M.S. Tikhov (Lobachevsky State University of Nizhny Novgorod, Russia).
Estimating distributions from samples of random size.

Let $\boldsymbol{X}^{(n)} = (X_1, X_2, ..., X_n)$ be independent and identically distributed random variables with an unknown distribution function (df) $F(x)$. The distribution of the Kolmogorov statistics (see [1], p.134-141) is based on the empirical df $F_n(x) = \frac{1}{n}\sum_{j=1}^{n} I(X_j < x)$, where $I(X_j < x)$ is an indicator of the event $(X_j < x)$, over the sample has *a fixed and a priori known size n*. The report considers the problem of estimating the df $F(x)$ and the asymptotic behavior of $F_n(x)$ using a complete (or censored) observations of ***a random sizes*** ν_p. The random variable ν_p has $\lambda-$negative binomial (**NB**) distribution that is the probability mass function of the $\lambda-$**NBD** is $\mathbf{P}(\nu_p = 0) = (1-\lambda q)^{\mu}$,

$$\mathbf{P}(\nu_p = k) = \frac{a\,(a+\lambda)\cdot ...\cdot (a+\lambda\,(k-a))}{\Gamma(k-a+2)}\,(1-\lambda q)^{\mu} q^{k-a},\ \ k \in \{a, a+1, ...\},$$

$$q = 1-p,\ 0 < p, \lambda < 1,\ \mu = a/\lambda,\ \nu_p \text{ and } X^{(n)} \text{ are independent.}$$

Let the segment [0,1] be the support of this distribution, $u_i = i/n,\ i = 0, 1, ..., n,$ be the division points on[0,1] and let $W_i = I(X_i < u_i)$. It is also required to construct estimates for the distribution function $F(x)$ and its quantiles $F^{-1}(\alpha),\ 0 < \alpha < 1$,based on the sample $\boldsymbol{W}^{(n)} = \{(u_i, W_i), i = 0, 1, ..., n\}$. If the sample size n is *not random* then the kernel-type Nadaraya-Watson estimator $\tilde{F}_n(x)$ with the kernel $A = A(x),\ \operatorname{supp}(A) = [-1, 1]$, and $||A||^2 = \int_{-1}^{1} A^2(x)\,dx < \infty$, was used as estimator of the distribution function $F(x)$ in [2], [3]. There the asymptotic distribution, at $n \to \infty$, of the normalized difference $n^{2/5}(\tilde{F}_n(x) - \mathbf{E}(\tilde{F}_n(x)))/b(x)$, where $b^2(x) = F(x)(1-F(x))\ ||A||^2$, was normal $\xi \in \mathcal{N}(0, 1)$. Let $K_\mu(x)$ be the Macdonald function (modified Bessel function of the second kind). Here are results of just a few of them.

Theorem 1. *If $F(x)$ is a continuous distribution function, then*

$$\text{(i) } F_{\nu_p}(x) \xrightarrow[\lambda q\to 1]{p} F(x), \quad \text{(ii) } \eta_\lambda = \sqrt{1-\lambda q}\cdot\frac{\nu_p\,(F_{\nu_p}(x) - F(x))}{b(x)} \xrightarrow[\lambda q\to 1]{d} \zeta_\mu, \quad \text{(iii) } \eta_\lambda \xrightarrow[\lambda\to 0]{d} \xi,$$

ς_μ has a characteristic function $\varphi_\mu(t) = \left(\dfrac{2}{2+t^2}\right)^{\mu}$, and the random variable $\dfrac{\varsigma_\mu}{\sqrt{2}}$ has a distribution density

$$f_\mu(x) = \frac{1}{\Gamma(\mu)\sqrt{\pi}}\left(\frac{|\,x\,|}{2}\right)^{\mu-1/2} K_{\mu-1/2}(|\,x\,|),\ \ x \in \mathbf{R}\,. \qquad (1)$$

The behavior of the kernel estimators and its quantiles based on the samples with random sizes is also considered.

REFERENCES

[1] Kolmogorov A.N. *On the empirical definition of the distribution law. In*:Select Works of A.N.Kolmogorov, Volume II: Theory of Probability and Mathematical Statistics. Springer, 1992 – 615p.
[2] Tikhov M.S. Statistical Estimation based on Interval Censored Data. – In: *Param. And Semiparam. Models with Appl. To Rel.,Surv. Analysis, and Qual. Of Life*: Springer-Verlag: Theor.&Meth., **XLIV** (2004), pp. 209-215.
[3] Okumura H., Naito K. Non-parametric kernel regression for multinomial data. *Journal Muktivariate Analysis*, **97** (2006), pp. 2009-2022.