第m个次序统计量 $X_{(m)}$ 的分布函数

$$F_m(x) = P(X_{(m)} \le x) = \sum_{i=m}^{n} \binom{n}{i} (F(x))^i (1-F(x))^{n-i} \quad ①$$

$$= m\binom{n}{m}\int_0^{F(x)} t^{m-1}(1-t)^{n-m}\,dt \quad ②$$

derivation: ① $P(X_i \le x) = F(x)$ (i个样本 $\le x$) (n个样本, i个 $\le x$) $P = F(x)$

二项分布: 一个 x_i 小于x的概率 $F(x)=P$、 i个小于x的概率

$$\therefore \sum_{i=m}^{n} \binom{n}{i}(F(x))^i(1-F(x))^{n-i}$$

②: ② $\sum_{i=m}^{n} \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{(n-m)!(m-1)!}\int_0^{p} t^{m-1}(1-t)^{n-m}\,dt$ $(p = F(x))$

第m个次序统计量 $X_{(m)}$ 的密度函数:

$$f_m(x) = F'_m(x) = m\binom{n}{m}(F(x))^{m-1}(1-F(x))^{n-m} f(x)$$

P(最小值小于x) → (1-最小值大于x)

样本极值 $X_{(1)}$ 的分布函数: $F_1(x) = P(X_{(1)}<x) = 1-(1-F(x))^n$

密度 $f_1(x) = n(1-F(x))^{n-1}f(x)$

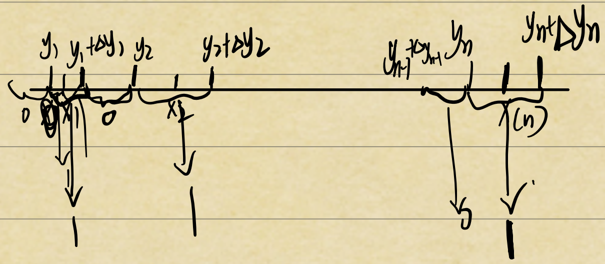
样本极值 $X_{(n)}$ 的分布,密度函数: $F_n(x) = P(X_{(n)}<x) = (F(x))^n$

密度: $n(F(x))^{n-1}f(x)$

n个次序统计量的联合密度:

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

$$g(y_1,\cdots,y_n) = \lim_{\Delta y_i \to 0} \frac{P(y_1<X_{(1)}<y_1+\Delta y_1,\cdots,y_n<X_{(n)}<y_n+\Delta y_n)}{\Delta y_1 \Delta y_2 \cdots \Delta y_n}$$



Beta(r, n-r+1)

均匀分布次序统计量的分布 ~ Beta(·, ·)

$F(x) = x$

$f_{Y_m}(x) = m\binom{n}{m} x^{m-1}(1-x)^{n-m} I_{(0,1)}(x) = \frac{n!}{(r-1)!(n-r)!} w^{r-1}(1-w)^{n-r}$

$E(W_r) = \frac{r}{n+1} \quad r = 1, \dots, n$

(原函数来自均匀分布) $F(Y_r) \sim Beta(r, n-r+1)$ (均匀分布的次序统计量分布函数)

Y_r ↓ order statistic

$E(F(Y_r)) = \frac{r}{n+1} \Rightarrow F(Y_r) \approx \frac{r}{n+1} \Rightarrow Y_r \approx F^{-1}\left(\frac{r}{n+1}\right)$

$\Rightarrow Y_r \approx F^{-1}\left(\frac{r}{n+1}\right)$

例 $X \sim Exp(\lambda) \quad F(x) = 1 - e^{-\lambda x} \quad Y = 1 - e^{-\lambda X} \sim U(0,1)$

$\tilde{m} \begin{cases} Y_{(n+1)/2} & n \text{ is odd} \\ \frac{Y_{\frac{n}{2}} + Y_{\frac{n}{2}+1}}{2} & n \text{ is even} \end{cases}$

$Y_{1\sim5}$ is order statistics

Then Y_3 is estimator of the median

$Pr(Y_1 < m < Y_5) =$ 　　S: $Pr(X_i < m) = \frac{1}{2}$

最小值 < m 　　F: $1 - \frac{1}{2} = \frac{1}{2}$

最小值小于 m → 1 − 最小值大于 m ⇒ 5个值都比m大

5次实验 　$Y_1 < m \Rightarrow$ 最小值小于 m

⇕

至少1次成功 (至少一个序素小于m)

⇓

$Pr(Y_1 < m < Y_5) \Rightarrow \underline{1 \le W \le 4}$ 　　$W \sim Bino(5, \frac{1}{2})$

$m < Y_5 \Rightarrow$ 最大值大于 m ⇒ 至少一个序素大于 m ⇒ 至少一次失败

CI by orderstatistics → contruct CI for quantile 最多4次成功

$P_r(y_1 < m < y_5) \approx 0.94$

$$P(y_i < m < y_j) = P(i-1 < w < j) = \sum_{k=i}^{j-1} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} \approx 1-\alpha$$

1. construct CI for quantiles by order statistics

① order sample $x_1 \cdots x_n \Rightarrow y_1 < y_2 < \cdots < y_n$ (y_i is ordered data)

② desity function of y_r $f_{y_r}(y) = \frac{n!}{(r-1)!(n-r)!} F(y)^{r-1} [1 - F(y)]^{n-r} f(y)$

$f(y)$, $F(y)$ comes from X_i (CDF, pdf of original data)

for original data: $U(0,1)$ $F(y) = y$ 、 $f(y) = 1 \Rightarrow f_{y_r}(y) \sim Beta(r, n-r+1)$ mean: $\frac{r}{n+1}$

tow estimate quantiles $X_i \sim F \Rightarrow F(X_i) \sim U_i \sim (0,1)$

$y_1 < y_2 < \cdots < y_n \Rightarrow F(y_1) < F(y_2) < \cdots F(y_n)$

$F(y_r) \sim Beta(r, n-r+1)$

$E[F(y_r)] = \frac{r}{n+1}$ $F(y_r) \approx \frac{r}{n+1} \Rightarrow y_r \approx F^{-1}\left(\frac{r}{n+1}\right)$

example: $0.4 = \frac{r}{n+1} \Rightarrow r = 0.4(n+1)$

$n=10$: $r = 0.4 \times 11 = 4.4$ 4.4 approximated quantile is between y_4

and $y_5 \Rightarrow F^{-1}(0.4) = w y_4 + (1-w) y_5$

$P_r(y_i < m < y_j)$ define success: $(x_i < m)$ value of X below media.

$y_i < m$ at least i th smallest element is smaller than median.

⇓

at least i successes

$w \sim Binomil(n, \frac{1}{2})$

$P_r(y_i < m < y_j) = P_r(i \le w \le j-1) = \sum_{k=i}^{j-1} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} \binom{n}{k} \approx 0.95$ (此区间可通过给i,j控制

$y_j > m$: jth smallest element is above the median.

$\Downarrow$

at most $j-1$ success

只要保证超过21...即可

次序统计量 只能用于估计 quantiles 不能估计 mean（如果对称 可以认为 mean = median）

y_1, y_2, y_3, y_4, y_5 small sample $\Rightarrow$ can't use asymptotic normality

① order statistic

② construct CI for quantile (eg. median).

$\tilde{y}_3$ is point estimate

Interval should be around $\tilde{y}_3$

$$P(\tilde{y}_1 < m < \tilde{y}_5) = P(1 \le w < 5) = \sum_{k=1}^{4} \binom{5}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{5-k} = 1 - 0.5^5 - 0.5^5 = \frac{15}{16}$$

$= P(1 \le w \le 4)$

$P(\tilde{y}_1 < m < \tilde{y}_4)$ 并不接近 0.95 $\therefore$ 最后利用 $(\tilde{y}_1, \tilde{y}_5)$ 构建区间）约 95

$\frac{k}{n+1} = \frac{1}{2}$ $\quad \frac{k}{6} = \frac{1}{2}$ $\quad k = 3$

So (y_1, y_5) is a 94% CI for m.

$$P(\tilde{y}_i < m < \tilde{y}_j) = P(i-1 < w < j) = \sum_{k=i}^{j-1} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} \approx 1-\alpha \ (b)$$

pick i, j suitable

length of fish: we can order it: $\tilde{y}_1 \cdots \tilde{y}_9$. point estimate is y_5.

find good estimate for median

we should pick

$$P(\tilde{y}_2 < m < \tilde{y}_8) = \sum_{k=2}^{7} \binom{9}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{9-k} \approx 96\%$$

one value greater than $\tilde{y}_5$

one value smaller than $\tilde{y}_5$

比较后发现 这区间最好

pbinom（累计概率） $\longrightarrow P(7) - P(1)$ $\quad (7 \le 10) - (1 \le 10)$ $\quad (7-2)$ R command.

$[y_2, y_7]$ 是 96%的CI. for median.

$$1-\alpha = P(Y_i < \pi_p < Y_j) = P(i-1 < W < j) = \sum_{k=i}^{j-1} \binom{n}{k} p^k (1-p)^{n-k}$$

$P(X_i < \pi_p) = p$

estimate $\pi_{0.25}$ using CI: $n=27$ choose y_k $\quad \frac{k}{n+1} = 0.25 = \frac{k}{28} \Rightarrow k=7.$

$y_7 = 183.$

$\frac{k}{n+1} = 0.5$

$k = 0.5 \times 28 = 14 \quad y_{14} = 200$

$Pr(y_i < \pi_{0.25} < y_j) = \quad [P(X_i < \pi_{0.25}) = 0.25]$

$\sum_{k=i}^{j-1} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{27-k} \binom{27}{k}$

$i = 4 \quad j = 10 \quad 0.815$

$Pr[i \le B \le j]$ asypototic normality.

sample size is large enough $\Rightarrow$ construct CI using normal.

$B \sim Binom(27, \frac{1}{4})$

$$Pr\left(\frac{i-1-\frac{27}{4}}{\sqrt{27 \times \frac{1}{4} \times \frac{3}{4}}} < \frac{B-\frac{27}{4}}{\sqrt{27 \times \frac{1}{4} \times \frac{3}{4}}} \le \frac{j-1-\frac{27}{4}}{\sqrt{27 \times \frac{1}{4} \times \frac{3}{4}}}\right) = \Phi\left(\frac{j-27/4-1}{\sqrt{27 \times \frac{1}{4} \times \frac{3}{4}}}\right) - \Phi\left(\frac{i-\frac{27}{4}}{\sqrt{27 \times \frac{1}{4} \times \frac{3}{4}}}\right)$$

$Z \sim N(0,1)$

pbinom

其实没必要估,直接用 ~~pbinom~~ 算就好