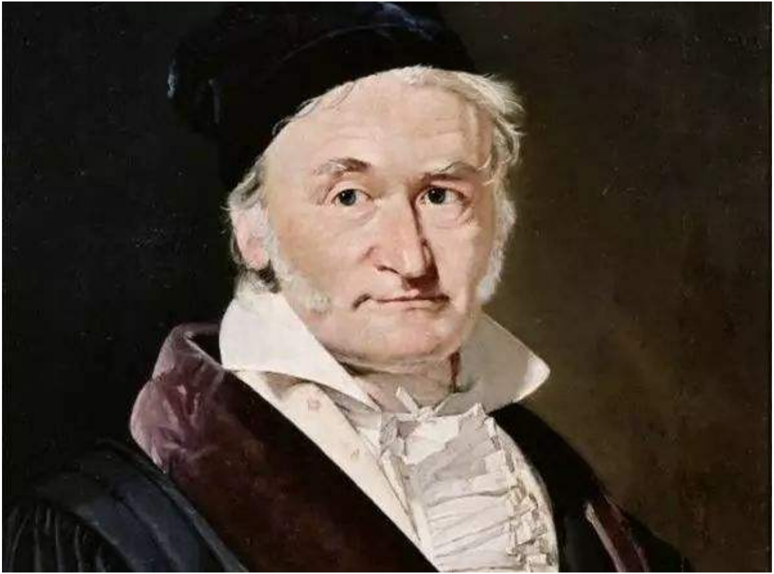
第 10 讲 函数的一阶泰勒展开之**四种形式**

提纲 (Outline) Part II

第 17 讲 从 Laplace 提出问题到 Gauss 解决问题: 正态分布的发现过程



蓝天白云智园楼, 绿草红毯典礼台



Johann Carl Friedrich Gauss (1777–1855): 德国数学家

17.1 μ 的 MLE 和矩估计量 (Moment Estimator)

$1^\bullet$ 正态分布的密度函数

- 设随机变量$X \sim N(\mu, \sigma^2)$, 则 X 的密度函数为

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty.$$

$2^\bullet$ μ 的 MLE 和矩估计量

- 设$X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ 且 $x_1, \ldots, x_n$ 是它们的实现值 (Realizations), 则 (μ, σ^2) 的似然函数为

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}.$$

- μ 的 MLE 和矩估计量均为$\hat{\mu} = \bar{X} = (1/n)\sum_{i=1}^{n} X_i$.

17.2 反问题是什么？如何提出？

- 如果μ的 MLE 是$\bar{X}$, 我们能否声称 (Claim): 该密度函数一定是**正态分布的密度函数**?

3° 高斯当时所面临的问题

- 为了测量行星与地球的真实距离 (μ)，在 17 世纪，第谷・布拉赫 (Tycho Brahe) 提倡采用重复测量，例如 $X_1, \ldots, X_n$。
- 当天文学家使用平均观测值 (**Average Observations**, 即$\bar{X}$) 来估计μ时，他们想知道误差的概率分布 (**Probability Distribution of Errors**) 是什么。

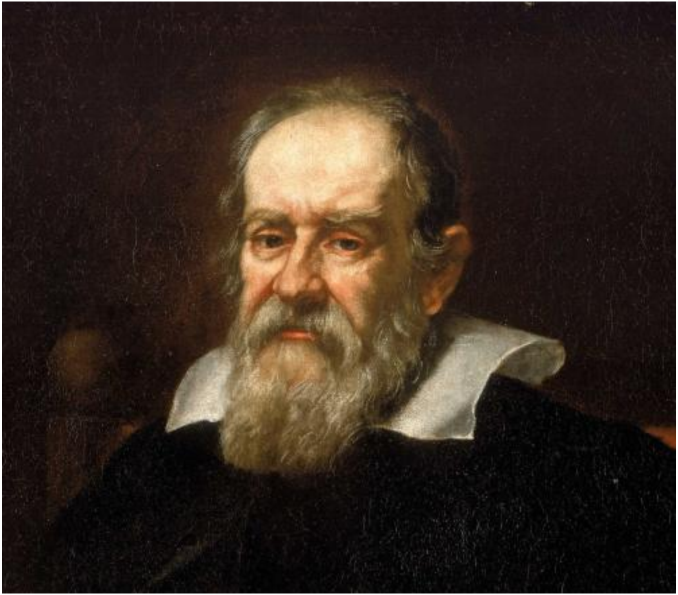
17.3 关于误差 (或误差曲线) 的概率分布的早期推理

4° 在 1632 年，伽利略 (Galileo) 提出：

- 存在一个真正的距离μ，它只是一个数字；
- 所有的观测$\{x_i\}_{i=1}^n$都有误差，用$\{x_i-\mu\}_{i=1}^n$来表示；
- 误差是关于真值 (True Value) 对称的 (Symmetric)；
- 小误差 (Small Errors) 比大误差更常见。

5° 在 1756 年，Simpson 提出了误差的概率分布

- 在 1756 年，Simpson 提出了误差 (或误差曲线) 的概率分布的**概念**。



Galileo (1564–1642): 意大利天文学家

6° 在 1774 年，拉普拉斯提出了 The First Law of Errors

- 第一定律指出，误差的频率 (Frequency of an Error) 可以表示为误差数值大小 (Numerical Magnitude of The Error) 的指数函数，而不考虑其符号:

$$f(x) = \frac{1}{2\sigma} e^{-|x/\sigma|}, \quad -\infty < x < \infty.$$

- 或者等价地, 不考虑其符号, 误差频率的对数 (Logarithm of the Frequency of an Error) 是误差的线性函数:

$$\log f(x) = c - |x/\sigma|.$$

Pierre-Simon de Laplace (1749–1827):

法国天体机械师

17.5 拉普拉斯 (Laplace) 第二误差律

7° 在 1778 年，拉普拉斯提出了 The Second Law of Errors

- 第二定律指出，误差的频率 (Frequency of the Error) 是误差平方 (Square of the Error) 的指数函数，或者等价地，误差频率的对数是误差的二次函数 (Quadratic Function of the Error):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-x^2/(2\sigma^2)}, \quad 或 \quad \log f(x) = c - x^2/(2\sigma^2).$$

- 拉普拉斯第二定律通常被称为正态分布 (**Normal Distribution**) 或高斯律 (**Gauss Law**)。

17.6 假设和高斯的证明

8° 基本的假设

(1) 用$f(x-\mu)$表示的误差密度函数 (Error Density Function) 是连续的且可微的; 即$f'(\cdot)$ 存在。

(2) 设$X_1,\ldots,X_n \overset{\text{iid}}{\sim} f(x-\mu)$ 且$\{x_i\}_{i=1}^n$ 是$\{X_i\}_{i=1}^n$的实现值。

(3) 设 μ 的 MLE 是$\bar{X}=(1/n)\sum_{i=1}^n X_i$。

(4) 误差密度函数是两次可微的; 即 $f''(\cdot)$ 存在。

9° 在假设 (1)–(3) 成立下，一个基本的结果

- μ 的似然函数由下式给出

$$L(\mu) = \prod_{i=1}^{n} f(x_i - \mu)$$

使得对数似然函数为

$$\ell(\mu) = \log L(\mu) = \sum_{i=1}^{n} \log f(x_i - \mu).$$

- 令$0 = \ell'(\mu) = \mathrm{d}\ell(\mu)/\mathrm{d}\mu$ 且定义 $g(t) = f'(t)/f(t)$, 我们得到

$$0 = \sum_{i=1}^{n} \frac{-f'(x_i - \mu)}{f(x_i - \mu)} \quad \Longrightarrow \quad 0 = \sum_{i=1}^{n} g(x_i - \mu). \tag{17.1}$$

- 因为 μ 的 mle 是 $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$, 方程 (17.1) 成为

$$0 = \sum_{i=1}^{n} g(x_i - \bar{x}). \tag{17.2}$$

$10^\bullet$ 高斯的证明

- 对方程的两边 (17.2), 通过对 $x_1, x_2, \ldots, x_n$ 取偏导数, 我们得到

$$
\begin{aligned}
0 &= g'(x_1-\bar{x})(1+a)+g'(x_2-\bar{x})a+\cdots+g'(x_n-\bar{x})a,\\
0 &= g'(x_1-\bar{x})a+g'(x_2-\bar{x})(1+a)+\cdots+g'(x_n-\bar{x})a,\\
&\vdots\\
0 &= g'(x_1-\bar{x})a+g'(x_2-\bar{x})a+\cdots+g'(x_n-\bar{x})(1+a),
\end{aligned}
$$

其中 $a \triangleq -1/n$.

- 以矩阵和向量的形式，我们有

$$\mathbf{0}_n = \begin{pmatrix} 1+a & a & \cdots & a \\ a & 1+a & \cdots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \cdots & 1+a \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

其中 $y_i \triangleq g'(x_i - \bar{x})$, $i = 1, \ldots, n$.

- 求解该齐次线性方程组 (Homogeneous Linear Equations)，我们得到

$$\begin{aligned} & y_1 = \cdots = y_n = c, \\ \Rightarrow\ & g'(x_1 - \bar{x}) = \cdots = g'(x_n - \bar{x}) = c, \\ \Rightarrow\ & g(t) = ct + b, \\ \Rightarrow\ & 0 \overset{(17.2)}{=} \sum_{i=1}^{n} g(x_i - \bar{x}) = \sum_{i=1}^{n} \left[c(x_i - \bar{x}) + b \right] = 0 + nb, \\ \Rightarrow\ & b = 0. \end{aligned}$$

◆ 因此，

$$g(t) = ct = \frac{f'(t)}{f(t)} \quad \Rightarrow \quad f(t) = K \exp\left(\frac{ct^2}{2}\right).$$

◆ 从

$$\int_{-\infty}^{+\infty} f(t) \, \mathrm{d}t = 1$$

并利用如下的积分恒等式

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-t^2} \, \mathrm{d}t = \sqrt{\pi},$$

我们得到

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t^2}{2\sigma^2}\right). \qquad \square$$

17.7 所包含的思政元素 (I)

- **反向思维**。正态分布母体的均值 μ 的 MLE 是样本均值 $\bar{X}$。反过来，如果μ的 MLE 是$\bar{X}$，则该密度函数是正态分布的密度函数。一般说来，学生们都不敢提这样的反问题。

- 我们的教育应该鼓励学生，**打破陈规、大胆提问**。在问及 2018 年埃尼“前沿能源奖”得主、中国科学院外籍院士王中林时，他表示，要培育顶尖科学家，最重要的一点，是在教育中，鼓励学生打破陈规、大胆提问。要鼓励学生独立且有创造性地思考，而不是只会解题或者是死记硬背。因为书本的内容可能会出错，或者有需要改进之处。必须要训练学生的逻辑思维，去**发现问题、解决问题**，此外还应培养他们的兴趣爱好。“那些在入学考试上答对了所有题目的人，未来未必会成为科学家。”王中林表示。

17.7 所包含的思政元素 (II)

- 了解**学科历史和大师成长之路**。正态分布有时称为**高斯分布**，是为了突出高斯在推导出该密度函数中的重要贡献。**能提出问题，等于解决了问题的一半，有时比解决问题之本身更重要**。该问题由法国人拉普拉斯在 1788 年首次提出，但未解决。21 年以后，在 1809 年，高斯在《天体运动理论》中首次推导出正态分布。

- **知识是相互联系的**。高斯的证明过程，既简单又天才，涉及到偏导数、齐次线性方程组、常微分方程、King 函数。这就说明了：**打好数学基础对其它学科的重要性**。