On Testing the Hypothesis of Equality of Two Bernoulli Regression Functions

Elizbar Nadaraya, Petre Babilua

Ivane Javakhishvili Tbilisi State University, Department of Mathematics
elizbar.nadaraya@tsu.ge; petre.babilua@tsu.ge

Let randon variables $Y^{(i)}$, $i=1,2$, take two values 1 and 0 with probabilities p_i (succes) and $1-p_i$, $i=1,2$ (failure), respectively. Assume that the probability of success p_i is the function of an independent variable $x\in[0,1]$, i.e. $p_i = p_i(x)=\mathbb{P}\{Y^{(i)}=1\,|\,x\}$ $(i=1,2)$ (see [1]-[3]). Let t_j, $j=1,\dots,n$, be the devision points of the interval $[0,1]$: $t_j=\frac{2j-1}{2n}$, $j=1,\dots,n$

Let further $Y_i^{(1)}$ and $Y_i^{(2)}$, $i=1,\dots,n$, be mutually independent random Bernoulli variables with $\mathbb{P}\{Y_i^{(k)}=1\,|\,t_i\}=p_k(t_i)$, $\mathbb{P}\{Y_i^{(k)}=0\,|\,t_i\}=1-p_k(t_i)$, $i=1,\dots,n$, $k=1,2$. Using the samples $Y_1^{(1)},\dots,Y_n^{(1)}$ and $Y_1^{(2)},\dots,Y_n^{(2)}$ we want to chek the hypothesis

$$H_0: p_1(x)=p_2(x)=p(x),\quad x\in[0,1],$$

against the sequence of "close" alternatives of the form

$$H_{1n}: p_k(x)=p(x)+\alpha_n u_k(x)+o(\alpha_n),\quad k=1,2.$$

where $\alpha_n\to 0$ relevantly, $u_1(x)\neq u_2(x)$, $x\in[0,1]$ and $o(\alpha_n)$ uniformly in $x\in[0,1]$.

We consider the crietrion of testing the hypothesis H_0 based on the statistic function

$$T_n=\frac{1}{2}\ nb_n\int_{\Omega_n(\tau)}\left[\hat{p}_{1n}(x)-\hat{p}_{2n}(x)\right]^2 p_n^2(x)\ dx=$$

$$=\frac{1}{2}\ nb_n\int_{\Omega_n(\tau)}\left[p_{1n}(x)-p_{2n}(x)\right]^2 dx,\qquad \Omega_n(\tau)=\left[\tau b_n,(1-\tau)b_n\right],\quad \tau>0,$$

Where $\hat{p}_{in}=p_{in}(x)p_n^{-1}(x)$, $p_{in}(x)=\frac{1}{nb_n}\sum_{j=1}^n K\left(\frac{x-t_j}{b_n}\right)Y_j^{(i)}$, $i=1,2,$ $p_n(x)=\frac{1}{nb_n}\sum_{i=1}^n K\left(\frac{x-t_i}{b_n}\right)$,

$K(x)$ is some distribution density and $b_n\to 0$ is a sequence of positive numbers, $\hat{p}_{in}(x)$ is the kernel estimator of the regression function.

We assume that a kernel $K(x)\geq 0$ is chosen so that it is a function of bounded variation and satisfies the conditions: $K(x)=K(-x)$, $K(x)=0$ for $|x|\geq\tau>0$, $\int K(x)dx=1$. The class of such functions is denoted by $H(\tau)$.

Theorem *Let* $K(x)\in H(\tau)$ *and* $p(x),u_1(x),u_2(x)\in C^1[0,1]$. *If* $nb_n^2\to\infty$, $\alpha_n b_n^{-1/2}\to 0$ *and* $nb_n^{1/2}\alpha_n^2\to c_0$, $0<c_0<\infty$, *then for the hypothesis* H_{1n}

$$b_n^{-1/2}\left(T_n-\Delta(p)\right)\sigma^{-1}(p)\xrightarrow{d}N(a,1),$$

where $\Delta(p)$ *and* $\sigma^2(p)$ *are defined in Lemma* 2 *and* $\xrightarrow{d}$ *denotes convergence in distribution and* $N(a,1)$ *is a random variable having the standard normal distribution with parameters* $(a,1)$,

$$a=\frac{c_0}{2\sigma(p)}\int_0^1\left(u_1(x)-u_2(x)\right)^2 dx.$$