KERNEL DISTRIBUTION ESTIMATORS
FROM A STATISTICIAN POINT OF VIEW

Ryszard Zieliński
Institute of Mathematics Polish Acad. Sc., Warszawa, Poland
R.Zielinski@impan.gov.pl

Abstract

Kernel estimators do not converge to the true distribution uniformly.

Key Words: kernel estimators, asymptotics

AMS classification: 62G20, 62G30, 62G07

The version of the Glivenko-Cantelli theorem in the form that we will exploit below states that

$$(GCT) \qquad (\forall\varepsilon)(\forall\eta)(\exists N)(\forall n \geq N)(\forall F \in \mathcal{F}) \quad P\{\sup_{x\in\mathbf{R}^1} |F_n(x) - F(x)| \geq \varepsilon\} \leq \eta$$

where

$$F_n(x) = \frac{1}{n}\sum_{j=1}^{n} 1_{(-\infty,x]}(X_j)$$

and $\mathcal{F}$ is the class of all distribution functions. The theorem is effective in the sense that for every $\varepsilon > 0$ and for every $\eta > 0$ one can effectively calculate $N = N(\varepsilon, \eta)$ such GCT holds. That can be done by, e.g., the inequality (Massart 1990) which states that for every n and $\varepsilon > 0$

$$(*) \qquad P\{\sup_{x\in\mathbf{R}^1} |F_n(x) - F(x)| \geq \varepsilon\} \leq 2e^{-2n\varepsilon^2}.$$

Due to the above, GCT is a genuinely statistical theorem; if all that a statistician knows is that an unknown distribution F belongs to $\mathcal{F}$, he is able to make a precise inference about F (testing hypothesis or constructing confidence intervals).

The standard kernel density estimator is of the form (e.g. Wegman 2006)

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n}k\Big(\frac{x-X_j}{h_n}\Big)$$

with appropriate $h_n, n = 1, 2, \ldots$. We shall consider kernel distribution estimator of the form

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n}K\Big(\frac{x-X_j}{h_n}\Big)$$

where $K(x) = \int_{-\infty}^{x} k(t)dt$, and we shall show that (GCT) does not hold if F_n is replaced by $\widehat{F}_n$, i.e. that the following is true

$$(\exists\varepsilon)(\exists\eta)(\forall N)(\exists n \geq N)(\exists F \in \mathcal{F}) \quad P\{\sup_{x\in\mathbf{R}^1}|\widehat{F}_n(x) - F(x)| \geq \varepsilon\} \geq \eta.$$

Concerning the kernel K we make the following natural assumptions: 1) $0<K(0)<1$; and 2) $K^{-1}(t) < 0$ for some $t \in (0, K(0))$. Concerning the sequence $(h_n, n = 1, 2, \ldots)$ we assume that $h_n > 0, n = 1, 2, \ldots$.

Obviously it is enough to demonstrate that

$$(\exists\varepsilon)(\exists\eta)(\forall n)(\exists F \in \mathcal{F}) \quad P\{\widehat{F}_n(0) > F(0) + \varepsilon\} \geq \eta$$

or that

$$(\exists\varepsilon)(\exists\eta)(\forall n)(\exists F \in \mathcal{F}) \quad P\Big\{\frac{1}{n}\sum_{j=1}^{n}K(\xi_j) > F(0) + \varepsilon\Big\} \geq \eta,$$

where $\xi_j = -X_j/h_n$.

We have

$$\bigcap_{j=1}^{n}\{K(\xi_j) > F(0) + \varepsilon\} \subset \{\frac{1}{n}\sum_{j=1}^{n}K(\xi_j) > F(0) + \varepsilon\}$$

hence it is enough to prove that

$$(\exists\varepsilon)(\exists\eta)(\forall n)(\exists F \in \mathcal{F}) \quad P\Big\{K(\xi_j) > F(0) + \varepsilon\Big\} \geq \eta^{1/n}.$$

Given K, take ε and η such that

$$1 - \eta^{1/n} < t - \varepsilon < t$$

and then choose F such that

$$F(0) = t - \varepsilon$$

and

$$F\Big(K^{-1}(t)\Big) < 1 - \eta^{1/n}. \qquad \square$$

It follows that for kernel estimators no inequality like (*) can be obtained which makes the estimators of a doubtful usefulness for statistical applications.

References

Massart, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Annals of Probability*, 18: 1269–1283

Wegman, E.J. (2006). Kernel estimators. In *Encyclopedia of statistical sciences.* Second Edition, Vol. 6, Wiley–Interscience