On a new principle to construct distribution free goodness of fit tests

Estate Khmaladze
Victoria University of Wellington

Let $p_1, \dots, p_m$ be a discrete probability distribution; all $p_i > 0$ and $\sum_{i=1}^m p_i = 1$. Denote $\nu_{1n}, \dots, \nu_{mn}$ the corresponding frequencies in a sample of size n and consider the vector Y_n of components of chi-square statistic

$$Y_{in} = \frac{\nu_{in} - np_i}{\sqrt{np_i}},\ i = 1, \dots, m.$$

As $n \to \infty$ the vector Y_n has the limit distribution of 0-mean Gaussian vector $Y = (Y_1, \dots, Y_m)^T$ such that

$$Y = X - \langle X, \sqrt{p} \rangle \sqrt{p},$$

where $X = (X_1, \dots, X_m)^T$ is a vector of independent $\mathcal{N}(0,1)$ random variables, $\sqrt{p} = (\sqrt{p}_1, \dots, \sqrt{p}_m)^T$, and $\langle a, b \rangle$ denotes inner product of vectors a and b. The distribution of Y depends on $\sqrt{p}$ – it is only its sum of squares $\langle Y, Y \rangle = \sum_{i=1}^m Y_i^2$, which is chi-square distributed and hence has distribution free from $\sqrt{p}$. It is for this reason that we do not have any other asymptotically distribution free goodness of fit test for the discrete distributions but the chi-square statistic. In this presentation we introduce a vector $Z_n = \{Z_{in}\}_{i=1}^m$ as follows: let r be the unit length "diagonal" vector with all coordinates $1/\sqrt{m}$, and put

$$Z_n = Y_n - \langle Y_n, r \rangle \frac{1}{1 + \langle \sqrt{p}, r \rangle} (r + \sqrt{p}).$$

Then (i) *any* statistic based on Z_n is asymptotically distribution free; (ii) asymptotically, the partial sums $\sum_{i=1}^k Z_{in}$, $k \leq m$, will behave as a discrete time analog of the standard Brownian bridge; (iii) the transformation from Y_n to Z_n is one-to-one. We will explain the nature of transformation of Y_n to Z_n, give different form of such transformations and then show that it produces new results also for continuous distributions in $\mathbb{R}^d$ and for both simple and parametric hypothesis. Therefore, a unified approach to distribution free goodness of fit testing for continuous and discrete distirbutions in $\mathbb{R}^d$ is now possible. This talk is based on Khmaladze, E.V., Note on distribution free testing for discrete distributions, Ann. Stat. 2013, 2979-2993.