统一框架下常见分布的参数估计及其优良性分析

——数理统计课程的教学补充

陈 银

扬州大学数学学院　江苏扬州

【摘要】参数估计的优良性是数理统计学中的核心问题，直接关系到统计推断的可靠性与决策质量。在统一框架下，本文逐一推导了常见分布（二项分布，泊松分布，正态分布，指数分布）的参数的最大似然估计（MLE）和矩估计（ME）这两种估计量的解析表达式，并基于无偏性准则和相合性准则对其进行了系统比较。主要结果表明：二项分布和泊松分布中参数的 MLE 与 ME 均为无偏且相合；正态分布的参数 μ 的 MLE 与 ME 均为无偏且相合的，参数 σ^2 的 MLE 与 ME 为相合但有偏的；指数分布的参数 λ 的 MLE 与 ME 是相合但有偏的。本研究为常见分布参数估计方法的选择提供了理论依据，亦可供数理统计教学参考。

【关键词】二项分布；泊松分布；正态分布；指数分布；无偏性；有效性；相合性

【收稿日期】2026 年 7 月 4 日　【出刊日期】2026 年 8 月 1 日　【DOI】10.12208/j.aam.20260010

Optimality analysis of parameter estimation for common distributions under a unified framework: A teaching supplement for mathematical statistics courses

Yin Chen

School of Mathematic, Yangzhou University, Yangzhou, Jiangsu

【**Abstract**】The optimality properties of parameter estimation are central topic in mathematical statistics as it directly affects the reliability of statistical inference and the quality of decision-making. Within a unified framework, this paper derives step by step the formulas for both the maximum likelihood estimators (MLE) and the moment estimators (ME) of the parameters of several common distribution (binomial, Poisson, normal, and exponential distributions) and systematically compares them based on the criteria of unbiasedness and consistency. The main results show that for the binomial and Poisson distributions, both MLE and ME of their parameters are unbiased and consistent; for the normal distribution, MLE and ME of the parameter μ are unbiased and consistent, whereas those of the parameter σ^2 are consistent but biased; for the exponential distribution, MLE and ME of the parameter λ is consistent but biased. This study provides a theoretical basis for choosing parameter estimation methods for common distributions and may also serve as a reference for the teaching of mathematical statistics.

【**Keywords**】Binomial distribution; Poisson distribution; Normal distribution; Exponential distribution; Unbiasedness; Efficiency; Consistency

1 引言

1.1 研究背景

参数估计是数理统计的核心问题之一，它的发展可以追溯到 18 世纪末，为解决天文观测中的误差问题，德国数学家 Gauss 和法国数学家 Legendre 几乎同时独立提出了最小二乘法，通过使误差的平方和最小来寻找参数的最佳估计，它主要用于线性统计模型中的参数估计问题。1894 年，英国统计学家 Pearson[12] 提出了矩估计法，主要思想是用样本矩去替换总体矩，从而估计出未知参数。这种方法计算简单，是早期的系统性估计方法之一。1912 年，英国统计学家 Fisher[5]，创造性地提出了最大似然估计法（MLE），引

入“似然”概念，认为能使得当前样本出现概率最大的那个参数值，就是最合理的估计。之后 Fisher[6]还系统提出了估计量的一致性、充分性、有效性等评价标准，他认为一个好的估计量应能充分利用样本信息，其渐近方差应达到一个理论下界，从而定义了 Fisher 信息量[7]。这一思想将统计推断从特定方法提升到了普遍原理的高度。随后在 Fisher 的理论基础上，Cramér[3]和 Rao[13]在 1940 年代中期独立地将其精确化，证明了任何无偏估计量的方差下界是 Fisher 信息量的倒数。这一结果将“有效性”这一优良性准则铸造成了一个可计算的数学工具，成为判别最优无偏估计的“黄金标准”。但是 Cramér-Rao 下界并不总能达到。Lehmann-Scheffé[10,11]在 1950 年代利用充分完备统计量，给出了寻找最小方差无偏估计（UMVUE）的系统性方法（Lehmann-Scheffé 定理）。这标志着对于“无偏”这一准则，理论达到了完善。与此同时，人们也认识到，固守无偏性有时会导致荒谬的估计，当有限样本最优难以企及，渐近理论成为主流，优良性标准转向相合性和渐近效率。Fisher 最初关于 MLE 渐近有效的猜想，由 Cramér[3]、Wald[15]等人得到严格证明。Hájek[8]在 1970 年建立的卷积定理和局部渐近极小极大定理，为渐近有效估计的方差下界提供了决策论基础。1964 年 Huber[8]提出了稳健估计理论，当数据有微小污染时，对理想模型最优的估计可能会急剧恶化。这拓展了优良性的维度，要求在理想模型下的效率与模型偏离时的稳定性之间做出权衡。当参数维数远超样本量时，经典理论不再适用，1996 年 Tibshirani[14]提出 Lasso 等正则化方法，通过有偏估计换取方差的大幅降低，实现了变量选择和参数估计的同时进行。2001 年 Fan-Li[4]提出 SCAD 惩罚函数，2006 年 Zou[17]提出的自适应 Lasso，系统地研究并实现了变量选择的 Oracle 性质，重新定义了高维下的渐近有效性。2014 年前后，Zhang-Zhang[16]等人发展的去偏 Lasso 技术，构造了渐近正态的估计量，从而使得在高维下进行单参数的假设检验和构造置信区间成为可能，这极大地拓展了 Hájek 推断理论的边界。这些参数估计的新理论提出和技术运用，使得参数估计从寻找单一“最优”解，逐步演变为一个在偏差与方差、效率与稳健、精度与隐私、最优性与可计算性等多重约束下进行适应性选择的复杂决策体系。“优良性”的内涵本身，也在这一历程中被不断丰富和重新定义。

鉴于此，在目前数理统计的教材中虽有个别分布的优良性结果，但缺少系统性横向比较。因此，在本文中，我们将四类常见分布，包括二项分布，泊松分布，正态分布，指数分布的参数估计方法纳入统一分析框架，并进行具体分析参数估计的优良性。首先，我们先给出一些参数估计的基本定义和优良性判定。

1.2 预备知识

定义 1.1[1]：设X_1，X_2，…，X_n为来自总体X，分布为$F(x;\theta)$的一组简单随机样本，$\hat{\theta}_n=\hat{\theta}_n(X_1,X_2,\cdots,X_n)$为未知数$\theta$的一个估计量，称$Bias(\hat{\theta}_n)\triangleq E(\hat{\theta}_n)-\theta$为参数$\theta$估计的偏差. 若$Bias(\hat{\theta}_n)=0$，即$E(\hat{\theta}_n)=\theta$，则称$\hat{\theta}_n$为参数$\theta$的无偏估计。 若$\lim\limits_{n\to\infty}E(\hat{\theta}_n)=\theta$，则称$\hat{\theta}_n$为参数$\theta$的渐进无偏估计。

定义 1.2[1]：设$\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$，$\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$都是$\theta$的无偏估计量，若有$Var(\hat{\theta}_1)<Var(\hat{\theta}_2)$，则称$\hat{\theta}_1$较$\hat{\theta}_2$更有效。

定义 1.3[1]：设$\hat{\theta}_n=\hat{\theta}(X_1,X_2,\cdots,X_n)$为未知数$\theta$的一个估计量，如果当$n\to\infty$时，$\hat{\theta}_n$依概率收敛于$\theta$，即任给$\varepsilon>0$，有$\lim\limits_{n\to\infty}P\{|\hat{\theta}_n-\theta|<\varepsilon\}=1$，则称$\hat{\theta}_n$为参数$\theta$的相合估计。

定理 1.1[2]：$\hat{\theta}_n=\hat{\theta}_n(X_1,X_2,\cdots,X_n)$为未知参数$\theta$的估计量，且有：

$$\lim_{n\to\infty}E(\hat{\theta}_n)=\theta,\quad \lim_{n\to\infty}Var(\hat{\theta}_n)=0,$$

即$\hat{\theta}_n$为渐进无偏估计且方差趋于 0，则$\hat{\theta}_n$为未知参数θ的相合估计。

2 常见分布的参数估计的其优良性

2.1 二项分布的参数估计

命题 2.1.1：二项分布$X\sim B(n,p)$，p为未知参数，最大似然估计$\hat{p}_{MLE}=\frac{\sum_{i=1}^{N}X_i}{Nn}$和矩估计$\hat{p}_{ME}=\frac{\sum_{i=1}^{N}X_i}{nN}$。

证明：随机变量$X\sim B(n，p)$，p为未知参数，取总体X的样本X_1，X_2，…，X_N，它们独立同分布于$B(n，p)$，样本容量为N，则似然函数为：

$$L(p)=\prod_{i=1}^{N}\binom{n}{X_i}p^{X_i}(1-p)^{n-X_i}，\ i=1，2，\cdots，N$$

对数似然函数为：

$$lnL(p)=\sum_{i=1}^{N}ln\binom{n}{X_i}+X_i lnp+(n-X_i)\ln(1-p)$$

求解：

$$\frac{d}{dp}lnL(p)=\sum_{i=1}^{N}\frac{X_i}{p}-\frac{n-X_i}{1-p}=0$$

解得：

$$\hat{p}_{MLE}=\frac{\sum_{i=1}^{N}X_i}{nN}$$

对于矩估计，由$\mu_1=EX=np$，则令：

$$EX=np=\bar{X}=\frac{\sum_{i=1}^{N}X_i}{N}$$

解得p的矩估计为：

$$\hat{p}_{ME}=\frac{\bar{X}}{n}=\frac{\sum_{i=1}^{N}X_i}{nN}$$

命题 2.1.2 在二项分布中，最大似然估计$\hat{p}_{MLE}$和矩估计$\hat{p}_{ME}$均是无偏估计量。

证明 由$\hat{p}_{MLE}=\hat{p}_{ME}=\frac{\sum_{i=1}^{N}X_i}{Nn}$，则：

$$E(\hat{p}_{MLE})=E(\hat{p}_{ME})=E\left(\frac{\sum_{i=1}^{N}X_i}{Nn}\right)=\frac{\sum_{i=1}^{N}E(X_i)}{E(Nn)}=\frac{Nnp}{Nn}=p$$

因此最大似然估计$\hat{p}_{MLE}$和矩估计$\hat{p}_{ME}$是二项分布的无偏估计量。

命题 2.1.3 在二项分布中，最大似然估计$\hat{p}_{MLE}$和矩估计$\hat{p}_{ME}$均是相合估计量。

证明：由

$$E(\hat{p}_{MLE})=E(\hat{p}_{ME})=p，\ \lim_{N\to\infty}Var(\hat{p}_{MLE})=\lim_{N\to\infty}Var(\hat{p}_{ME})=\lim_{N\to\infty}\frac{p(1-p)}{Nn}=0$$

则最大似然估计$\hat{p}_{MLE}$和矩估计$\hat{p}_{ME}$是二项分布的相合估计量。

2.2 泊松分布的参数估计

命题 2.2.1 泊松分布$X\sim P(\lambda)$，λ为未知参数，最大似然估计$\hat{\lambda}_{MLE}=\frac{\sum_{i=1}^{N}X_i}{N}$和矩估计$\hat{\lambda}_{ME}=\frac{\sum_{i=1}^{N}X_i}{N}$。

证明：随机变量$X\sim P(\lambda)$，λ为未知参数，取总体X的样本X_1，X_2，…，X_N，它们独立同分布于$P(\lambda)$，样本容量为N，则似然函数为：

$$L(\lambda)=\prod_{i=1}^{N}P（X_i；\ \lambda）=\prod_{i=1}^{N}\frac{e^{-\lambda}\lambda^{X_i}}{X_i!}，\ i=1，2，\cdots，N$$

对数似然函数为：

$$lnL（\lambda）= -N\lambda + \sum_{i=1}^{N} X_i\, ln(\lambda) - \sum_{i=1}^{N} ln \quad （X_i!）$$

求解：

$$\frac{d}{d\lambda} lnL(\lambda) = -N + \frac{\sum_{i=1}^{N} X_i}{\lambda} = 0$$

得到：

$$\hat{\lambda}_{MLE} = \frac{\sum_{i=1}^{N} X_i}{N}$$

对于矩估计，由$\mu_1 = EX = \lambda$，则令：

$$EX = \lambda = \bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

解上述方程，得λ的矩估计为：

$$\hat{\lambda}_{ME} = \bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

命题 2.2.2 在泊松分布中，最大似然估计$\hat{\lambda}_{MLE}$和矩估计$\hat{\lambda}_{ME}$是无偏估计量。

证明 由$\hat{\lambda}_{MLE} = \hat{\lambda}_{ME} = \bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$，则

$$E(\hat{\lambda}_{MLE}) = E(\hat{\lambda}_{ME}) = E\left(\frac{\sum_{i=1}^{N} X_i}{N}\right) = \frac{\sum_{i=1}^{N} E(X_i)}{N}$$

又因为$E(X_i) = \lambda$，所以：

$$E(\hat{\lambda}_{MLE}) = E(\hat{\lambda}_{ME}) = \frac{N\lambda}{N} = \lambda$$

因此，最大似然估计$\hat{\lambda}_{MLE}$和矩估计$\hat{\lambda}_{ME}$是泊松分布的无偏估计量。

命题 2.2.3 在泊松分布中，最大似然估计$\hat{\lambda}_{MLE}$和矩估计$\hat{\lambda}_{ME}$都是相合估计。

证明：由

$$E(\hat{\lambda}_{MLE}) = E(\hat{\lambda}_{ME}) = \lambda, \quad \lim_{N\to\infty} Var(\hat{\lambda}_{MLE}) = \lim_{N\to\infty} Var(\hat{\lambda}_{ME}) = \lim_{N\to\infty} \frac{\lambda}{N} = 0$$

因此，最大似然估计$\hat{\lambda}_{MLE}$和矩估计$\hat{\lambda}_{ME}$泊松分布的相合估计量。

2.3 正态分布的参数估计

命题 2.3.1 正态分布$X\sim N(\mu，\sigma^2)$，μ，σ^2为未知参数，最大似然估计$\hat{\mu}_{MLE}$和矩估计$\hat{\mu}_{ME} = \bar{X} = \hat{\mu}_{MLE}$，最大似然估计$\hat{\sigma}^2_{MLE}$和矩估计$\hat{\sigma}^2_{ME} = \frac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{N} = \hat{\sigma}^2_{MLE}$。

证明：随机变量$X\sim N(\mu，\sigma^2)$，μ，σ^2为未知参数，取总体X的样本X_1，X_2，…，X_N，它们独立同分布于$N(\mu，\sigma^2)$，样本容量为N，则似然函数为：

$$L(\mu，\sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i-\mu)^2\right\}$$

对数似然函数为：

$$lnL = -\frac{N}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i-\mu)^2$$

求解：

$$\begin{cases}\dfrac{\partial lnL}{\partial \mu}=\dfrac{1}{\sigma^2}\left(\sum\limits_{i=1}^{N}X_i-N\mu\right)=0,\\ \dfrac{\partial lnL}{\partial \sigma^2}=-\dfrac{N}{2\sigma^2}+\dfrac{1}{2(\sigma^2)^2}\sum\limits_{i=1}^{N}(X_i-\mu)^2=0,\end{cases}$$

得到：

$$\begin{cases}\hat{\mu}_{MLE}=\bar{X},\\ \hat{\sigma}^2{}_{MLE}=\dfrac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{N}\end{cases}$$

由$\mu_1=EX=\mu$，$\mu_2=EX^2=\mu^2+\sigma^2$，则：

$$\begin{cases}\mu=\bar{X},\\ \mu^2+\sigma^2=\dfrac{\sum_{i=1}^{N}X_i^2}{N}\end{cases}$$

解上述方程组，得μ和σ^2的矩估计分别为：

$$\hat{\mu}_{ME}=\bar{X},\ \ \hat{\sigma}^2_{ME}=\frac{\sum_{i=1}^{N}X_i^2-\bar{X}^2}{N}=\frac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{N}$$

命题 2.3.2 在正态分布中，最大似然估计$\hat{\mu}_{MLE}$和矩估计$\hat{\mu}_{ME}$是μ的无偏估计量，但最大似然估计$\hat{\sigma}^2_{MLE}$和矩估计$\hat{\sigma}^2_{ME}$是σ^2的有偏估计量。

证明：由$\hat{\mu}_{MLE}=\hat{\mu}_{ME}=\bar{X}$，$\hat{\sigma}^2_{MLE}=\hat{\sigma}^2_{ME}=\frac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{N}=\frac{(N-1)S^2}{N}$，其中$S^2=\frac{\sum_{i=1}^{N}（X_i-\bar{X}）^2}{N-1}$为样本方差，则：

$$E(\hat{\mu}_{MLE})=E(\hat{\mu}_{ME})=E(\bar{X})=\mu$$

又因为由于$\frac{(N-1)S^2}{\sigma^2}$服从自由度为$N-1$的卡方分布，则$E(S^2)=\sigma^2$，$E(\hat{\sigma}^2_{MLE})=E(\hat{\sigma}^2_{ME})=E\left(\frac{(N-1)S^2}{N}\right)=\frac{(N-1)\sigma^2}{N}\neq\sigma^2$，因此，最大似然估计$\hat{\mu}_{MLE}$和矩估计$\hat{\mu}_{ME}$是正态分布中$\mu$的无偏估计量，但最大似然估计$\hat{\sigma}^2_{MLE}$和矩估计$\hat{\sigma}^2_{ME}$是正态分布中$\sigma^2$的有偏估计量。

命题 2.3.3 在正态分布中，最大似然估计$\hat{\mu}_{MLE}$，$\hat{\sigma}^2_{MLE}$和矩估计$\hat{\mu}_{ME}$，$\hat{\sigma}^2_{ME}$是相合估计量。

证明 由$E(\hat{\mu}_{MLE})=E(\hat{\mu}_{ME})=E(\bar{X})=\mu$，则：

$$\lim_{N\to\infty}E(\hat{\sigma}^2_{MLE})=\lim_{N\to\infty}E(\hat{\sigma}^2_{ME})=\lim_{N\to\infty}\frac{(N-1)\sigma^2}{N}=\sigma^2$$

又因为：

$$Var(\hat{\sigma}^2_{MLE})=Var(\hat{\sigma}^2_{ME})=Var\left(\frac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{N}\right)=Var\left(\frac{(N-1)S^2}{N}\right)$$

$$=\frac{(N-1)^2}{N^2}VarS^2=\frac{2\sigma^4(N-1)^2}{N^2(N-1)}=\frac{2\sigma^4(N-1)}{N^2}$$

则有：

$$\lim_{N\to\infty}Var(\hat{\mu}_{MLE})=\lim_{N\to\infty}Var(\hat{\mu}_{ME})=0$$

$$\lim_{N\to\infty}Var(\hat{\sigma}^2_{MLE})=\lim_{N\to\infty}Var(\hat{\sigma}^2_{ME})=0$$

因此，最大似然估计$\hat{\mu}_{MLE}$，$\hat{\sigma}^2_{MLE}$和矩估计$\hat{\mu}_{ME}$，$\hat{\sigma}^2_{ME}$是正态分布的相合估计量。

2.4 指数分布的参数估计

命题 2.4.1 指数分布$X\sim Exp(\lambda)$，λ为未知参数，最大似然估计$\hat{\lambda}_{MLE}=\frac{\sum_{i=1}^{N}X_i}{N}$和矩估计$\hat{\lambda}_{ME}=\frac{\mathrm{N}}{\sum_{i=1}^{N}X_i}$.

证明 随机变量$X\sim Exp(\lambda)$，λ为未知参数，取总体X的样本设X_1，X_2，…，X_n，它们独立同分布于$Exp(\lambda)$，样本容量为N，则似然函数为：

$$L(\lambda)=\prod_{i=1}^{N}\lambda e^{-\lambda X_i}$$

对数似然函数为：

$$lnL(\lambda)=-\lambda\sum_{i=1}^{N}X_i+Nln(\lambda)$$

求解：

$$\frac{dlnL(\lambda)}{d\lambda}=-\sum_{i=1}^{N}X_i+N\frac{1}{\lambda}=0$$

得到：

$$\hat{\lambda}_{MLE}=\frac{1}{\bar{X}}=\frac{\mathrm{N}}{\sum_{i=1}^{N}X_i}$$

由$EX=\frac{1}{\lambda}$，则令：

$$EX=\frac{1}{\lambda}=\bar{X}=\frac{\sum_{i=1}^{N}X_i}{N}$$

求解λ的矩估计为：

$$\hat{\lambda}_{ME}=\frac{1}{\bar{X}}=\frac{\mathrm{N}}{\sum_{i=1}^{N}X_i}$$

命题 2.4.2 在指数分布中，最大似然估计$\hat{\lambda}_{MLE}$和矩估计$\hat{\lambda}_{ME}$是有偏估计量。

证明：由$\hat{\lambda}_{MLE}=\hat{\lambda}_{ME}=\frac{1}{\bar{X}}=\frac{\mathrm{N}}{\sum_{i=1}^{N}X_i}$，令$Y_N=X_1+X_2+\cdots+X_N$，又$X\sim Exp(\lambda)=\Gamma$（$\lambda$，1），由Γ分布的可加性，则$Y_N\sim\Gamma$（$\lambda$，N），且：

$$f_{Y_N}(x)=\begin{cases}\dfrac{\lambda^N}{\Gamma(N)}x^{N-1}e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$$

令$Z_N=\frac{1}{Y_N}$，则$F_{Z_N}(z)=P(Z_N\le z)=P\left(\frac{1}{Y_N}\le z\right)$，由$X_1$，$X_2$，…，$X_N>0$，则$Y_N>0$，故$z\le 0$时，$F_{Z_N}(z)=P\left(\frac{1}{Y_N}\le z\right)=P(\emptyset)=0$；$z>0$时，$F_{Z_N}(z)=P(Y_N\ge\frac{1}{z})=\int_{\frac{1}{z}}^{+\infty}f_{Y_N}(x)\,dx=\int_{\frac{1}{z}}^{+\infty}\frac{\lambda^N}{\Gamma(N)}x^{N-1}e^{-\lambda x}\,dx$。

因此：

$$f_{Z_N}(z)=\begin{cases}e^{-\frac{\lambda}{z}}\dfrac{\lambda^N}{\Gamma(N)z^{N+1}}, & z>0\\ 0, & z\le 0\end{cases}$$

则有：

$$E\left(\hat{\lambda}_{MLE}\right)=E\left(\hat{\lambda}_{ME}\right)=E\left(\frac{N}{\sum_{i=1}^{N}X_i}\right)=NE(Z_N)=N\int_{-\infty}^{+\infty}zf_{Z_N}(z)\,dz$$

$$= N\int_0^{+\infty} e^{-\frac{\lambda}{z}}\frac{\lambda^N}{\Gamma(N)}\frac{1}{z^N}dz = N\lambda\frac{\Gamma(N-1)}{\Gamma(N)} = \frac{N\lambda}{N-1}$$

因此，最大似然估计$\hat{\lambda}_{MLE}$和矩估计$\hat{\lambda}_{ME}$是指数分布的有偏估计量。

命题 2.4.3 在指数分布中，最大似然估计$\hat{\lambda}_{MLE}$和矩估计$\hat{\lambda}_{ME}$是相合估计量。

证明：由$\hat{\lambda}_{MLE} = \hat{\lambda}_{ME} = \frac{1}{\bar{X}} = \frac{N}{\sum_{i=1}^{N} X_i}$，$E\left(\hat{\lambda}_{MLE}{}^2\right) = E\left(\hat{\lambda}_{ME}{}^2\right) = N^2E\left(Z_N{}^2\right) = N^2\int_{-\infty}^{+\infty} z^2 f_{Z_N}(z)\,dz = N^2\int_0^{+\infty} e^{-\frac{\lambda}{z}}\frac{\lambda^N}{\Gamma(N)}\frac{1}{z^{N-1}}dz = N^2\lambda^2\frac{\Gamma(N-2)}{\Gamma(N)} = \frac{N^2\lambda^2}{(N-1)(N-2)}$，则有：$Var\left(\hat{\lambda}_{ME}\right) = E\left(\hat{\lambda}_{ME}{}^2\right) - E\left(\hat{\lambda}_{ME}\right)^2 = \frac{N^2\lambda^2}{(N-1)(N-2)} - \frac{N^2\lambda^2}{(N-1)^2}$，此外：$\lim_{N\to\infty} E\left(\hat{\lambda}_{ME}\right) = \lim_{N\to\infty}\frac{N\lambda}{N-1} = \lambda$，$\lim_{N\to\infty} Var\left(\hat{\lambda}_{ME}\right) = \lim_{N\to\infty}\frac{N^2\lambda^2}{(N-1)(N-2)} - \frac{N^2\lambda^2}{(N-1)^2} = 0$，因此，$\hat{\lambda}_{MLE}$和$\hat{\lambda}_{ME}$是指数分布的相合估计量。

3 结论与展望

结合上一节内容，我们进行对比不同分布的参数估计的优良性结果。首先，在二项分布$X\sim B(n,\ p)$，参数 p 的 MLE 与 ME 相同，均为无偏且相合的。在泊松分布$X\sim P(\lambda)$中，参数 λ 的 MLE 与 ME 相同，均为无偏且相合的。从而在实际问题中，符合二项分布和泊松分布的模型，可以采用收集大量样本，计算样本矩来近似代替参数真值。

在正态分布$X\sim N(\mu,\ \sigma^2)$中，参数 μ 的 MLE 与 ME 相同，均为无偏且相合的；参数σ^2的 MLE 与 ME 相同，均为相合但有偏的，实际上有偏性，是因为在估计方差时，我们不知道真实的均值 μ，只能用样本均值$\overline{X}$去替代，而这个替代过程“消耗”了一个自由度，导致估计值系统性地偏小。另一方面，样本方差$S^2 = \sum_{i=1}^{n}(X_i - \overline{X_i})^2 \Big/ n-1$，且$E(S^2) = \sigma^2$，$S^2$就是参数$\sigma^2$的一个无偏估计。若进一步计算均方误差，我们发现：

$$Var(\sigma^2{}_{MLE}) = Var(\sigma^2{}_{ME}) = Var(\sum_{i=1}^{n}(X_i - \overline{X_i})^2 \Big/ n) < Var(S^2)$$

即$\sigma^2{}_{MLE} = \sigma^2{}_{ME}$的均方误差比样本方差$S^2$的均方误差来的小，因此参数$\sigma^2$的 MLE 与 ME 虽然是有偏的，但更有效，更稳定；因此在样本量较小时，我们优先选择作$\sigma^2{}_{MLE} = \sigma^2{}_{ME}$为参数$\sigma^2$的估计，而当样本量 N 足够大时，样本方差S^2与$\sigma^2{}_{MLE} = \sigma^2{}_{ME}$相差不大，也可作为参数$\sigma^2$的估计。

在指数分布$X\sim Exp(\lambda)$中，参数λ的 MLE 和 ME 相同，均是是相合但有偏的。根据命题 2.4.2，有偏性是指系统性地偏大，这主要是因为$\hat{\lambda}_{MLE} = \hat{\lambda}_{ME} = 1/\overline{X}$是样本均值的倒数，利用凸函数的性质，有 Jensen 不等式$E(\hat{\lambda}_{MLE}) = E(\hat{\lambda}_{ME}) = E(1/\overline{X}) \geq 1/E(\overline{X}) = \lambda$,从而参数$\lambda$的 MLE 和 ME 是有偏估计。此外若对$\hat{\lambda}_{MLE} = \hat{\lambda}_{ME}$进行系数修正（乘以$N-1/N$），则$E(\hat{\lambda}_{MLE}\ N-1/N) = E(\hat{\lambda}_{MLE}\ N-1/N) = \lambda$，且：

$$Var(\frac{N-1}{N}\hat{\lambda}_{MLE}) = E\left[(\frac{N-1}{N}\hat{\lambda}_{MLE})^2\right] - \left[E(\frac{N-1}{N}\hat{\lambda}_{MLE})\right]^2 = \frac{1}{N-2}\lambda^2 < Var(\hat{\lambda}_{MLE})$$

故$\frac{N-1}{N}\hat{\lambda}_{MLE}$是一个无偏估计，且比$\hat{\lambda}_{MLE}$更稳定有效。这在现有教材往往仅给出结论而缺少细节推导。本文旨在填补这一教学与参考层面的空缺。此外，在进行参数优良性判定时，不能只看一个维度，需要把无偏性，稳定性，相合性结合在一起。在以上例子中，从均方误差角度看，有些参数估计是有偏的但是更稳定，有的则相反，因此“有偏估计与无偏估计的权衡”是一个复杂更深层次的问题，值得进一步探讨。

参考文献

[1] 茆诗松, 程依明, 濮晓龙, 概率论与数理统计教程. 北京:高等教育出版社, (2019).

[2] 陈希孺, 倪国熙, 数理统计学教程. 安徽:中国科学技术大学出版社, (2009).

[3] Cramér, H., Mathematical Methods of Statistics. Princeton University Press. (1999).

[4] Fan, J., Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96(456), 1348-1360. (2001).

[5] Fisher, R. A., On an absolute criterion for fitting frequency curves. Messenger of Mathematics, 41, 155-160, (1912).

[6] Fisher, R. A., On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London, Series A, 222, 309-368, (1922).

[7] Fisher, R. A., Theory of statistical estimation. Mathematical Proceedings of the Cambridge Philosophical Society, 22(5), 700-725, (1925).

[8] Hájek, J., A characterization of limiting distributions of regular estimates. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 14(4), 323-330, (1970).

[9] Huber, P. J., Robust estimation of a location parameter. The Annals of Mathematical Statistics, 35(1), 73-101. (1964).

[10] Lehmann, E. L., Scheffé, H., Completeness, similar regions, and unbiased estimation. Sankhyā: The Indian Journal of Statistics, 10(4), 305-340, (1950)

[11] Lehmann, E. L., Scheffé, H. Completeness, similar regions, and unbiased estimation: Part II. Sankhyā: The Indian Journal of Statistics, 15(3), 219-236, (1955).

[12] Pearson, K., Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London, Series A, 185, 71-110, (1894).

[13] Rao, C. R., Information and accuracy attainable in the estimation of statistical parameters. Bulletin of the Calcutta Mathematical Society, 37, 81-91, (1945).

[14] Tibshirani, R., Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267-288, (1996).

[15] Wald, A., Note on the consistency of the maximum likelihood estimate. The Annals of Mathematical Statistics, 20(4), 595-601, (1949)

[16] Zhang, C. H., Zhang, S. S., Confidence intervals for low dimensional parameters in high dimensional linear models. Journal of the Royal Statistical Society: Series B, 76(1), 217-242, (2014).

[17] Zou, H., The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101(476), 1418-1429, (2006).