Journal of Mathematical Research and Exposition
Vol. 11, No.2, May. 1991

The Accurate Distribution of the Kolmogorov Statistic With Application to Bootstrap Approximation*

Zhu Lixing

(Institute of Systems Seience Academia Sinica, Beijing)

Suppose that $X_1, \cdots, X_m$ come from the population F, F_m be the empirical measure determined by $X_1, \cdots, X_m$. Let

$$D_m = \sup_{-\infty<t<\infty} |F_m(t) - F(t)| \tag{1}$$

It is well-known that when F is continuous the limit distribution of $\sqrt{m}\,D_m$ is as follows:

$$\lim_{m\to\infty} P\{\sqrt{m}\,D_m \geq \lambda\} = 2\sum_{j=1}^{\infty}(-1)^{j-1}\exp\{-2j^2\lambda^2\} \tag{2}$$

And the accurate distribution of D_m was obtained by Zhang Liqian (1956). But when F is discontinuous, the results concerning that of D_m is very rarely in literature.

In this paper, we obtain the following results.

Put $p_r(\beta) = (r+\beta)^r/r!$ $\quad u_l(\beta) = \beta\, p_l(\beta)/(l+\beta)$

$$W_k^*(\beta) = \sum_{l=0}^{[k-\beta]} u_l(\beta)\, p_{k-l}(-\beta)$$

$$W_k(\beta) = p_k(-\beta) + \sum_{l=1}^{[k-\beta]} \frac{(\beta^2 - l)\, p_l(\beta)\, p_{k-l}(-\beta)}{(l+\beta)^2}$$

where $W_k^*(\beta)$, $W_k(\beta)$ are well-defined for $k, l \geq \beta > 0$.

Theorem 1 Suppose F is the probability measure which hase mass $\frac{1}{n}$ on each point of $\{y_1, \cdots, y_n\}$. $\frac{n}{m} = p$ is integer. Then for any $\lambda \in (0,1)$, there exists an integer. k satisfying $\frac{k-1}{n} < \lambda \leq \frac{k}{n}$ such that

$$P\{D_m \geq \lambda\} = P\{D_m \geq \frac{k}{n}\} = \frac{2 \cdot m!}{m^m} \sum_{i=1}^{[(n+k)/2k]} \sum_1 W_{m_1}(\frac{2k}{p}), \cdots, W_{m_{i-1}}(\frac{2k}{p}) W_{m_i}^*(\frac{k}{p})$$

$$+ \frac{m!}{m^m} \sum_{i=1,}^{[n/2k]} \sum_1 W_{m_1}(\frac{2k}{p}), \cdots, W_{m_{i-1}}(\frac{2k}{p}) \{W_{m_i}^*(\frac{2k}{p}) + \sum_{l_1+l_2=m_i} W_{l_1}(\frac{k}{p}) W_{l_2}^*(\frac{k}{p})\}$$

and
$$P\{D_m \geq \lambda\} = \begin{cases} 0 & \lambda \geq 1, \\ 1 & \lambda = 0. \end{cases}$$

* Received Oct. 17, 1989.

where $\sum\limits_{1}$ sum all of summands such that $m_1 + \cdots + m_i = m$, $W_k(\beta)$ and $W_k^*(p)$ are equal to zero when $k < \beta$, $[x]$ is denoted as the integer part of x.

Theorem 2 Suppose F is the considered probability measure as in theorem 1. Then for given positive constant ε, when $0 < \varepsilon \leq z = O(m^{1/6})$, we have

$$P\{\sqrt{m}\, D_m \geq z\} = P\{\sqrt{m}\, D_m \geq \sqrt{m}\frac{k}{n}\}$$
$$= 2\sum_{l=1}^{\infty}(-1)^{l+1} e^{-2l^2(\sqrt{m}\frac{k}{n})^2}\{1 - \frac{2l^2 k}{3n} - \frac{1}{18\,m}[f_1 - 4(f_1+3)l^2(\sqrt{m}\frac{k}{n})^2$$
$$+ 8l^2(\sqrt{m}\frac{k}{n})^2] + \frac{l^2 k}{27\,mn}[\frac{f_2}{5} - (4(f_2+45)l^2(\sqrt{m}\frac{k}{n})^2)/15 + 8l^4(\sqrt{m}\frac{k}{n})^4]$$
$$+ O((\sqrt{m}\,k)^{13}/m^2 n^{13})\}. \qquad (4)$$

where $\sqrt{m}(k-1)/n < z \leq \sqrt{m}\,k/n$, $f_1 = l^2 - \frac{1}{2}(1-(-1)^l)$, $f_2 = 5l^2 + 22 - \frac{15}{2}(1+(-1)^l)$.

Applying the above results to the bootstrap approximation for the distribution of Kolmogorov statistic, we can derive a very satisfactory result.

Suppose $Y_1, \cdots, Y_n$ be iid sample drawn from a continuous population F. F_n is empirical distribution based on this sample. A resample $Y_1^*, \cdots, Y_m^*$ comes from F_n. F_m^* is empirical distribution concerning $Y_1^*, \cdots, Y_m^*$.

Define $\quad D_n = \sup\limits_{-\infty < t < \infty} |F_n(t) - F(t)|, \qquad D_{m,n}^* = \sup\limits_{-\infty < t < \infty} |F_m^*(t) - F_n(t)|$.

$P\{\sqrt{m}\,D_{m,n}^* \geq \lambda \,|\, Y_1, \cdots, Y_n\}$ is denoted as the conditional distribution of $\sqrt{m}\,D_{m,n}^*$ given $Y_1, \cdots, Y_n$.

Theorem 3 For some $0 < \delta < 1$ such that $[n^\delta] < m \leq n$, $n/m = p$ is integer. when $\frac{1}{2} \leq \lambda = O(n^{\delta/q})$, then

$$|P\{\sqrt{n}\,D_n \geq \lambda\} - P\{\sqrt{m}\,D_{m,n}^* \geq \lambda \,|\, Y_1, \cdots, Y_n\}|$$
$$\leq \frac{A(\lambda)}{3\sqrt{n}} + \frac{4}{\sqrt{np}} + o(\frac{1}{\sqrt{np}}) \leq \frac{2\sqrt{2}(\sqrt{p}-1)}{3e\sqrt{n}} + \frac{4}{\sqrt{np}} + o(\frac{1}{\sqrt{np}}) \quad \text{a.s.} \ (5)$$

where $A(\lambda) = |2\sum_{l=1}^{\infty}(-1)^{l+1} 2l^2 \lambda e^{-2l^2\lambda}|$.

Kolmogorov统计量的精确分布及其在Bootstrap逼近中的应用

朱 力 行

（中国科学院系统科学研究所）

摘　　要

假设 x_1, x_m iid服从分布 F，当 F 连续时，Kolmogorov 统计量的精确分布已由张里千 (1956) 获得．本文考虑在 n 个点上取概率为 $\frac{1}{n}$ 的离散分布．得到的精确分布与张里千的结果有相同的形式．利用这个结果，得到 Kolmogorov 统计量分布的 Bootstrap 逼近的收敛速度为 $n^{-\frac{1}{2}}$ 的阶．这是对统计量的极限分布形式复杂甚至未知的情况下 Bootstrap 逼近的收敛速度问题的一个初步的探讨．