2025 April

数学月刊

Mathematics Monthly

Author :
Guanzhong Yang(student)
Yan Wang(teacher)

Mathematics changes our lives

PREFACE

This month, we are going to talk about two important statistics knowledge: normal distribution and central limit theory. They are the foundations of many further conclusions, whereas their proofs are not provided by our textbooks.

If you have other brilliant questions or knowledge willing to learn, email to anmiciuangray@163.com **for surprising rewards!**

1. 正态分布

Introduction

正态分布来源于误差的分布. 在我们观察多次测量x_i的测量误差ϵ_i的时候, 我们发现其有几个特征:

(1) 所有的观测值都可以有误差, 其误差可能来源于观测者/仪器/实验条件等.

(2) 观测误差对称地分布在 0 的两侧.

(3) 小误差的出现频率比大误差的出现频率更高.

因此我们用数学的语言对其中的(1) (2)进行翻译:

(1) 误差$\epsilon = x - \mu$, 其中 x 是观察值, μ是真实值.

(2) 如果 $f(\epsilon)$表示误差分布, 则 $f(\epsilon) = f(-\epsilon)$.

Proof

下面我们来证明一下正态分布的公式.

假设我们有 n 个独立的观测值$x_1, x_2, ..., x_n$, 其对应的误差是$\epsilon_i = x_i - \mu$. 那么我们得到观测值的概率就是

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

根据自然想让我们得到观测值的概率最大化, 因此, 我们就将这个问题转化成了最大似然估计的问题.

$$\frac{\partial}{\partial\mu}\left[\sum_{i=1}^{n} \ln f(\epsilon_i)\right] = 0,$$

$$\sum_{i=1}^{n} \frac{f'(\epsilon_i)}{f(\epsilon_i)} = 0.$$

在这里, 高斯做出了第一个很重要的假设:

$$\mu = \sum_{i=1}^{n} x_i,$$

当其为μ的最大似然估计时:

$$\sum_{i=1}^{n} \epsilon_i = \sum_{i=1}^{n} x_i - \mu = 0.$$

随后, 高斯做出了第二个很重要的假设:

$$\frac{f'(\epsilon_i)}{f(\epsilon_i)} = k\epsilon_i,$$

这是因为其不但满足我们所提到的“(2) 如果 $f(\epsilon)$表示误差分布, 则 $f(\epsilon) = f(-\epsilon)$”:

$$\frac{f'(-\epsilon_i)}{f(-\epsilon_i)} = -k\epsilon_i,$$

也满足我们最大似然估计问题的关键证明:

$$\sum_{i=1}^{n} \frac{f'(\epsilon_i)}{f(\epsilon_i)} = \sum_{i=1}^{n} k\epsilon_i = 0.$$

至此, 我们便只需要求解微分方程即可.

$$\frac{f'(\epsilon)}{f(\epsilon)} = k\epsilon,$$

$$\ln f(\epsilon) = \frac{k}{2}\epsilon^2 + C,$$

$$f(\epsilon) = e^{C} e^{\frac{k}{2}\epsilon^2}.$$

因为概率密度函数归一化:

$$\int_{-\infty}^{\infty} f(\epsilon)d\epsilon = e^{C} \int_{-\infty}^{\infty} e^{\frac{k}{2}\epsilon^2} d\epsilon = 1.$$

为了使函数收敛, 高斯做出了第三个重要的假设:

$$k = \frac{1}{\sigma^2},$$

这样做既确保了函数一定收敛, 同时也让我们可以更好地调节误差分布的特征.

随后经过计算, 我们得到:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}.$$

2. 中心极限定理

Auxiliary Result

为了证明我们的结论, 我们需要一些傅立叶变化/卷积/特征函数的知识.

(1) 对于函数 $f(x)$, 如果其周期为 T, $w_n = \frac{n\pi}{T}$, 我们可以用傅立叶变化将其用一堆三角函数表示出来:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos w_n x + b_n \sin w_n x).$$

其中各项系数我们可以利用正交性进行求解:

$$a_n = \frac{1}{T}\int_{-T}^{T} f(x)\cos w_n x \, dx,\ b_n = \frac{1}{T}\int_{-T}^{T} f(x)\sin w_n x \, dx.$$

为了让形式更加通用, 我们将傅立叶变化转化成复数形式. 通过展开 $\cos w_n x$ 与 $\sin w_n x$:

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{iw_n x}, \text{ where } c_n = \frac{1}{2T}\int_{-T}^{T} f(x) e^{-iw_n x} \, dx,$$

我们之所以可以这样求c_n是因为任何包含$e^{w_k x}$的项都将被转化成 $\cos w_k x$ 和 $\sin w_k x$ 的线性组合, 最后通过积分消失, 因为积分域内包含其包含其的多个完整周期.

因此:

$$f(x) = \frac{1}{2T}\sum_{n=-\infty}^{\infty} \int_{-T}^{T} f(t) e^{-iw_n t} \, dt \; e^{iw_n x}.$$

记$\Delta w = w_n - w_{n-1} = \frac{\pi}{T}$, 当 $T \to \infty$时$\Delta w \to 0$, 此时我们可以对一个无限的 domain 的函数作傅立叶变化.

$$\lim_{T\to\infty} f(x) = \lim_{\Delta w \to 0} \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} \int_{-T}^{T} f(t) e^{-iw_n t} \, dt \; e^{iw_n x} \, \Delta w = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(t) e^{-iw_n t} dt\right] e^{iw_n x} \, dw.$$

此时, 内层积分被称为傅立叶变化, 外层积分被称为傅立叶逆变化.

(2) 随机变量 X 的特征函数定义为:

$$\phi_X(t) = E[e^{itX}].$$

对于连续性随机变量 X:

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx.$$

对于离散型随机变量 X:

$$\phi_X(t) = \sum e^{itx} p(x).$$

不难看出特征函数是概率分布的傅里叶变换的另一种形式, 因此特征函数与概率分布一一对应也就不奇怪了.

Result

设$X_1, X_2, \ldots, X_n$为独立同分布的随机变量, 均值为μ, 方差为σ^2. 定义它们的和为:

$$S_n = \sum_{i=1}^{n} X_i.$$

同时也定义当 n 趋近于无穷大时, 标准化后的和为:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

我们先求出正态分布的特征函数, 随后求出Z_n的特征函数, 并证明两者相同.

正态分布的特征函数为:

$$f_N(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

$$\phi_N(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2-\frac{t^2}{2}} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2} dx.$$

注意到$\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2}$为一个高斯积分, $\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} = \sqrt{2\pi}$, 而(x - it)仅仅代表一个水平的平移, 不影响积分:

$$\phi_N(t) = e^{-\frac{t^2}{2}}.$$

我们随后求Z_n的特征函数:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}},\ S_n = \sigma\sqrt{n}Z_n + n\mu.$$

$$\phi_{S_n}(t) = E[e^{it(\sigma\sqrt{n}Z_n + n\mu)}] = e^{itn\mu}\ E[e^{it\sigma\sqrt{n}Z_n}] = e^{itn\mu}\phi_{Z_n}(t\sigma\sqrt{n}).$$

又因为$X_1, X_2, \ldots, X_n$为独立同分布的随机变量:

$$\phi_{S_n}(t) = E[e^{itS_n}] = E[e^{it\sum_{i=1}^{n} X_i}] = \{E[e^{itX}]\}^n = [\phi_X(t)]^n.$$

所以:

$$[\phi_X(t)]^n = e^{itn\mu}\phi_{Z_n}(t\sigma\sqrt{n}).$$

让 $u = t\sigma\sqrt{n}$:

$$\phi_{Z_n}(u) = [\phi_X(\frac{u}{\sigma\sqrt{n}})]^n e^{-i\frac{u\mu\sqrt{n}}{\sigma}}.$$

又因为对$\phi_X(t)$先在零点进行泰勒展开, 再考虑 $E(X) = \mu$, $E(X^2) = \mu^2 + \sigma^2$, 最后将 $t = \frac{u}{\sigma\sqrt{n}}$带入:

$$\phi_X(t) = 1 + itE(X) + \frac{(it)^2}{2!}E(X^2) + \frac{(it)^3}{3!}E(X^3) + \ldots = 1 + it\mu - \frac{t^2\sigma^2}{2} + o(t^2).$$

$$\phi_X(\frac{u}{\sigma\sqrt{n}}) = 1 + i\frac{u\mu}{\sigma\sqrt{n}} - \frac{u^2}{2n} + o(\frac{u^2}{n}).$$

当 $n \to \infty$时, $\frac{u}{\sigma\sqrt{n}} \to 0$, 因此泰勒展开是有效的. 考虑到 $o(\frac{u^2}{n})$是高阶无穷小可以忽略, $i\frac{u\mu}{\sigma\sqrt{n}}$代表虚部趋近于 0, 我们可以近似地得出:

$$\lim_{n\to\infty} [\phi_X(t)]^n = \lim_{n\to\infty} [\phi_X(\frac{u}{\sigma\sqrt{n}})]^n = \lim_{n\to\infty} [1 - \frac{u^2}{2n}]^n = e^{-\frac{u^2}{2}}.$$

同时因为当 $n \to \infty$时, $e^{-i\frac{u\mu\sqrt{n}}{\sigma}}$代表虚部趋近于 0, $e^{-i\frac{u\mu\sqrt{n}}{\sigma}} \to 1$:

$$\lim_{n\to\infty} \phi_{Z_n}(u) = \lim_{n\to\infty} [\phi_X(\frac{u}{\sigma\sqrt{n}})]^n e^{-i\frac{u\mu\sqrt{n}}{\sigma}} = e^{-\frac{u^2}{2}}.$$

我们注意到, 正态分布的特征函数与Z_n的特征函数相同, 说明当 $n \to \infty$, $Z_n \sim N(0,1)$.

Afterwords

说白了, 傅立叶变化和特征函数都是在干一个事情: 那就是找每一项的系数, 而我们得到的答案实际上是系统的通项公式. 如此, 也就不奇怪为什么傅立叶变化和特征函数都是 bijection