概率论与数理统计第三次习题课题目解答

题 1 $\frac{4}{3}$

题 2 $\frac{1}{3}$

题 3 见书本习题3.4.32

题 4 将编号为 1 至 n 的 n 个球随机投入编号为 1 至 n 的 n 个盒子中，并限制每一个盒子中只能放入一个球，设球与盒子的号码一致的个数为 S_n ，求证：

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0, \quad n \to \infty$$

证法 1. 记

$$X_k = \begin{cases} 1, & \text{如果编号为 } k \text{ 的球被放入编号为 } k \text{ 的盒子} \\ 0, & \text{否则} \end{cases}$$

则

$$S_n = X_1 + \cdots + X_n,$$

于是对任意 $\varepsilon > 0$

$$P\left(\left|\frac{S_n - ES_n}{n}\right| > \varepsilon\right) \leq \frac{\mathrm{Var}S_n}{n^2\varepsilon^2} = \frac{1}{n^2\varepsilon^2}\left(\sum_{i=1}^{n}\mathrm{Var}(X_i) + \sum_{1\leq i\neq j\leq n}\mathrm{Cov}(X_i, X_j)\right).$$

而

$$P(X_k = 1) = \frac{1}{n}, \quad P(X_i = 1, X_j = 1) = \frac{1}{n(n-1)}, \quad \forall 1 \leq k \leq n, \quad \forall 1 \leq i \neq j \leq n,$$

故

$$EX_k = \frac{1}{n}, \quad \mathrm{Var}X_k = \frac{n-1}{n^2}, \quad \mathrm{Cov}(X_k, X_l) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}.$$

所以，

$$\frac{1}{n^2\varepsilon^2}\left(\sum_{i=1}^{n}\mathrm{Var}(X_i) + \sum_{1\leq i\neq j\leq n}\mathrm{Cov}(X_i, X_j)\right) = \frac{1}{n^2\varepsilon^2}\left(\frac{n-1}{n} + \frac{1}{n}\right) \to 0, \quad n \to \infty.$$

因此 $\frac{S_n - ES_n}{n} \xrightarrow{P} 0, \quad n \to \infty$。 □

证法 2. 符号同上，因为

$$\mathrm{Cov}(X_k, X_l) \leq \sqrt{\mathrm{Var}X_k} \cdot \sqrt{\mathrm{Var}X_l} = \frac{n-1}{n^2}$$

由证法 1

$$\begin{aligned} P\left(\left|\frac{S_n - ES_n}{n}\right| > \varepsilon\right) &\leq \frac{1}{n^2\varepsilon^2}\left(\sum_{i=1}^{n}\mathrm{Var}(X_i) + \sum_{1\leq i\neq j\leq n}\mathrm{Cov}(X_i, X_j)\right) \\ &\leq \frac{1}{n^2\varepsilon^2}\left(n \cdot \frac{n-1}{n^2} + n(n-1) \cdot \frac{n-1}{n^2}\right) \\ &= \frac{n-1}{n^2\varepsilon^2} \to 0, \quad n \to \infty. \end{aligned}$$

结论成立。 □

<u>证法 3</u>. 仍然是对 S_n 用 Chebyshev 不等式。由于 $0 \le S_n \le n$ 恒成立，所以

$$\mathrm{Var}(S_n) \le ES_n^2 \le nE(S_n) = n,$$

因此

$$\frac{\mathrm{Var}(S_n)}{n^2} \le \frac{1}{n} \to 0, \quad n \to +\infty.$$

□

<u>证法 4</u>. 由于 $ES_n = 1$，所以

$$\frac{ES_n}{n} = \frac{1}{n} \to 0, \quad n \to \infty.$$

由于 $S_n \ge 0$，所以对任何 $\varepsilon > 0$，由 Markov 不等式

$$P\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) = P\left(\frac{S_n}{n} > \varepsilon\right) \le \frac{1}{\varepsilon}E\left(\frac{S_n}{n}\right) = \frac{1}{n\varepsilon} \to 0, \quad n \to \infty.$$

所以

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0, \quad n \to +\infty.$$

□

题 5 一报贩发现每个路过他的报摊的行人向他买报的概率为 $\frac{1}{3}$ 。求在他出售了 100 份报纸时的过路人的数目在 280 人到 320 人之间的概率。（用两种不同的估计方法，并比较它们的优劣）

<u>解法 1</u>. 记

$$Y_n = \begin{cases} 1, & \text{若第 } n \text{ 个人买报纸}, \\ 0, & \text{否则} \end{cases}$$

于是 $Y_1, Y_2, \ldots, Y_n, \ldots$ 独立同分布，服从 Bernoulli 分布 $B(1, 1/3)$。

买到第 100 份报纸的行人的序号应该是

$$N = \min\{n : Y_1 + \cdots + Y_n = 100\},$$

它是个随机变量。

由 N 的定义，我们知道，事件 $\{N > m\}$ 就是“前 m 个人都没有买到第 100 份报纸”，也就是

$$Y_1 + \cdots + Y_m < 100.$$

由中心极限定理，

$$U_n = \frac{Y_1 + \cdots + Y_n - \frac{n}{3}}{\sqrt{\frac{2n}{9}}} = \frac{3(Y_1 + \ldots + Y_n) - n}{\sqrt{2n}} \dot{\sim} N(0, 1),$$

注意到

$$\{280 \le N \le 320\} = \{N > 279\} - \{N > 320\},$$

所以

$$
\begin{aligned}
&P(280 \leq N \leq 320) \\
=&P\left(Y_1+Y_2+\cdots+Y_{279}<100\right)-P\left(Y_1+\cdots+Y_{320}<100\right) \\
=&P\left(U_{279}<\frac{99.5-93}{\sqrt{62}}\right)-P\left(U_{320}<-\frac{3\times 99.5-320}{8\sqrt{10}}\right) \quad \text{这里为什么做 0.5 修正？} \\
\approx&\Phi\left(0.826\right)-\Phi\left(-0.850\right) \\
\approx&0.5981.
\end{aligned}
$$

□

解法 2. “买，还是不买，这的确是个问题。”每个路过报摊的行人都在**相互独立地重复着**这样的抉择，这是一个 Bernoulli试验。则第 k张报纸到底是卖给了第几个路经报摊的行人，那个行人的序号 X_k 服从的正是负二项概率分布，在这样的 Bernoulli试验中相继卖出任何两份报纸之间，路过的行人数 $Z_1, Z_2, \ldots, Z_k, \ldots$ 相互独立，都服从参数为 1/3 的几何分布，并且 $X_k=\sum_{j=1}^{k} Z_j$。

采用 Chebyshev不等式，由于

$$
E(Z_1+\cdots+Z_{100})=100\times 3=300, \quad \operatorname{Var}(Z_1+\cdots+Z_{100})=100\times 3^2\times\frac{2}{3}=600,
$$

所以用 Chebyshev 不等式估计上述概率，那么

$$
\begin{aligned}
&P\left(280 \leq Z_1+Z_2+\cdots+Z_{100} \leq 320\right) \\
=&1-P\left(|Z_1+\cdots+Z_{100}-E(Z_1+\cdots+Z_{100})| \geq 20\right) \\
\geq&1-\frac{\operatorname{Var}(Z_1+\cdots+Z_{100})}{20^2}=1-\frac{6}{4}=-\frac{1}{2},
\end{aligned}
$$

结果估计十分粗糙，采用中心极限定理进行估计，有

$$
\begin{aligned}
&P\left(280 \leq Z_1+Z_2+\cdots+Z_{100} \leq 320\right) \\
=&P\left(\frac{280-300}{\sqrt{600}} \leq \frac{Z_1+\cdots+Z_{100}-E(Z_1+\cdots+Z_{100})}{\sqrt{\operatorname{Var}(X_1+\cdots+X_{100})}} \leq \frac{320-300}{\sqrt{600}}\right) \\
=&2\Phi\left(\frac{20.5}{\sqrt{600}}\right)-1=2\Phi(0.837)-1=0.5973.
\end{aligned}
$$

□

题6 设某城市有 N 辆机动车，牌号依次是 $1,2,\cdots,N$ 。一个人将他一天内看到的所有机动车牌号（包括重复出现的牌号）都记录下来，得到 $X_1, X_2, \cdots, X_n$ 。如果用最大牌号 $X_{(n)}$ 作为对N的一个估计（即近似值），我们采取以下方式来评价这个估计：

1. 当n充分大时，$X_{(n)}$是否近似等于N？并且试证明 $X_{(n)}$ 是 N 的 MLE
2. 试给出 N 的一个矩估计，并与其 MLE, 即 $X_{(n)}$ 进行比较。
3. 如果这样的观察方式被多次重复进行，每次得到$X_{(n)}$的一个观测值，那么根据大数定律，$X_{(n)}$观测值的算术平均值将以$EX_{(n)}$为极限，求$EX_{(n)}-N$（称为这种近似方式的“偏”，即系统误差）的值。

4. 如果$X_{(n)}$存在系统误差（有偏，即$EX_{(n)}-N\neq 0$），那么你有什么办法可以消除这个系统误差？

如果不重复记录的话，如何用观测值$X_1,X_2,\cdots,X_n$给出 N 的一个估计？分析你给出的估计的性质，并与重复情况下的估计进行比较。

解. 重复记录的情形：

(a) 记$X_1,\ldots,X_n$独立同分布，都服从$\{1,2,\ldots,N\}$上的离散均匀分布。于是

$$P(X_{(n)}\leq x)=P(X_1\leq x,X_2\leq x,\ldots,X_n\leq x)=\left[P(X_1\leq x)\right]^n=\frac{x^n}{N^n},\quad x=1,2\ldots,N.$$

从而

$$P(X_{(n)}=N)=P(X_{(n)}\leq N)-P(X_{(n)}\leq N-1)=1-\left(\frac{N-1}{N}\right)^n\to 1,\quad n\to\infty.$$

因此$X_{(n)}$依概率收敛于N。不仅如此，事实上，由于

$$P(X_{(n)}<N)=P(X_{(n)}\leq N-1)=\left(\frac{N-1}{N}\right)^n,\quad \forall n\geq 1,$$

所以

$$P\left(\bigcup_{n\geq k}\{X_{(n)}<N\}\right)\leq\sum_{n\geq k}P(X_{(n)}<N)=\sum_{n\geq k}\left(\frac{N-1}{N}\right)^n=N\left(\frac{N-1}{N}\right)^k\to 0,\quad k\to+\infty,$$

从而

$$P\left(\bigcap_{k\geq 1}\bigcup_{n\geq k}\{X_{(n)}<N\}\right)=0,$$

这说明，事件“从某个$k\geq 1$开始，对任意$n\geq k$，最大值$X_{(n)}=N$”以概率1发生，因此$X_{(n)}$以概率1收敛于N。

为证明 $X_{(n)}$ 是 N 的 MLE, 考虑如下似然函数：

$$L(N,\ x_1,\ x_2,\ \cdots,x_n)=P\left(X_1=x_1,\ X_2=x_2,\ \cdots,X_n=x_n\right)=\left(\frac{1}{N}\right)^n\cdot I_{1\leq x_{(1)}\cdots\leq x_{(n)}\leq N}$$

观察发现:当 N 向下趋于 $x_{(n)}$ 时， L 递增. 所以 $X_{(n)}$ 是 N 的 $MLE.$

(b)

$$E[X]=\sum_{k=1}^{N}k\cdot\frac{1}{N}=\frac{1}{N}\sum_{k=1}^{N}k=\frac{1+N}{2}$$

所以 N 的矩估计可写为 $2\bar{X}-1$. 这显然是一个无偏估计，而且由辛钦大数定律知该矩估计强相合于 N.

(c)由于

$$P(X_{(n)}=x)=P(X_{(n)}\leq x)-P(X_{(n)}\leq x-1)=\frac{x^n-(x-1)^n}{N^n},\quad x=1,2,\ldots,N,$$

故

$$
\begin{aligned}
EX_{(n)} &= \sum_{x=1}^{N} xP(X_{(n)}=x) = \sum_{x=1}^{N} x\frac{x^n-(x-1)^n}{N^n} \\
&= \sum_{x=1}^{N} \frac{x^{n+1}-(x-1)^{n+1}-(x-1)^n}{N^n} \\
&= N\sum_{x=1}^{N} \frac{x^{n+1}-(x-1)^{n+1}}{N^{n+1}} - \sum_{x=1}^{N}\frac{(x-1)^n}{N^n} \\
&= N - \sum_{x=1}^{N}\frac{(x-1)^n}{N^n}.
\end{aligned}
$$

另外一个办法是：

$$
\begin{aligned}
EX_{(n)} &= \sum_{x=0}^{+\infty} P(X_{(n)}>x) \\
&= \sum_{x=0}^{N-1}\left(1-\frac{x^n}{N^n}\right) \\
&= N - \sum_{x=1}^{N}\frac{(x-1)^n}{N^n}.
\end{aligned}
$$

注意，虽然

$$
E\left(X_{(n)} + \sum_{x=1}^{N}\frac{(x-1)^n}{N^n}\right) = N,
$$

但$X_{(n)} + \sum\limits_{x=1}^{N}\frac{(x-1)^n}{N^n}$不是$N$的无偏估计，因为它不是统计量（它仍然依赖未知参数$N$）。

由于

$$
\sum_{x=1}^{N}\frac{(x-1)^n}{N^n}\cdot\frac{1}{N} \to \int_0^1 t^n dt = \frac{1}{n+1}, \quad N\to\infty,
$$

于是当N充分大时，

$$
EX_{(n)} \approx N - \frac{N}{n+1} = \frac{n}{n+1}N.
$$

这表明用$X_{(n)}$作为N的近似值，这种近似方法是存在系统误差的，因为这样的近似值平均意义下总是比真值小。

(d) 我们考虑全体顺序统计量

$$
X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}.
$$

由于

$$
P(X_{(k)}>x) = \sum_{j=0}^{k-1}\binom{n}{j}\frac{x^j(N-x)^{n-j}}{N^n},
$$

故

$$
EX_{(k)} = \sum_{x=0}^{\infty} P(X_{(k)}>x) = \sum_{x=0}^{N}\sum_{j=0}^{k-1}\binom{n}{j}\frac{x^j(N-x)^{n-j}}{N^n},
$$

于是

$$EX_{(1)} = \sum_{x=0}^{N} \frac{(N-x)^n}{N^n},$$

从而

$$EX_{(n)} + EX_{(1)} - 1 = N,$$

而

$$P(X_{(1)} = 1) = 1 - P(X_{(1)} > 1) = 1 - \left(\frac{N-1}{N}\right)^n,$$

因此可以类似证明$X_{(1)}$依概率收敛/以概率1收敛于1。从而，$X_{(1)} + X_{(n)} - 1$依概率收敛/以概率1收敛于N，并且$E\big(X_{(1)} + X_{(n)} - 1\big) = N$。这表明作为$N$的近似值，$X_{(1)} + X_{(n)} - 1$没有系统误差（无偏）。

事实上有一个更简便的方法，由于离散均匀分布的对称性，X与$N+1-X$具有相同的概率分布，所以$N+1-X_{(1)} = \max\limits_{1\le i\le n}\{N+1-X_i\}$与$X_{(n)}$具有相同的概率分布，因此

$$E(N+1-X_{(1)}) = E(X_{(n)}),$$

从而

$$E(X_{(n)} + X_{(1)} - 1) = N,$$

即$X_{(n)} + X_{(1)} - 1$是N的无偏估计。其直观含义是，我们试图用$X_{(1)}$到左端点1的距离去弥补$X_{(n)}$到N的距离。

不重复记录的情形： 这时$X_1, X_2, \ldots, X_n$的联合分布为

$$P(X_1 = x_1, \ldots, X_n = x_n) = \frac{(N-n)!}{N!}, \quad x_1, \ldots, x_n \in \{1, \ldots, N\}, \quad x_i \ne x_j (\forall i \ne j).$$

顺序统计量$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$的联合分布为

$$P(X_{(1)} = x_1, X_{(2)} = x_2, \ldots, X_{(n)} = x_n) = n!\frac{(N-n)!}{N!} = \frac{1}{\binom{N}{n}},$$
$$1 \le x_1 < x_2 < \cdots < x_n \le N.$$

$X_{(n)}$的概率分布为

$$P(X_{(n)} \le x) = \frac{\binom{x}{n}}{\binom{N}{n}}, \quad x = n, n+1, \ldots, N.$$

于是

$$P(X_{(n)} = x) = \frac{\binom{x-1}{n-1}}{\binom{N}{n}}, \quad x = n, n+1, \ldots, N$$

从而

$$EX_{(n)} = \sum_{k=n}^{N} k\frac{\binom{k-1}{n-1}}{\binom{N}{n}} = \frac{n(N+1)}{n+1}\sum_{k=n+1}^{N+1}\frac{\binom{k+1-1}{n+1-1}}{\binom{N+1}{n+1}} = \frac{n}{n+1}(N+1) < N, \quad \forall n < N.$$

这表明作为N的近似值，$X_{(n)}$存在系统误差（有偏）。

我们考虑

$$X_{(n)} + \frac{1}{n}\sum_{i=1}^{n}(X_{(i)} - X_{(i-1)} - 1) = \frac{n+1}{n}X_{(n)} - 1, \quad X_{(0)} = 0.$$

则

$$E\left(\frac{n+1}{n}X_{(n)}-1\right)=\frac{n+1}{n}\cdot\frac{n}{n+1}(N+1)-1=N.$$

这表明作为N的近似值，$\frac{n+1}{n}X_{(n)}-1$没有系统误差（无偏）。

<u>无论重不重复，都可以构造无偏的矩估计</u>

因为两种情况下都有

$$EX=\sum_{k=1}^{N}\frac{k}{N}=\frac{N+1}{2}$$

从而得到N的矩估计，

$$\widehat{N}=2\overline{X}-1.$$

这个矩估计是无偏的。

但是矩估计有其自身的局限性，比如，如果样本值为$2,4,6,100$，矩估计给出的结果为$\widehat{N}=55$，这显然无法解释样本值中的100，而前面两个估计方法给出的N的近似值分别是101和124，都没有矩估计遇到的矛盾。所以选择那种估计办法要根据具体问题作具体分析。当然，读者可以自己比较一下上述无偏估计的方差。 □

附加题

题7 设总体分布为$U[\theta-1,\theta+1]$，其中θ是未知参数。设$X_1,\dots,X_n$是来自该总体的简单随机样本。

1. 求θ的矩估计量$\hat{\theta}$，判断它的相合性和无偏性，计算均方误差$\mathrm{MSE}(\hat{\theta})$；
2. 证明对任何$0\le t\le 1$，$\hat{\theta}_t:=tX_{(n)}+(1-t)X_{(1)}+1-2t$都是$\theta$的极大似然估计量；此例表明极大似然估计可以不唯一。
3. 求$X_{(1)}$和$X_{(n)}$的概率分布以及数学期望$EX_{(1)}$、$EX_{(n)}$；
4. 问$\hat{\theta}_t$是否为θ的相合估计和无偏估计？
5. 求$X_{(1)},X_{(n)}$的联合分布，以及$X_{(1)}+X_{(n)}$的概率分布，并计算方差$\mathrm{Var}(\hat{\theta}_{1/2})$；对比第1问的结果，你有何结论？

<u>解</u>. (a) 由$EX=\theta$得到矩估计$\hat{\theta}=\bar{X}$。根据大数定律，它是θ的（强）相合估计。$E\bar{X}=EX=\theta$，故矩估计是无偏估计，这时

$$\mathrm{MSE}(\hat{\theta})=\mathrm{Var}\bar{X}=\frac{\mathrm{Var}X}{n}=\frac{2^2}{12n}=\frac{1}{3n}.$$

(b) 似然函数

$$L(\theta;x_1,\dots,x_n)=p(x_1,\dots,x_n;\theta)=\left(\frac{1}{2}\right)^n I_{x_{(n)}-1\le\theta\le x_{(1)}+1},$$

因此它在区间$[x_{(n)}-1,x_{(1)}+1]$上处处取得最大值，因此对一切$0\le t\le 1$，

$$\hat{\theta}_t:=t(X_{(n)}-1)+(1-t)(X_{(1)}+1)=tX_{(n)}+(1-t)X_{(1)}+1-2t$$

都是θ的极大似然估计。

(c) 由

$$P(X_{(n)}\le x)=\begin{cases}1, & 若x>\theta+1;\\ \left(\frac{x-\theta+1}{2}\right)^n, & 若\theta-1\le x\le\theta+1;\\ 0, & 若x<\theta-1.\end{cases}$$

得到$X_{(n)}$的概率密度为

$$f_{X_{(n)}}(x) = n\left(\frac{x-\theta+1}{2}\right)^{n-1}\frac{1}{2}I_{\theta-1\le x\le\theta+1}.$$

进而得到

$$\begin{aligned}EX_{(n)} &= \int_{\theta-1}^{\theta+1} x\cdot n\left(\frac{x-\theta+1}{2}\right)^{n-1}\frac{1}{2}dx \\ &= \int_0^1 (2y+\theta-1)ny^{n-1}dy \qquad (y=\frac{x-\theta+1}{2}) \\ &= \frac{2n}{n+1}+(\theta-1) = \theta+\frac{n-1}{n+1}.\end{aligned}$$

类似（或由对称性）可得

$$EX_{(1)} = \theta-\frac{n-1}{n+1}.$$

(d) 于是

$$E\hat{\theta}_t = \theta+\frac{2-4t}{n+1},$$

因此当且仅当$t=1/2$时，$\hat{\theta}_t$是θ的无偏估计。由X_n的概率分布函数知，对任何$\varepsilon>0$，

$$P(|X_{(n)}-(\theta+1)|>\varepsilon) = P(X_{(n)}<(\theta+1)-\varepsilon) \le \left(\frac{2-\varepsilon}{2}\right)^n I_{0<\varepsilon<2} \to 0, \quad n\to\infty,$$

因此

$$X_{(n)} \xrightarrow{P} \theta+1, \quad n\to\infty.$$

类似可证

$$X_{(1)} \xrightarrow{P} \theta-1, \quad n\to\infty.$$

因此

$$\hat{\theta}_t \xrightarrow{P} \theta, \quad n\to\infty,$$

即$\hat{\theta}_t$是θ的相合估计。

(e) 令$Y_k = X_k-\theta$，则$Y_1,\ldots,Y_n \overset{i.i.d.}{\sim} U(-1,1)$，由

$$P(Y_{(1)}\ge u, Y_{(n)}\le v) = P(u\le Y_k\le v, k=1,2,\ldots,n) = \left(\frac{v-u}{2}\right)^n I_{-1\le u\le v\le 1},$$

因此

$$f_{Y_{(1)},Y_{(n)}}(u,v) = -\frac{\partial^2}{\partial u\partial v}\left(\frac{v-u}{2}\right)^n I_{-1\le u\le v\le 1} = n(n-1)\left(\frac{v-u}{2}\right)^{n-2}\frac{1}{4}I_{-1\le u\le v\le 1}.$$

于是

$$
\begin{aligned}
f_{Y_{(1)}+Y_{(n)}}(z) &= \int_{-\infty}^{+\infty} f_{Y_{(1)},Y_{(n)}}(u, z-u)du \\
&= \int_{-\infty}^{+\infty} n(n-1)\left(\frac{z-u-u}{2}\right)^{n-2}\frac{1}{4}I_{-1\le u\le z-u\le 1}du \\
&\qquad\qquad (\max\{-1, z-1\} \le u \le \frac{z}{2}) \\
&= I_{\max\{-1,z-1\}\le\frac{z}{2}}\int_{\max\{-1,z-1\}}^{\frac{z}{2}} n(n-1)\left(\frac{z-2u}{2}\right)^{n-2}\frac{1}{4}du \\
&= I_{|z|\le 2}\int_0^{1-\frac{|z|}{2}} n(n-1)w^{n-2}\frac{1}{4}dw \qquad (w=\frac{z}{2}-u) \\
&= \frac{n}{4}\left(1-\frac{|z|}{2}\right)^{n-1} I_{|z|\le 2}.
\end{aligned}
$$

于是

$$
\begin{aligned}
\mathrm{Var}(Y_{(1)}+Y_{(n)}) = E(Y_{(1)}+Y_{(n)})^2 &= \int_{-2}^{2} z^2\cdot\frac{n}{4}\left(1-\frac{|z|}{2}\right)^{n-1}dz \\
&= 2\int_0^2 z^2\cdot\frac{n}{4}\left(1-\frac{|z|}{2}\right)^{n-1}dz \\
&= 4\int_0^1 (1-w)^2 n w^{n-1}dw \quad (w=1-\frac{z}{2}) \\
&= \frac{8}{(n+1)(n+2)},
\end{aligned}
$$

因此

$$
\mathrm{Var}\hat{\theta}_{1/2} = \mathrm{Var}\left(\frac{X_{(1)}+X_{(n)}}{2}\right) = \frac{1}{4}\mathrm{Var}\left(Y_{(1)}+Y_{(n)}\right) = \frac{2}{(n+1)(n+2)} \le \frac{1}{3n},
$$

即$\hat{\theta}_{1/2}$比$\bar{X}$有效。 □

题8 设总体分布为$U[\theta,2\theta]$，其中$\theta>0$是未知参数。设$X_1,\ldots,X_n$是来自该总体的简单随机样本。

1. 利用矩估计方法求θ的无偏估计量$\hat{\theta}_1$，计算其方差；
2. 求θ的极大似然估计量$\hat{\theta}_{\mathrm{MLE}}$，并由它构造$\theta$的一个无偏估计$\hat{\theta}_2$，并计算$\hat{\theta}_2$的方差；
3. 把$X_{(1)}$当作θ的一个点估计，由它构造θ的一个无偏估计$\hat{\theta}_3$，并计算$\hat{\theta}_3$的方差；
4. 试比较上述无偏估计的有效性。

解. (a) 由$EX=\frac{3}{2}\theta$得到矩估计$\hat{\theta}_1=\frac{2}{3}\bar{X}$。它是$\theta$的无偏估计，

$$
\mathrm{Var}(\hat{\theta}_1) = \frac{4}{9}\mathrm{Var}\bar{X} = \frac{4}{9n}\mathrm{Var}X = \frac{\theta^2}{27n}.
$$

(b) 似然函数

$$
L(\theta;x_1,\ldots,x_n) = p(x_1,\ldots,x_n;\theta) = \left(\frac{1}{\theta}\right)^n I_{\theta\le x_{(1)}\le x_{(n)}\le 2\theta} = \frac{1}{\theta^n}I_{\frac{1}{2}x_{(n)}\le\theta\le x_{(1)}},
$$

因此它在$\frac{1}{2}x_{(n)}$处取得最大值，因此θ的极大似然估计是

$$\hat{\theta}_{\text{MLE}} = \frac{1}{2}X_{(n)}.$$

由

$$P(X_{(n)} \le t) = \begin{cases} 1, & t \ge 2\theta; \\ \left(\frac{t-\theta}{\theta}\right)^n, & \theta \le t < 2\theta; \\ 0, & t < \theta \end{cases}$$

得到$X_{(n)}$的概率密度函数为

$$f_{X_{(n)}}(t) = n\left(\frac{t-\theta}{\theta}\right)^{n-1}\frac{1}{\theta}I_{\theta<t<2\theta}.$$

所以，

$$\begin{aligned}
EX_{(n)} &= \int_\theta^{2\theta} t\cdot n\left(\frac{t-\theta}{\theta}\right)^{n-1}\frac{1}{\theta}dt = \int_0^1 (1+u)\theta\cdot nu^{n-1}du = \frac{2n+1}{n+1}\theta, \\
EX_{(n)}^2 &= \int_\theta^{2\theta} t^2\cdot n\left(\frac{t-\theta}{\theta}\right)^{n-1}\frac{1}{\theta}dt = \int_0^1 (1+u)^2\theta^2\cdot nu^{n-1}du = \frac{4n^2+8n+2}{(n+2)(n+1)}\theta^2, \\
\text{Var}X_{(n)} &= \frac{n}{(n+1)^2(n+2)}\theta^2,
\end{aligned}$$

于是

$$\hat{\theta}_2 = \frac{n+1}{2n+1}X_{(n)}$$

是θ的无偏估计，它的方差为

$$\text{Var}\hat{\theta}_2 = \frac{(n+1)^2}{(2n+1)^2}\text{Var}X_{(n)} = \frac{n}{(2n+1)^2(n+2)}\theta^2.$$

(c) 由

$$P(X_{(1)} \le t) = 1 - P(X_{(1)} > t) = \begin{cases} 1, & \text{若}t \ge 2\theta; \\ 1-\left(\frac{2\theta-t}{\theta}\right)^n, & \text{若}\theta \le t \le 2\theta; \\ 0, & \text{若}t < \theta. \end{cases}$$

得到$X_{(1)}$的概率密度函数为

$$f_{X_{(1)}}(t) = n\left(\frac{2\theta-t}{\theta}\right)^{n-1}\frac{1}{\theta}I_{\theta<t<2\theta}.$$

所以，

$$\begin{aligned}
EX_{(1)} &= \int_\theta^{2\theta} t\cdot n\left(\frac{2\theta-t}{\theta}\right)^{n-1}\frac{1}{\theta}dt = \int_0^1 (2-u)\theta\cdot nu^{n-1}du = \frac{n+2}{n+1}\theta, \\
EX_{(1)}^2 &= \int_\theta^{2\theta} t^2\cdot n\left(\frac{2\theta-t}{\theta}\right)^{n-1}\frac{1}{\theta}dt = \int_0^1 (2-u)^2\theta^2\cdot nu^{n-1}du = \frac{n^2+5n+8}{(n+2)(n+1)}\theta^2, \\
\text{Var}X_{(1)} &= \frac{n}{(n+1)^2(n+2)}\theta^2,
\end{aligned}$$

于是

$$\hat{\theta}_3 = \frac{n+1}{n+2} X_{(1)}$$

是θ的无偏估计，它的方差为

$$\mathrm{Var}\hat{\theta}_3 = \frac{(n+1)^2}{(n+2)^2} \mathrm{Var} X_{(1)} = \frac{n}{(n+2)^3}\theta^2.$$

(d) 由于$2n+1 \geq n+2$，所以$\hat{\theta}_2$总比$\hat{\theta}_3$有效。

当$n \geq 4$时，

$$(2n+1)^2(n+2) - 27n^2 = 4n^3 - 15n^2 + 9n + 2 \geq n^2(4n-15) + 9n + 2 \geq 0,$$

当$n = 3$时

$$7^2 \times 5 - 27 \times 3^2 = 245 - 243 > 0,$$

当$n = 2$时

$$5^2 \times 4 - 27 \times 2^2 = -8 < 0,$$

当$n = 1$时

$$3^2 \times 3 - 27 = 0,$$

所以当$n \geq 3$时，$\hat{\theta}_2$比$\hat{\theta}_1$有效，当$n = 2$时，$\hat{\theta}_1$比$\hat{\theta}_2$有效。 □

题5 甲乙两位编辑独立地对同一段文字进行校对，甲发现了n_1处错误，乙发现了n_2处错误，并且其中有n_3处错误是甲乙共同发现的。试用矩估计法和极大似然估计法估计这段文字的错误个数。

<u>解</u>. 记X, Y, Z分别表示甲发现的错误个数、乙发现的错误个数、甲乙共同发现的错误个数。设这段文字共有N处错误，一个错误被甲发现的概率为p_1，被乙发现的概率为p_2。则由于甲乙是独立校对的，所以一个错误被甲乙同时发现的概率为p_1p_2。于是$X \sim B(N, p_1)$，$Y \sim B(N, p_2)$，$Z \sim B(N, p_1p_2)$。

$$EX = Np_1, \quad EY = Np_2, \quad EZ = Np_1p_2,$$

由矩估计方法

$$n_1 = \hat{N}\hat{p}_1, \quad n_2 = \hat{N}\hat{p}_2, \quad n_3 = \hat{N}\hat{p}_1\hat{p}_2,$$

因此

$$n_3 = \hat{N} \cdot \frac{n_1}{\hat{N}} \cdot \frac{n_2}{\hat{N}} = \frac{n_1 n_2}{\hat{N}},$$

于是得到N的矩估计量

$$\hat{N} = \frac{n_1 n_2}{n_3}.$$

通常取最接近$\frac{n_1 n_2}{n_3}$的整数值为$\hat{N}$的值。

用极大似然估计方法，似然函数为

$$\begin{aligned}
&L(N, p_1, p_2) = P_{N,p_1,p_2}(X = n_1, Y = n_2, Z = n_3) \\
=&\frac{N!}{n_3!(n_1-n_3)!(n_2-n_3)!(N-n_1-n_2+n_3)!} \\
&\qquad \cdot (p_1p_2)^{n_3}[p_1(1-p_2)]^{n_1-n_3}[p_2(1-p_1)]^{n_2-n_3}[(1-p_1)(1-p_2)]^{N-n_1-n_2+n_3} \\
=&\frac{N!}{n_3!(n_1-n_3)!(n_2-n_3)!(N-n_1-n_2+n_3)!} p_1^{n_1} p_2^{n_2} (1-p_2)^{N-n_2}(1-p_1)^{N-n_1}.
\end{aligned}$$

由

$$\frac{\partial \ln L(N,p_1,p_2)}{\partial p_1} = \frac{n_1}{p_1} - \frac{N-n_1}{1-p_1},$$
$$\frac{\partial \ln L(N,p_1,p_2)}{\partial p_2} = \frac{n_2}{p_2} - \frac{N-n_2}{1-p_2},$$

解得对任意$N \geq 1$，$L(N,p_1,p_2)$在

$$\hat{p_1} = \frac{n_1}{N}, \hat{p_2} = \frac{n_2}{N}$$

处的值

$$L^*(N) := L(N,\hat{p_1},\hat{p_2}) \geq L(N,p_1,p_2).$$

而

$$\begin{aligned}\frac{L^*(N+1)}{L^*(N)} &= \frac{N+1}{N+1-n_1-n_2+n_3}(1-\hat{p_1})(1-\hat{p_2}) \\ &= \frac{N+1}{N+1-n_1-n_2+n_3}\frac{N-n_1}{N}\frac{N-n_2}{N} \\ &= 1 + \frac{-n_3N^2+(n_1n_2-n_1-n_2)N+n_1n_2}{N^3-(n_1+n_2-n_3-1)N^2},\end{aligned}$$

取

$$N^* = \frac{(n_1n_2-n_1-n_2)+\sqrt{(n_1n_2-n_1-n_2)^2+4n_1n_2n_3}}{2n_3},$$

于是，当$N > N^*$时，$L^*(N+1) < L^*(N)$；当$N < N^*$时，$L^*(N+1) > L^*(N)$。因此N的极大似然估计$\hat{N}$为不小于

$$\frac{(n_1n_2-n_1-n_2)+\sqrt{(n_1n_2-n_1-n_2)^2+4n_1n_2n_3}}{2n_3}$$

的最小整数。不难发现，通常这个估计值小于矩估计值。

如果$n_1 = 24$，$n_2 = 25$，$n_3 = 20$，则N的矩估计值为30，极大似然估计值为29。 □