.

GLIVENKO-CANTELLI THEOREM AND KERNEL ESTIMATORS

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

Presented to

International Seminar on Stability Problems for Stochastic Models

Nahariya (Israel) Oct 22 - 26, 2007

.

Summary

Standard kernel estimators do not converge to the true distribution uniformly. A consequence is that no inequality like Dvoretzky-Kiefer-Wolfowitz one can be constructed, and as a result it is impossible to answer the question how many observations are needed to guarantee a prescribed level of accuracy of the estimator. A remedy is to adapt the bandwidth to the sample at hand.

.

Dvoretzky-Kiefer-Wolfowitz inequality (Massart 1990)

$$P\{\sup_{x\in\mathbf{R}} |F_n(x) - F(x)| \geq \epsilon\} \leq 2e^{-2n\epsilon^2}$$

Glivenko-Cantelli theorem

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

where

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

Here $N = N(\epsilon, \eta)$ does not depend of $F \in \mathcal{F}$!

Standard kernel density estimator

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

Kernel distribution estimator

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n}K\Big(\frac{x-X_j}{h_n}\Big), \qquad K(x) = \int_{-\infty}^{x}k(t)dt$$

Glivenko-Cantelli theorem does not hold:

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

It is enough to demonstrate that

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

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

Proof that

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

Recall the assumption that $K^{-1}(t)<0$ for some $t\in(0,F(0))$. Take $\epsilon\in(0,t)$ and $\eta\in(t-\epsilon,1)$. Given ϵ, η, and n, take F such that $F(0)=t-\epsilon$ and $F\big(-h_nK^{-1}(t)\big)>\eta^{1/n}$. Then

$$P\{X_j<-h_nK^{-1}(t)\}\quad \textit{and}\quad P\{K\left(-\frac{X_j}{h_n}\right)>t\}>\eta^{1/n}$$

Due to the fact that

$$\bigcap_{j=1}^{n}\left\{K\left(-\frac{X_j}{h_n}\right)>t\right\}\subset\left\{\frac{1}{n}\sum_{j=1}^{n}K\left(-\frac{X_j}{h_n}\right)>t\right\}$$

we have

$$P\left\{\frac{1}{n}\sum_{j=1}^{n}K\left(-\frac{X_j}{h_n}\right)>t\right\}=P\left\{\underbrace{\frac{1}{n}\sum_{j=1}^{n}K\left(-\frac{X_j}{h_n}\right)}_{\widehat{F}_n(0)}>F(0)+\epsilon\right\}>\eta$$

QED

RANDOM BANDWIDTH

$X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ - order statistics

Define

$$H_n = \min\{X_{j:n} - X_{j-1:n},\ j = 2, 3, \ldots, n\}$$

Define the kernel estimator

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

where for K we assume:

$$K(t) = \begin{cases} 0, & \text{for } t \leq -\frac{1}{2} \\ \frac{1}{2}, & \text{for } t = 0 \\ 1, & \text{for } t \geq \frac{1}{2} \end{cases}$$

$K(t)$ continuous and increasing in $(-\frac{1}{2}, \frac{1}{2})$

For a fixed k and $j = 1, 2, \ldots, n$ we have

$$K\Big(\frac{X_{k:n} - X_{j:n}}{H_n}\Big) =$$

$$= \begin{cases} 0, & \text{for } \frac{X_{k:n} - X_{j:n}}{H_n} \leq -\frac{1}{2} \Leftrightarrow X_{j:n} > X_{k:n} + \frac{1}{2} H_n \Leftrightarrow j > k \\ \frac{1}{2}, & \text{for } t = 0 \\ 1, & \text{for } j < k \end{cases}$$

It follows that

$$\widetilde{F}_n(X_{k:n}) = \frac{1}{n} \sum_{j=1}^{n} K\Big(\frac{X_{k:n} - X_{j:n}}{H_n}\Big) = \frac{k-1}{n} + \frac{1}{2n}$$

$$= F_n(X_{k-1:n}) + \frac{1}{2n} = F_n(X_{k:n}) - \frac{1}{2n}$$

Hence, for $k = 1, 2, \ldots, n$, we have $|\widetilde{F}_n(X_{k:n}) - F_n(X_{k:n})| \leq \dfrac{1}{2n}$.

For $k = 1, 2, \ldots, n$, we have $|\widetilde{F}_n(X_{k:n}) - F_n(X_{k:n})| \leq \frac{1}{2n}$.
Kernel estimator $\widetilde{F}_n(x)$ is continuous and increasing, empirical distribution function $F_n(x)$ is a step function, and in consequence $|\widetilde{F}_n(x) - F_n(x)| \leq \frac{1}{2n}$ for all $x \in (-\infty, \infty)$.
By the triangle inequality

$$|\widetilde{F}_n(x) - F(x)| \leq |F_n(x) - F(x)| + \frac{1}{2n}$$

we obtain

$$P\{\sup_{x \in \mathbf{R}} |\widetilde{F}_n(x) - F(x)| \geq \epsilon\} \leq P\{\sup_{x \in \mathbf{R}} |F_n(x) - F(x)| + \frac{1}{2n} \geq \epsilon\}$$

and Dvoretzky-Kiefer-Wolfowitz inequality takes on the form:

$$P\{\sup_{x \in \mathbf{R}} |\widetilde{F}_n(x) - F(x)| \geq \epsilon\} \leq 2e^{-2n(\epsilon - 1/2n)^2}, \quad n > \frac{1}{2\epsilon}$$

which enables us to calculate $N = N(\epsilon, \eta)$ that guarantees the prescribed accuracy of the kernel estimator $\widetilde{F}_n(x)$.

.

COMMENT.

The smallest $N = N(\epsilon, \eta)$ that guarantees the prescribed accuracy is somewhat greater for kernel estimator $\widetilde{F}_n$ than that for crude empirical step function F_n.

For example, $N(0.1, 0.1) = 150$ for F_n and $= 160$ for $\widetilde{F}_n$;

$N(0.01, 0.01) = 26,492$ for F_n and $= 26,592$ for $\widetilde{F}_n$.

.

COMMENT

Another disadvantage of kernel smoothing has been discovered by Hjort and Walker (2001):

"*kernel density estimator with optimal bandwidth lies outside any confidence interval, around the empirical distribution function, with probability tending to 1 as the sample size increases*".

.

Perhaps a reason is that smoothing adds to observations
something which is rather arbitrarily chosen
and which may spoil the inference.

.

A GENERALIZATION.

Inequality

$$P\{\sup_{x\in\mathbf{R}} |\widetilde{F}_n(x) - F(x)| \geq \epsilon\} \leq 2e^{-2n(\epsilon-1/2n)^2}, \quad n > \frac{1}{2\epsilon}$$

holds for every smoothed nondecreasing distribution function that satisfies $|\widetilde{F}_n(X_{k:n}) - F_n(X_{k:n})| \leq \dfrac{1}{2n}$, $k = 1, 2, \ldots, n$.

REFERENCES

Hjort, N.L., and Walker, S.G. (2001). A note on kernel density estimators with optimal bandwidths. Statistics & Probability Letters 54, 153-159

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