Kolmogorov-Smirnov two-samples statistics in a chi-squared approximation

Kolmogorov test[1] measures the maximum absolute deviation between the empirical distribution function F_n of a sample and the empirical distribution function of a theoretical function F:

$$D_n(x) = sup_x |F(x) - F_n(x)|$$

Smirnov generalised Kolmogorov statistics to the comparison of two samples (Kolmogorov-Smirnov test)[2]. Goodman statistics approximates Kolmogorov-Smirnov statistics to a chi-squared statistics[3]. The aim of this test is to compare two distributions, f and g with respectively n and m observations, in order to test whether the two independent samples involved have been drawn from the same population.
Deriving from Kolmogorov test statistics, Goodman test statistics computation involves the two empirical distribution functions F and G, translating it into a chi-squared statistics χ^2_G. If the two samples have been drawn from the same population distribution, the cumulative distributions of both samples may be expected to be close to each other; if the two sample cumulative distributions differ, this suggests that the samples come from different populations. If D_{mn} represents Kolmogorov-Smirnov test statistics, Goodman test statistics χ^2_G is defined as follows:

$$\chi^2_G = 4D^2 \frac{nm}{n+m}$$

Goodman test statistics χ^2_G is approximatively distributed as a χ^2 variable with $\nu = 2$ degrees of freedom; its limit distribution is:

$$P(\chi^2 > \chi^2_{obs}) = \frac{1}{2\Gamma(\frac{\nu}{2})} \int_{\chi^2_{obs}}^{\infty} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \mathrm{dx}$$

[1]A.Kolmogorov, *On the empirical determination of a distribution function*, Giornale dell'Istituto Italiano degli Attuari, 1933.

[2]N.V.Smirnov, *On the estimation of the discrepancy between empirical curves of distributions for two independent samples*, Bull. Math. Univ. Moscou, 1939.

[3]L.A.Goodman, *Kolmogorov-Smirnov tests for psychological research*, Psychol. Bull., 1954.

Deriving from Kolmogorov statistics, Goodman test is based on the same assumptions: it can be applied only in case of unbinned data.