概率论与数理统计 (Fall 2024)
习题课讲义

2024 年 11 月 13 日

聂嵘琢, 祝国隽

B01GB006Y-02/Note9

---

1. 随机抛掷一枚质地均匀的散子 1200 次. 试利用中心极限定理来估计出现 6 点的次数介于 180 和 240 之间的概率 (给出有关 $\Phi(\cdot)$ 的表达式即可, 其中 $\Phi(\cdot)$ 为标准正态分布的分布函数).

Ans: 记 $X_i = 1$ 若第 i 次抛出 6 点, 反之 $X_i = 0$.

则 $S = \sum_{i=1}^{1200} X_i$. 可知 $S \sim B(1200, 1/6)$.

从而$ES/1200 = 1/6, \mathrm{var}(S/1200) = \frac{1}{1200}\frac{1}{6}\left(1-\frac{1}{6}\right) = \frac{1}{240\times 36}$. 故

$$\frac{S/1200 - 1/6}{\sqrt{1/(240\times 36)}} \xrightarrow{d} N(0,1).$$

$$P(180 \le S \le 240) = P\left(\frac{180/1200 - 1/6}{\sqrt{1/(240\times 36)}} \le \frac{S/1200 - 1/6}{\sqrt{1/(240\times 36)}} \le \frac{240/1200 - 1/6}{\sqrt{1/(240\times 36)}}\right)$$
$$= \Phi(4\sqrt{15}/5) - \Phi(-2\sqrt{15}/n).$$

2. 令 $X_1, X_2, \cdots$ 为独立同分布, 服从标准正态分布的随机变量序列. 求证:

$$\frac{nX_{n+1}}{\sum_{i=1}^n X_i^2} \xrightarrow{d} N(0,1).$$

Ans: 记 $Y = \frac{nX_{n+1}}{\sum_{i=1}^n X_i^2} = \frac{X_{n+1}}{\sum_{i=1}^n X_i^2/n}$. 则

$$E\left(\sum_{i=1}^{n} X_i^2/n\right) = EX_1^2 = 1$$

$$\operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}\left(X_i^2\right) = \frac{2}{n} \to 0, n \to \infty.$$

由 Chebyshev 不等式知

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i^2 - 1\right| > \varepsilon\right) \leq \frac{\operatorname{var}\left(\sum_{i=1}^{n} X_i/n\right)}{\varepsilon^2} \to 0, n \to \infty$$

即 $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} 1$. 另一方面 $X_{n+1} \sim N(0,1)$, 由 Slustky 定理知 $Y \xrightarrow{d} N(0,1)$.

3. 利用 Poisson 分布和中心极限定理证明:

$$\lim_{n\to\infty}\sum_{i=0}^{n}\frac{1}{i!}n^i \mathrm{e}^{-n} = \frac{1}{2}.$$

Ans: 记 $X_1, X_2, \cdots, X_n \cdots$ iid $\sim$ Poisson(1), 则 $EX_n = 1, \operatorname{var}(X_n) =$ Poisson 分布可加性, 得 $Y_n = \sum_{i=1}^{n} X_i \sim$ Poisson(n), 且

$$P\left(Y_n \leqslant n\right) = \sum_{i=0}^{n} P\left(Y_n = i\right) = \sum_{i=0}^{n}\frac{n^i}{i!}\mathrm{e}^{-n}.$$

由中心极限定理, 当 $n \to \infty$ 时, 有

$$\frac{Y_n - n}{\sqrt{n}} \xrightarrow{d} N(0,1),$$

从而

$$\sum_{i=0}^{n}\frac{n^i}{i!}\mathrm{e}^{-n} = P\left(Y_n \leqslant n\right) = P\left(\frac{Y_n - n}{\sqrt{n}} \leqslant \frac{n-n}{\sqrt{n}}\right) \to \Phi(0) = \frac{1}{2}.$$

4. 记 $\alpha \in (0,1)$, 同时删去样本中 $\lfloor n\alpha/2 \rfloor$ 个数值小的和 $\lfloor n\alpha/2 \rfloor$ 个数值大的观测值, 对

剩余的观测值求均值, 记作 $\bar{x}_\alpha$, 即

$$\bar{x}_\alpha = \frac{x_{(\lfloor n\alpha/2+1\rfloor)} + x_{(\lfloor n\alpha/2+2\rfloor)} + \cdots + x_{(n-\lfloor n\alpha/2\rfloor)}}{n-2\lfloor n\alpha/2\rfloor}.$$

称 $\bar{x}_\alpha$ 为 α-截尾均值. 现有 12 名用户对一电视节目打分, 分值为

$$83, 88, 91, 86, 53, 92, 100, 93, 85, 84, 82, 79.$$

求样本均值和 0.2 -截尾均值.

Ans: 样本均值为

$$\bar{X} = \frac{1}{12}(83+88+91+86+53+92+100+93+85+84+82+79) \approx 84.667$$

由于 $\left\lfloor \frac{n\alpha}{2} \right\rfloor = \left\lfloor \frac{12*0.2}{2} \right\rfloor = \lfloor 1.2 \rfloor = 1.$

12 个样本去掉最小值 53 和最大值 100 , 再取平均, 则 0.2-截尾均值为:

$$\bar{X}_{0.2} = \frac{1}{10}(83+88+91+86+92+93+85+84+82+79) = 86.3.$$

5. 考察两组简单样本 $X_1, X_2, \cdots, X_n$ 和 $Y_1, Y_2, \cdots, Y_m$, 记它们的样本均值分别为 $\bar{X}$ 和 $\bar{Y}$, 样本方差分别为 S_1^2 和 S_2^2,

$$\bar{Z} = \frac{1}{n+m}\left(\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j\right)$$

$$S_3^2 = \frac{1}{n+m-1}\left[\sum_{i=1}^n \left(X_i - \bar{Z}\right)^2 + \sum_{j=1}^m \left(Y_j - \bar{Z}\right)\right]$$

求证: $\bar{Z} = \dfrac{n\bar{X}+m\bar{Y}}{n+m}, S_3^2 = \dfrac{(n-1)S_1^2+(m-1)S_2^2}{n+m-1} + \dfrac{nm(\bar{X}-\bar{Y})^2}{(n+m)(n+m+1)}.$

Ans: 由 $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \bar{Y} = \frac{1}{m}\sum_{j=1}^m Y_j \Rightarrow \sum_{i=1}^n X_i = n\bar{X}, \sum_{j=1}^m Y_j = m\bar{Y}$ 则

$$\begin{aligned}\bar{Z} &= \frac{1}{n+m}\left(\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j\right) \\ &= \frac{1}{n+m}(n\bar{X}+m\bar{Y})\end{aligned}$$

由 $S_1^2=\dfrac{1}{n-1}\sum_{i=1}^n\left(X_i-\bar{X}\right)^2, S_2^2=\dfrac{1}{m-1}\sum_{j=1}^m\left(Y_j-\bar{Y}\right)^2$, 则

$$
\begin{aligned}
S_3^2 &= \frac{1}{n+m-1}\left[\sum_{i=1}^n\left(X_i-\bar{Z}\right)^2+\sum_{j=1}^m\left(Y_j-\bar{Z}\right)^2\right] \\
&= \frac{1}{n+m-1}\left[\sum_{i=1}^n\left(X_i-\bar{X}+\bar{X}-\bar{Z}\right)^2+\sum_{j=1}^m\left(Y_j-\bar{Y}+\bar{Y}-\bar{Z}\right)^2\right] \\
&= \frac{1}{n+m-1}\left[\sum_{i=1}^n\left(X_i-\bar{X}\right)^2+2\sum_{i=1}^n\left(X_i-\bar{X}\right)(\bar{X}-\bar{Z})+n(\bar{X}-\bar{Z})^2\right]+ \\
&\frac{1}{n+m-1}\left[\sum_{j=1}^m\left(Y_j-\bar{Y}\right)^2+2\sum_{j=1}^m\left(Y_j-\bar{Y}\right)(\bar{Y}-\bar{Z})+m(\bar{Y}-\bar{Z})^2\right] \\
&= \frac{1}{n+m-1}\left[\sum_{i=1}^n\left(X_i-\bar{X}\right)^2+\sum_{j=1}^m\left(Y_j-\bar{Y}\right)^2\right] \\
&+\frac{1}{n+m-1}\left[n\left(\bar{X}-\frac{n\bar{X}+m\bar{Y}}{n+m}\right)^2+m\left(\bar{Y}-\frac{n\bar{X}+m\bar{Y}}{n+m}\right)^2\right], \\
&= \frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-1}+\frac{1}{n+m-1}\left[\frac{nm^2}{(n+m)^2}(\bar{X}-\bar{Y})^2+\frac{mn^2}{(n+m)^2}(\bar{Y}-\bar{X})^2\right] \\
&= \frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-1}+\frac{nm(\bar{X}-\bar{Y})^2}{(n+m-1)(n+m)}.
\end{aligned}
$$

6. 记 $X_1,X_2,\cdots,X_n,X_{n+1},X_{n+2},\cdots,X_{2n}$ iid $\sim N(0,1)$,

$$
Y=\left(X_1+X_2+\cdots+X_n\right)^2+\left(X_{n+1}+X_{n+2}+\cdots+X_{2n}\right)^2.
$$

求常数 c 使得 cY 服从 χ^2 分布.

Ans: 令

$$
Z_1=X_1+X_2+\cdots+X_n\sim N(0,n),
$$
$$
Z_2=X_{n+1}+X_{n+2}+\cdots+X_{2n}\sim N(0,n).
$$

Z_1,Z_2 相互独立.

$$
\frac{1}{\sqrt{n}}Z_1\sim N(0,1),
$$

$$\frac{1}{\sqrt{n}} Z_2 \sim N(0,1).$$

因此, 由卡方分布的性质,

即$c = 1/n$. $$\frac{1}{n}\left(Z_1^2 + Z_2^2\right) = \frac{1}{3} Y \sim \chi_2^2.$$

这次补充材料参考张伟平老师 (USTC) 的讲义

2 次序统计量

次序统计量: 即若$X_1, X_2, \cdots, X_n$ i.i.d. $\sim F$, 将其按大小排列为$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$, 则称$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$为次序统计量, 它的任一部分, 如$X_{(i)}$,和$(X_{(i)}, X_{(j)})$ $(1 \leq i < j \leq n)$等也称为次序统计量.

次序统计量的分布

1. $X_{(n)}$ 的分布

$$P(X_{(n)} \leq x) = P(X_1 \leq x, \cdots, X_n \leq x) = F^n(x)$$

2. $X_{(1)}$ 的分布

$$P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) = 1 - P(X_1 > x, \cdots, X_n > x) = 1 - (1 - F(x))^n$$

3. $X_{(m)}$ 的分布$(1 < m < n)$

$$\begin{aligned} f_{X_{(m)}}(x)dx &\approx P(x < X_{(m)} \leq x + dx \\ &= \frac{n!}{(m-1)!1!(n-m)!} F^{m-1}(x) f(x) dx [1 - F(x + dx)]^{n-m} \end{aligned}$$

因此两边同时除以dx, 并令$dx \to 0$,得到

$$f_{X_{(m)}}(x) = m \binom{n}{m} F^{m-1}(x) f(x) [1 - F(x)]^{n-m}$$

.

4. $X_{(i)}, X_{(j)}$ 的联合密度

$$f_{ij}(x, y) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (F(x))^{i-1} (F(y) - F(x))^{j-i-1} & \\ \qquad \times (1 - F(y))^{n-j} f(x) f(y), & \text{当} x < y, \\ 0, & \text{其它}. \end{cases}$$

5. $(X_{(1)}, \cdots, X_{(n)})$的联合密度为

$$g(x_{(1)}, \cdots, x_{(n)}) = \begin{cases} n! f(x_{(1)}) \cdots f(x_{(n)}), & \text{当} x_{(1)} < x_{(2)} < \cdots < x_{(n)}, \\ 0, & \text{其它}. \end{cases}$$

6. 极差$X_{(n)} - X_{(1)}$ 的分布

作下列变换

$$\begin{cases} V = X_{(j)} - X_{(i)} \\ Z = X_{(i)} \end{cases} \Longleftrightarrow \begin{cases} X_{(i)} = Z \\ X_{(j)} = V + Z \end{cases}$$

变换的Jacobian 行列式为:$|J| = \left| \frac{\partial(X_{(i)}, X_{(j)})}{\partial V \partial Z} \right| = 1$, $(X_{(i)}, X_{(j)})$的联合分布密度由前给出,故(V, Z)的联合密度为

$$g_{ij}(v, z) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (F(z))^{i-1} (F(v+z) - F(z))^{j-i-1} & \\ \qquad \times (1 - F(v+z))^{n-j} f(z) f(v+z), & \text{当} v > 0, \\ 0, & \text{其它}. \end{cases} \tag{2.1}$$

从而易知V的密度为

$$f_V(v) = \int_{-\infty}^{\infty} g_{i,j}(v,z)dz.$$

特别, 当取$i=1,\ j=n$得到$(R, X_{(1)})$的联合密度

$$g_{1,n}(v,z) = \begin{cases} \frac{n!}{(n-2)!}(F(v+z)-F(z))^{n-2}f(v+z)f(z), & \text{当}v>0, \\ 0, & \text{当}v \le 0. \end{cases} \tag{2.2}$$

而R的边缘密度为$\int_{-\infty}^{\infty} g_{1,n}(v,z)dz$.

均匀分布情形

设$X_1, X_2, \cdots, X_n$ i.i.d. ~均匀分布$U(0,1)$,其分布函数和密度函数分别为

$$F(x) = \begin{cases} 0, & x \le 0, \\ x, & 0<x<1, \\ 1, & x \ge 1. \end{cases} \quad \text{和} \quad f(x) = \begin{cases} 1, & 0<x<1, \\ 0, & \text{其它} \end{cases}$$

设$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$为样本$X_1, X_2, \cdots, X_n$的次序统计量, 次序统计量$X_{(m)}$ 的密度函数为

$$f_m(x) = m\binom{n}{m}x^{m-1}(1-x)^{m-n}I_{[0,1]}(x). \tag{2.3}$$

由前可知$(X_{(i)}, X_{(j)})$的联合密度为

$$f_{ij}(x) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!}x^{i-1}(y-x)^{j-i-1}(1-x)^{n-j}, & 0<x<y<1, \\ 0, & \text{其它}. \end{cases}$$

而$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$的联合密度为

$$f_{1,2,\cdots,n}(x_{(1)}, x_{(2)}, \cdots, x_{(n)}) = \begin{cases} n!, & \text{当}0<x_{(1)}<\cdots<x_{(n)}<1, \\ 0, & \text{其它}. \end{cases}$$

令$F(z)=z,\ 0<z<1,\ F(v+z)=v+z,\ 0<v+z<1$,得到在均匀分布$U(0,1)$ 场合(V,Z)的联合密度

$$g_{ij}(v,z) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!}z^{i-1}v^{j-i-1}[1-(v+z)]^{n-j}, & \\ \qquad\qquad \text{当 } 0<z<1,\ 0<v+z<1, & \\ 0, & \text{其它}. \end{cases}$$

此时$V=X_{(j)}-X_{(i)}$的边缘密度, 通过计算积分$\int_0^{1-v} g_{i,j}(v,z)dz$得

$$g_{nij}(v) = \begin{cases} \frac{n!}{(j-i-1)!(n-j+i)!}v^{j-i-1}(1-v)^{n-j+1}, & \text{当 } 0<v<1, \\ 0, & \text{其它}. \end{cases}$$

特别极差 $R=X_{(n)}-X_{(1)}$的密度函数$g_{1n}(r)$为将前式中的v换成r,将j和i分别用n和1代替得到

$$g_{1n}(r) = \begin{cases} n(n-1)r^{n-2}(1-r), & \text{当 } 0<r<1, \\ 0, & \text{其它}. \end{cases}$$

3 统计量的极限分布

本章引言中已指出, 在许多情形下统计量的精确分布很难求出, 因此我们要研究统计量的极限分布. 首先给出下列定义.

定义 1. *当样本大小趋向无穷时, 统计量的分布趋于一确定分布, 则后者的分布称为统计量的极限分布. 也常称为大样本分布.*

当样本大小n 充分大时, 极限分布可作为统计量的近似分布.

研究统计量的极限分布有下列意义: (1) 为了获得统计推断方法的优良性, 常常要知道统计量的分布. 但统计量的精确分布一般很难求得, 建立统计量的极限分布, 提供了一种近似方法, 总比什么方法没有要好. (2) 有时统计量的精确分布虽可求出, 但表达式过于复杂, 使用不方便. 若极限分布较简单, 宁可使用极限分布. (3) 有些统计推断方法的优良性本身就是研究其极限性质, 如相合性, 渐近正态性等.

定义 2. *t当样本大小$n \to \infty$ 时, 一个统计量或统计推断方法的性质称为大样本性质 (Large sample properties). 当样本大小固定时, 统计量或统计推断方法的性质称为小样本性质 (Small sample properties).*

在此要强调的是, 大样本性质 和小样本性质 的差别不在于样本个数的多少, 而是在于所讨论的问题是在样本大小$n \to \infty$ 时去考虑, 还是在样本大小 n 固定时去研究, 关于大样本性质 的研究构成了数理统计的一个很重要的部分, 叫做统计大样本理论. 统计大样本理论, 近几十年来发展很快, 成为二次世界大战后数理统计发展的重要特点之一. 有些统计分支, 如非参数统计, 其中大样本理论占据了主导地位.