参数估计

大数律和中心极限定理

强大数律

> **记号**：若$P(A)=1$，则记A a.s.，即A几乎一定（almost surely）会发生。

如果X_i是独立同分布的随机变量，且$\mu=EX_1$，则

$$\lim_{n\to\infty}\overline{X_n}=\mu \text{ a.s.}$$

因为概率等于1的事件在实际中必然发生，所以在强大数律中，如果用 x_n 表示 X_n 的观测值，则有

$$\lim_{n\to\infty}\frac{x_1+x_2+\cdots+x_n}{n}=\mu.$$

因为强大数律的数学证明并不需要概率的频率定义，所以它从理论上保证了概率的频率定义是正确的。

> 利用强大数律的关键是构造独立同分布的一组随机变量。

弱大数律

依概率收敛

设 $U,U_1,U_2,\cdots$ 是随机变量。如果对任何 $\varepsilon>0$，$\lim_{n\to\infty}P\left(\left|U_n-U\right|\geqslant\varepsilon\right)=0$，则称 U_n 依概率收敛到 U, 记做 $U_n\xrightarrow{p}U$。

切比雪夫不等式

随机变量X的数学期望是μ，方差是σ^2，则对常数$\varepsilon>0$，有

$$P(|X-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{\varepsilon^2}$$

弱大数律

设随机变量 $X_1,X_2,\cdots$ 独立同分布，$\mu=\mathrm{E}X_1$，则对任何 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\left(\left|\bar{X}_n-\mu\right|\geqslant\varepsilon\right)=0$$

> **对于强大数律和弱大数律的理解**
>
> 设甲某是一位新的专职司机，用 U_n 表示他在第 n 个工作日的交通事故造成的损失。因为他的开车经验在不断提高，所以随着时间的推移，他的 U_n 会向老司机的 $U=0$ 收敛。如果 $U_n\xrightarrow{P}U$，则 $n\to\infty$ 时，我们只能得到
>
> $$P\left(U_n\geqslant\varepsilon\right)=P\left(\left|U_n-U\right|\geqslant\varepsilon\right)\to0$$
>
> 所以对任意大的 n，都不能保证 $P\left(U_n\geqslant\varepsilon\right)=0$。也就是说，无论有多长的开车经验，这位新司机因交通事故造成较大损失的概率都是正数，从而都有可能造成较大的损失。用 u_n 表示 U_n 的观测值。如果 $U_n\to U$ a.s.，则实际中有 $u_n\to0$。说明存在 n_0，使得 $n\geqslant n_0$ 时，$u_n<\varepsilon$。也就是说，从某天开始，这位新司机就再也不会发生有较大损失的交通事故了。

中心极限定理

中心极限定理研究的是当n较大时，随机变量的部分和$S_n=\sum X_j$的概率分布问题。实验表明，独立同分布随机变量和的分布近似于正态分布，这就是将要介绍的中心极限定理。

设随机变量 $X_1,X_2,\cdots$ 独立同分布，$\mathrm{E}X_1=\mu,\operatorname{Var}\left(X_1\right)=\sigma^2>0$。用 $S_n=\sum_{j=1}^nX_j$ 表示部分和，用 $Z_n=\dfrac{S_n-n\mu}{\sqrt{n\sigma^2}}$表示 S_n 的标准化，用 $\Phi(x)$ 表示服从 $N(0,1)$ 的分布函数。

在上述条件下，当n趋于正无穷时，有

$$P(Z_n\leq x)\to\Phi(x),x\in(-\infty,\infty)$$

即Z_n依分布收敛到$N(0,1)$，记作$Z_n\xrightarrow{d}N(0,1)$。

推论：在上述条件下，对较大的n，有

$$P\left(\frac{S_n-n\mu}{\sqrt{n\sigma^2}}\leqslant x\right)\approx\Phi(x),\quad P\left(\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\leqslant x\right)\approx\Phi(x).$$

在一些实际问题中，随机变量的方差 σ^2 是未知的，这时可以用下面两个统计量来估计σ^2。

$$\hat{\sigma}^2=\frac{1}{n-1}\sum_{j=1}^{n}\left(X_j-\bar{X}_n\right)^2 \text{ [or] } \hat{\sigma}^2=\frac{1}{n}\sum_{j=1}^{n}\left(X_j-\bar{X}_n\right)^2$$

参数估计（上）

一些术语

【总体假设】设总体含有 N 个个体，第 i 个个体是 y_i 。

总体均值：$\mu=\frac{y_1+y_2+\cdots+y_N}{N}$。

总体方差/方差：$\sigma^2=\frac{(y_1-\mu)^2+(y_2-\mu)^2+\cdots+(y_N-\mu)^2}{N}$。

总体标准差/标准差：$\sigma=\sqrt{\sigma^2}$。

总体参数/参数：描述总体特性的指标，包括但不限于上面提到的三者，讲到参数的时候要明确它是哪个总体的参数。

【样本假设】设总体的一个抽样有n个样本，第i个样本是x_i。

样本均值：$\bar{x}$。

样本方差：$s^2=\frac{1}{n-1}\left[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2\right]$。

样本标准差：$s=\sqrt{s^2}$。

估计

估计是利用**样本**计算出对**参数**的估计值。估计不是唯一的，对相同的观测数据，不同的方法可以给出不同的估计结果。

> 在总体中任取一个个体X，X是随机变量，$EX=\mu$是总体均值，这说明随机抽样是无偏的。
>
> 如果用 $X_1,X_2,\cdots,X_n$ 表示依次随机抽取的样本，则样本均值$\bar{X}$是总体均值 μ 的估计，且 $E\bar{X}=\mu$.

设 $\hat{\theta}$ 是 θ 的估计。

（1）如果 $\mathrm{E}\hat{\theta}=\theta$, 则称 $\hat{\theta}$ 是 θ 的无偏估计；

（2）如果当样本量 $n\to\infty, \hat{\theta}$ 依概率收敛到 θ，则称 $\hat{\theta}$ 是 θ 的相合估计；

（3）如果当样本量 $n\to\infty, \hat{\theta}$ 以概率 1 收敛到 θ，则称 $\hat{\theta}$ 是 θ 的强相合估计。

样本均值

设总体均值 $\mu=EX$ 存在，$X_1,X_2,\cdots,X_n$ 是总体 X 的样本。均值 μ 的估计定义为

$$\bar{X}_n=\frac{X_1+X_2+\cdots+X_n}{n}$$

样本均值 $\bar{X}_n$ 有如下的性质：

（1）$\bar{X}_n$ 是 μ 的无偏估计，这是因为 $\mathrm{E}\bar{X}_n=\mu$;

（2）$\bar{X}_n$ 是 μ 的强相合估计，从而是相合估计。这是因为从强大数律得到

$$\lim_{n\to\infty}\bar{x}_n=\mu \text{ a.s.}$$

样本方差

给定总体 X 的样本 $X_1,X_2,\cdots,X_n$，用 $\hat{\mu}$ 表示样本均值。总体方差 $\sigma^2=\mathrm{Var}(X)$ 的估计由下式定义。

$$S^2=\frac{1}{n-1}\sum_{j=1}^{n}\left(X_j-\hat{\mu}\right)^2$$

样本方差的性质：

（1）样本方差 S^2 是总体方差的无偏估计，$\mathrm{E}S^2=\frac{1}{n-1}\sum_{j=1}^{n}\mathrm{E}\left(X_j-\hat{\mu}\right)^2=\frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2=\sigma^2$.

> **具体证明**：取定j。因为 $\mathrm{E}\left(X_j-\hat{\mu}\right)=\mu-\mu=0$，所以从$X_1,X_2,\cdots,X_n$ 的独立性得到

> $$
> \begin{aligned}
> \mathrm{E}\left(X_j-\hat{\mu}\right)^2 &= \operatorname{Var}\left(X_j-\frac{1}{n}\sum_{i=1}^{n}X_i\right)\\
> &= \operatorname{Var}\left[\left(1-\frac{1}{n}\right)X_j-\frac{1}{n}\sum_{i\neq j}X_i\right]\\
> &= \left(\frac{n-1}{n}\right)^2\sigma^2+\frac{1}{n^2}\sum_{i\neq j}\sigma^2\\
> &= \frac{\left[\left(\frac{n-1}{n}\right)^2+\frac{n-1}{n^2}\right]\sigma^2}{n}\sigma^2\\
> &= \frac{n-1}{n}\sigma^2
> \end{aligned}
> $$

（2）样本方差 S^2 是总体方差 σ^2 的强相合估计。

> **具体证明**：利用强大数律 $\hat{\mu}\to\mu$ a.s. 和$\frac{1}{n-1}\sum_{j=1}^{n}X_j^2\to\mathrm{E}X^2$ a.s.，得到
>
> $$
> \begin{aligned}
> \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j-\hat{\mu}\right)^2 &= \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j^2-2X_j\hat{\mu}+\hat{\mu}^2\right)\\
> &= \frac{1}{n-1}\left(\sum_{j=1}^{n}X_j^2-2n\frac{1}{n}\sum_{j=1}^{n}X_j\hat{\mu}+n\hat{\mu}^2\right)\\
> &= \frac{1}{n-1}\sum_{j=1}^{n}X_j^2-\frac{n}{n-1}\hat{\mu}^2\\
> &\to EX^2-\mu^2=\sigma^2 \text{ a.s.}
> \end{aligned}
> $$

样本标准差

由于 S^2 是 σ^2 的估计，所以定义标准差 σ 的估计为

$$
S=\sqrt{S^2}=\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(X_j-\hat{\mu}\right)^2}
$$

称 S 为样本标准差。

（1）由于 $S^2\to\sigma^2$ a.s.，所以 $S\to\sigma$ a.s. 成立，说明 S 是 σ 的强相合估计。

（2）当 $\sigma>0, S$ 不是 σ 的无偏估计，也就是说 $ES=\sigma$ 不成立。这是因为**没有不全为零的常数a,b使得** $P(aS+b=0)=1$，所以由内积不等式得到

$$
\mathrm{E}S=\mathrm{E}(S\cdot 1)<\sqrt{\mathrm{E}S^2\cdot\mathrm{E}1^2}=\sqrt{\sigma^2}=\sigma.
$$

这时称 S 低估了 σ。

> 因为上面三个的强相合性，如果 $x_1,x_2,\cdots,x_n$ 是总体 X 的样本，则如下事实得到保证：
>
> $$
> \lim_{n\to\infty}\bar{x}_n=\mu,\ \lim_{n\to\infty}s^2=\sigma^2,\ \lim_{n\to\infty}s=\sigma.
> $$

样本矩

因为 $X_1^k,X_2^k,\cdots,X_n^k$ 独立同分布，且和 X^k 同分布，所以是总体 X^k 的样本。并且

$$
\hat{\mu}_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k
$$

是 μ_k 的估计。所以 $\hat{\mu}_k$ 是 μ_k 的**强相合无偏估计**。称 $\mu_k=\mathrm{E}X^k$ 为 X 的 k 阶原点矩, 称 $\hat{\mu}_k$ 为 k 阶样本 原点矩。

> **估计的技巧**
>
> （1）如果总体 X 的分布函数 $F(x;\theta)$ 只有一个未知参数 θ，则 $\mu_1=\mathrm{E}X$ 常和 θ 有关。如果 $g(s)$ 是已知函数，并且能从$\mu_1=\mathrm{E}X$得到$\theta=g\left(\mu_1\right)$，则 $\hat{\theta}=g\left(\hat{\mu}_1\right)$ 是 θ 的矩估计，其中 $\hat{\mu}_1$ 是样本均值。
>
> （2）如果总体 X 的分布函数 $F\left(x;\theta_1,\theta_2\right)$ 有 2 个未知参数 θ_1,θ_2，则 $\mu_1=\mathrm{E}X$ 和 $\mu_2=\mathrm{E}X^2$ 常和 θ_1,θ_2 有关。如果 $g_1(s,t),g_2(s,t)$ 是 已知函数，并且能从
>
> $$
> \begin{cases}\mu_1=\mathrm{E}X,\\ \mu_2=\mathrm{E}X^2\end{cases}\quad[\text{get}]\quad\begin{cases}\theta_1=g_1\left(\mu_1,\mu_2\right),\\ \theta_2=g_2\left(\mu_1,\mu_2\right),\end{cases}
> $$
>
> 则$\hat{\theta}_1=g_1\left(\hat{\mu}_1,\hat{\mu}_2\right),\hat{\theta}_2=g_2\left(\hat{\mu}_1,\hat{\mu}_2\right)$分别是 θ_1,θ_2 的矩估计。

最大似然估计

离散情况

设离散随机变量 $X_1, X_2, \cdots, X_n$ 有联合分布

$$p\left(x_1, x_2, \cdots, x_n ; \theta\right)=P\left(X_1=x_1, X_2=x_2, \cdots, X_n=x_n\right),$$

其中 θ 是未知参数，给定观测数据 $x_1, x_2, \cdots, x_n$ 后，称 θ 的函数

$$L(\theta)=p\left(x_1, x_2, \cdots, x_n ; \theta\right)$$

为似然函数，称 $L(\theta)$ 的最大值点 $\hat{\theta}$ 为 θ 的最大似然估计.

> 许多情况下，为了方便计算，一般取$l(\theta)=\ln L(\theta)$进行最值点的求解。

连续情况

类似离散的情况，此时$P(X=x)=f(x;\theta)dx$。

设随机向量 $X=\left(X_1, X_2, \cdots, X_n\right)$ 有联合密度 $f(\boldsymbol{x} ; \boldsymbol{\theta})$，其中 $\boldsymbol{\theta}=\left(\theta_1, \theta_2, \cdots, \theta_m\right)$ 是未知参数。得到 $\boldsymbol{X}$ 的观测值 $\boldsymbol{x}=\left(x_1, x_2, \cdots, x_n\right)$ 后，称

$$L(\boldsymbol{\theta})=f(\boldsymbol{x} ; \boldsymbol{\theta})$$

为 $\boldsymbol{\theta}$ 的似然函数，称 $L(\boldsymbol{\theta})$ 的最大值点 $\hat{\boldsymbol{\theta}}$ 为 $\boldsymbol{\theta}$ 的最大似然估计（MLE）。

设总体 X 有概率密度 $f(x ; \boldsymbol{\theta})$，则 X 的样本 $X_1, X_2, \cdots, X_n$ 有联合密度

$$f\left(x_1, x_2, \cdots, x_n ; \boldsymbol{\theta}\right)=\prod_{j=1}^n f\left(x_j ; \boldsymbol{\theta}\right)$$

基于观测值 $\boldsymbol{x}=\left(x_1, x_2, \cdots, x_n\right)$ 的似然函数是

$$L(\boldsymbol{\theta})=\prod_{j=1}^n f\left(x_j ; \boldsymbol{\theta}\right)$$

同样取$l(\vec{\theta})=\ln L(\vec{\theta})$，求最大值点可以通过解方程组$\frac{\partial l(\vec{\theta})}{\partial \theta_j}=0, j=1,2,\dots,m$来获得。

> 这里要深刻理解，还需要做一些题，也可以看一下书上的例子。比如估计$N(\mu,\sigma^2)$，这里σ^2是参数而不是σ，要对σ^2求偏微分。