Note on the KOLMOGOROV Statistic for a General Discontinuous Variable

By DONALD J. WHEELER [1])

Summary: A simple demonstration of the conservative character of the KOLMOGOROV test in the case of general discontinuous distributions is given.

Zusammenfassung: Ein einfacher Beweis des konservativen Charakters des KOLMOGOROFFschen Tests bei allgemeinen unstetigen Verteilungen wird gebracht.

Let Z take on values $\{x_\alpha\}$ with probabilities $\{p_\alpha\}$ for α in some countable indexing set A, and let the distribution function of Z, $F(z)$, be continuous everywhere else on $-\infty < z < \infty$. Denote the empirical distribution function for a sample of size N from $F(z)$ by $F_N(z)$. The KOLMOGOROV statistic, $D_N = \sup_z |F_N(z) - F(z)|$, is not distribution free, but D_N is invariant under the probability integral transformation. Denote the distribution of $U = F(z)$ by $G(u)$ and let $G_N(u)$ correspond to $F_N(z)$. Also let $Pr[Z < x_\alpha] = q_{2\alpha-1}$ and $Pr[Z \leq x_\alpha] = q_{2\alpha}$. The variable U is restricted to values in the set $R = \bigcup_{\alpha \in A} [q_{2\alpha}, q_{2\alpha+1}] \subset [0,1] = I$. Let $F(z_0) = u_0 \in [q_{2\alpha}, q_{2\alpha-1}]$ for some α, then

$$Pr[q_{2\alpha} < U \leq u_0] = Pr[x_\alpha < Z \leq z_0] = Pr[Z \leq z_0] - Pr[Z \leq x_\alpha] = u_0 - q_{2\alpha}$$

and

$$Pr[U = q_{2\alpha}] = Pr[Z = x_\alpha] = p_\alpha = q_{2\alpha} - q_{2\alpha-1}$$

while

$$Pr[q_{2\alpha-1} < U < q_{2\alpha}] = Pr[Z < x_\alpha] - Pr[Z < x_\alpha] = 0$$

therefore $G(u)$ increases according to a uniform distribution on the intervals in R and is constant of the set $I - R$, with discontinuities occuring at the points $u = q_{2\alpha}$ for every $\alpha \in A$.

Now let $H(z)$ be continuous on $-\infty < z < \infty$ such that $H(x_\alpha) = F(x_\alpha)$ for every $\alpha \in A$. The KOLMOGOROV statistic is $D_N^* = \sup_z |H_N(z) - H(z)|$. Letting $U = H(z)$ we have $U \sim$ Uniform $[0,1]$ and $D_N^* = \sup_u |G_N^*(u) - G^*(u)|$ where $G^*(u)$ is the uniform distribution function. Now if we use a random sample

---

[1]) DONALD J. WHEELER, Southern Methodist University, University Park, Dallas, Texas 75222.

$u_1^*,\ldots,u_N^*$ from $G^*(u)$ to generate a random sample $u_1,\ldots,u_N$ from $G(u)$ by setting $u_k^* = u_k$ if $u_k^* \in R$ and $u_k^* = \min[q_{2\alpha}|q_{2\alpha} \geq u_k^*]$ if $u_k^* \in I - R$, we can see that on R, $G_N(u) = G_N^*(u)$, and from above $G(u) = G^*(u)$ on R. Thus

$$|G_N^*(u) - G^*(u)| = |G_N(u) - G(u)| \quad \text{on} \quad R$$

which implies

$$D_N^* = \sup_{u \in I} |G_N^*(u) - G^*(u)| \geq \sup_{u \in R} |G_N(u) - G(u)| = D_N .$$

Therefore the cumulative distribution of D_N is not less than the cumulative distribution of D_N^* for every value of $D > 0$. Hence tests of hypotheses and confidence regions based on D_N^* are conservative when the population is discontinuous.

Reference

WALSH, JOHN E.: Bounded Probability Properties of KOLMOGOROV-SMIRNOV and Similar Statistics for Discrete Data, Annals of the Institute of Statistical Mathematics, XV, 153, 1963.

Eingegangen im Februar 1969