第二十章 中心极限定理的证明

20.1 证明的关键思路

一些独立同分布的随机变量 $X_1, X_2, \cdots$ 均值为 μ，方差为 σ^2，那么标准化随机变量 $Z_n = \dfrac{X_1+X_2+\cdots+X_n - n\mu}{\sqrt{n}\cdot\sigma}$

会服从标准正态分布的随机变量 Z.

条件：存在 $\delta > 0$，当 $|t| < \delta$ 时，$M_X(t)$ 收敛（复分析证明）

思路是：

当 $|t| < \delta$ 时，$M_{Z_n}(t) = M_Z(t)$

$$\Longleftrightarrow \lim_{n\to\infty} Z_n = Z$$

需证明：$\int_{-\infty}^{+\infty} h(x) f(x)\, dx = \int_{-\infty}^{+\infty} h(x) g(x)\, dx$

在连续的条件下可推 $f(x) = g(x)$

证：　不失一般性地，我们假设

$f(0)=1$，$g(0)=-1$，并且在区间

$[-\eta,\eta]$ 上，$h(x)=\frac{1}{2\eta}$，其它为 0.

限制在 $[-\eta,\eta]$ 上都有 $f(x)>\frac{1}{2}$，

$g(x)<-\frac{1}{2}$.

则 $\int_{-\infty}^{+\infty} h(x)f(x)dx \geq \int_{-\eta}^{\eta} h(x)f(x)dx$

$$= \int_{-\eta}^{\eta} \frac{1}{2\eta}\cdot\frac{1}{2}dx$$

$$= \frac{1}{2}$$

同理可得 $\int_{-\infty}^{+\infty} h(x)g(x)dx \leq -\frac{1}{2}$

这是矛盾的，由此推出 $f=g$！

证明两个函数相等的方法可以是证明其关于一大类测试函数有相同的积分

20.2 中心极限定理的陈述

定理 20.2.2 (**中心极限定理 (CLT)**) 设 $X_1, \cdots, X_N$ 是独立同分布的随机变量，存在某个 $\delta > 0$，使得当 $|t| < \delta$ 时它们的矩母函数收敛 (这意味着所有矩均存在且有限). 用 μ 表示其均值，σ^2 表示方差. 令

$$\overline{X}_N = \frac{X_1 + \cdots + X_N}{N}$$

且

$$Z_N = \frac{\overline{X}_N - \mu}{\sigma/\sqrt{N}}.$$

那么，当 $N \to \infty$ 时，Z_n 的分布会收敛于标准正态分布 (相关陈述参见定义 20.2.1).

当进行了大量重复实验后，任一次实验结果的值的落点服从一个正态分布

解释：

$$E(\overline{X}_n) = E\left(\frac{\sum_{i=1}^{N} x_i}{N}\right)$$

$$= \frac{1}{N} \cdot N\mu$$

$$= \mu$$

$$Var(\overline{X}_n) = Var\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)$$

$$= \frac{1}{n^2} \cdot n\sigma^2$$

$$= \frac{\sigma^2}{n}$$

标准差为 $\frac{\sigma}{\sqrt{n}}$. 当 $n \to \infty$. $\frac{\sigma}{\sqrt{n}} \to 0$.

$\Rightarrow$ 随着实验次数增加. 平均值会趋向真实的均值 (大数定律理解)

20.3 均值、方差与标准差.

20.4 标准化

定义 20.4.1 (随机变量的标准化) 设 X 是一个均值为 μ 且标准差为 σ 的随机变量, 并且 μ 和 σ 都是有限的. X 的标准化变量 Z 被定义为

$$Z := \frac{X - \mathbb{E}[X]}{\mathrm{StDev}(X)} = \frac{X-\mu}{\sigma}.$$

注意

$$\mathbb{E}[Z] = 0 \quad 且 \quad \mathrm{StDev}(Z) = 1.$$

X 服从某均值为 μ. 方差为 σ^2 的分布.

先取 $Y = X - \mu$.

$F_Y(y) = \mathrm{Prob}(Y \leq y)$

$$= \mathrm{Prob}(X \le y + \mu)$$

$$= F_X(y+\mu)$$

$$\Rightarrow \quad f_Y(y) = f_X(y+\mu)$$

$$\Rightarrow \quad E(y) = \int_{-\infty}^{+\infty} y\, f_Y(y)\, dy$$

$$= \int_{-\infty}^{+\infty} (x-\mu) f_X(y+\mu)\, d(x-\mu)$$

$$= \int_{-\infty}^{+\infty} x f_X(x)\, dx - \mu \int_{-\infty}^{+\infty} f_X(x)\, dx$$

$$= \mu - \mu$$

$$= 0.$$

再取 $Z = \dfrac{Y}{6} = \dfrac{X-\mu}{6}$

易知 $E(Z) = \dfrac{1}{6} \cdot E(Y) = 0.$

$$\mathrm{Var}(Z) = E(Z^2) - E^2(Z)$$

$$= E(Z^2)$$

$$= \int_{-\infty}^{+\infty} z^2 f_z(z)\, dx$$

$$F_Z(z) = \text{Prob}(Z \le z)$$

$$= \text{Prob}(Y \le 6z)$$

$$= F_Y(6z)$$

$$= F_X(6z+\mu)$$

$$f_Z(z) = 6 f_X(6z+\mu)$$

$$= 6 f_X(x)$$

$$Var(Z) = \int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{6^2} \cdot 6 f_X(x)\, d\left(\frac{x-\mu}{6}\right)$$

$$= \frac{1}{6^2} \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)$$

$$= \frac{1}{6^2} \cdot 6^2$$

$$= 1$$

得证.

20.5 矩母函数的相关结果

定理 20.5.1 (正态分布的矩母函数) 设随机变量 X 服从均值为 μ 且方差为 σ^2 的正态分布. 它的矩母函数是

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}.$$

特别地, 如果 Z 服从标准正态分布, 那么它的矩母函数就是

$$M_Z(t) = e^{t^2/2}.$$

$X \sim N(\mu, \sigma^2)$

设 $z = \frac{X-\mu}{\sigma}$

有 $M_X(t) = M_{\sigma z+\mu}(t)$

$$= e^{\mu t} M_z(\sigma t)$$

先计算 $M_z(t)$

$$= \int_{-\infty}^{+\infty} e^{zt} \cdot \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\, dz$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z-t)^2}{2} + \frac{t^2}{2}\right) dz$$

$= \cdots$

$= e^{-\frac{t^2}{2}}$

由此可得 $M_X(t) = e^{\mu t} \cdot e^{-\frac{\sigma^2 t^2}{2}}$

$$= e^{\mu t - \frac{\sigma^2 t^2}{2}}$$

在一个不太强的条件约束下，由 $n \to \infty$ 函数列的矩母函数收敛到正态分布的矩母函数，可以得到该数列本身收敛到正态分布函数。

20.6 特殊情形：服从泊松分布的随机变量之和。

定理 20.6.1 设 $X_1, \cdots, X_N$ 是相互独立的随机变量，它们均服从参数为 λ 的泊松分布。令

$$\overline{X}_N = \frac{X_1 + \cdots + X_N}{N}.$$

当 $N \to \infty$ 时，$\overline{X}_N$ 会收敛于均值为 λ 且方差为 λ/n 的正态分布。

考虑 $z = \dfrac{\overline{X}_N - \mu}{\sigma}$

$$= \frac{\frac{X_1 + X_2 + \cdots + X_N}{N} - \mu}{\left(\sigma/\sqrt{N}\right)}$$

$$= \frac{X_1 + X_2 + \cdots + X_N - N\mu}{\sqrt{N}\,\sigma}$$

有 $M_Z(t) = M_{\frac{\sum_{i=1}^{N} X_i - N\mu}{\sqrt{N}\sigma}}(t)$

$$= \prod_{i=1}^{N} M_{\frac{X_i-\mu}{\sqrt{N}\sigma}}(t)$$

※ $$= \prod_{i=1}^{N} e^{\frac{-\mu t}{\sigma\sqrt{N}}} M\left(\frac{t}{\sqrt{N}\sigma}\right)$$

$$= \prod_{i=1}^{N} e^{\frac{-\mu t}{\sigma\sqrt{N}}} \cdot e^{\mu\left(e^{\frac{t}{\sigma\sqrt{N}}}-1\right)}$$

$$= \prod_{i=1}^{N} e^{\frac{-\mu t}{\sigma\sqrt{N}}} \cdot e^{\lambda\cdot\left(\frac{t}{\sigma\sqrt{N}} + \frac{t^2}{2\sigma^2 N} + \cdots\right)}$$

$$O\left(\frac{t^3}{3!\,\sigma^3 N^{\frac{3}{2}}}\right)$$

$$= \prod_{i=1}^{N} e^{\frac{t^2}{2\sigma^2 N} + O\left(\frac{t^3}{3!\,\sigma^3 N^{\frac{3}{2}}}\right)}$$

$$= e^{\frac{t^2}{2} + O\left(\frac{t^3}{\sqrt{N}}\right)}$$

当 $N \to \infty$ 时，皮亚诺余项 $\to 0$。

得证

20.7 利用MGF证明一般的CLT

※ 之式适用于一般的pdf.

$$\text{原式} = e^{\frac{-\mu t\sqrt{N}}{\sigma}} M_X\left(\frac{t}{\sqrt{N}\sigma}\right)^N \triangleq u.$$

$$\ln u = -\mu t\sqrt{N} + N \ln M_X\left(\frac{t}{\sqrt{N}\sigma}\right)$$

$$\begin{cases} M_X = 1 + \mu t + \dfrac{\mu_2' t^2}{2!} + \cdots \\ \ln(1+u) = u - \dfrac{u^2}{2!} + \dfrac{u^3}{3!} + \cdots \end{cases}$$

把上面的两个式子结合起来可得

$$\begin{aligned}\log M_X(t) &= t\left(\mu + \frac{\mu_2' t}{2} + \cdots\right) - \frac{t^2\left(\mu + \frac{\mu_2' t}{2} + \cdots\right)^2}{2} + \cdots \\ &= \mu t + \frac{\mu_2' - \mu^2}{2} t^2 + t^3 \text{ 或更高次项.}\end{aligned}$$

于是

$$\log M_X(t) = \mu t + \frac{\mu_2' - \mu^2}{2} t^2 + t^3 \text{ 或更高次项.}$$

但我们想求的是 M_X 在 $t/\sigma\sqrt{N}$ 处的值, 而不是在 t 处的值. 于是

$$\log M_X\left(\frac{t}{\sigma\sqrt{N}}\right) = \frac{\mu t}{\sigma\sqrt{N}} + \frac{\mu_2' - \mu^2}{2}\frac{t^2}{\sigma^2 N} + t^3/N^{3/2} \text{ 项或 } N \text{ 的次数更低的项.}$$

因此, 我们把 N 的次数更低的项记作 $O(N^{-3/2})$. 当乘以 N 之后, 它们就变成了新的误差项 $O(N^{-1/2})$.

$$\ln u = -\mu t\sqrt{N} + N\cdot\left(\frac{\mu t}{\sigma\sqrt{N}} + \frac{\mu_2' - \mu^2}{2}\cdot\frac{t^2}{\sigma^2 N}\right) + O(N^{-\frac{1}{2}})$$

$$= \frac{\mu_2' - \mu^2}{2}\cdot\frac{t^2}{\sigma^2}$$

$$= \frac{t^2}{2} \qquad (\sigma^2 = E(x^2) - E^2(x))$$

$$\Rightarrow \text{原式} = u = e^{\frac{t^2}{2}}$$

得证.

20.8 使用中心极限定理

20.9 中心极限定理与蒙特卡罗积分