On the asymptotic properties of a nonparametric L_1-test statistic of homogeneity

Gérard Biau *   László Györfi †

June 28, 2005

Abstract

We present two simple and explicit procedures for testing homogeneity of two independent multivariate samples of size n. The nonparametric tests are based on the statistic T_n, which is the L_1 distance between the two empirical distributions restricted to a finite partition. Both tests reject the null hypothesis of homogeneity if T_n becomes large, i.e., if T_n exceeds a threshold. We first discuss Chernoff-type large deviation properties of T_n. This results in a distribution-free strong consistent test of homogeneity. Then the asymptotic null distribution of the test statistic is obtained, leading to an asymptotically α-level test procedure.

AMS Classification: 62G10.

Key words and phrases: homogeneity testing, partitions, large deviations, consistent testing, central limit theorem, Poissonization.

*Corresponding author. Institut de Mathématiques et de Modélisation de Montpellier, Equipe de Probabilités et Statistique, Université Montpellier II, UMR CNRS 5149, CC 51 – Place Eugène Bataillon, 34095 Montpellier Cedex 5, France. e-mail: biau@math.univ-montp2.fr .

†Department of Computer Science and Information Theory, Budapest University of Technology and Economics, H-1521 Stoczek u. 2, Budapest, Hungary.
email: gyorfi@szit.bme.hu . His research has been supported by the Hungarian Academy of Sciences (MTA SZTAKI and MTA IEKCS).

1 Introduction

Consider two mutually independent samples of $\mathbb{R}^d$-valued random vectors $X_1, \ldots, X_n$ and $X'_1, \ldots, X'_n$ with i.i.d. components defined on the same probability space and distributed according to unknown probability measures μ and μ'. We are interested in testing the null hypothesis that the two samples are homogeneous, that is

$$\mathcal{H}_0 : \mu = \mu'.$$

Such tests have been extensively studied in the statistical literature for special parametrized models, e.g. for linear or loglinear models. For example, the analysis of variance provides standard tests of homogeneity when μ and μ' belong to a normal family on the Borel line. For multinomial models these tests are discussed in common statistical textbooks, together with the related problem of testing independence in contingency tables. For testing homogeneity in more general parametric models, we refer the reader to the monograph of Greenwood and Nikulin [12] and further references therein.

However, in many real life applications, the parametrized models are either unknown or too complicated for obtaining asymptotically α-level homogeneity tests by the classical methods. As explained in Pardo, Pardo and Vajda [14], this is typically the case in electroencephalographic (EEG) and electrocardiographic (ECG) biosignal analysis, or in speech source characterization. In such situations parametric families cannot be adopted with confidence, nonparametric tests should be used. For $d = 1$, there are nonparametric procedures for testing homogeneity, for example the Cramer-von Mises, Kolmogorov-Smirnov, or Wilcoxon tests. The problem of $d > 1$ is much more complicated, but nonparametric tests based on finite partitions of $\mathbb{R}^d$ may provide a welcome alternative. In this context, Pardo, Pardo and Vajda [14] recently presented a partition-based generalized likelihood ratio test of homogeneity and derived its asymptotic distribution under the null hypothesis, enabling to control the asymptotic test size. The results of these authors extend former results of Read and Cressie [17], and Pardo, Pardo and Zografos [15] on disparity statistics.

In the present paper, we discuss a simple approach based on a L_1 distance test statistic. The advantage of our test procedure is that, besides being explicit and relatively easy to carry out, it requires very few assumptions on the partition sequence, and it is consistent. Let us now describe our test statistic.

Denote by μ_n and μ'_n the empirical measures associated with the samples $X_1, \ldots, X_n$ and $X'_1, \ldots, X'_n$, respectively, so that

$$\mu_n(A) = \frac{\#\{i : X_i \in A, i = 1, \ldots, n\}}{n}$$

for any Borel subset A, and, similarly,

$$\mu'_n(A) = \frac{\#\{i : X'_i \in A, i = 1, \ldots, n\}}{n}.$$

Based on a finite partition $\mathcal{P}_n = \{A_{n1}, \ldots, A_{nm_n}\}$ of $\mathbb{R}^d$ $(m_n \in \mathbb{N}^*)$, we let the test statistic comparing μ_n and μ'_n be defined as

$$T_n = \sum_{j=1}^{m_n} |\mu_n(A_{nj}) - \mu'_n(A_{nj})|.$$

The paper is organized as follows. We first discuss in Section 2 Chernoff-type large deviation properties of T_n. This results in a distribution-free strong consistent test of homogeneity, which rejects the null hypothesis if T_n becomes large, i.e., T_n is larger than a threshold. In Section 3, we derive the asymptotic null distribution of T_n. This yields another – in fact, a smaller – threshold resulting in a consistent asymptotically α-level test procedure.

2 Large deviation properties

For testing a simple hypothesis versus a composite alternative, Györfi and van der Meulen [13] introduced a related goodness of fit test statistic L_n defined as

$$L_n = \sum_{j=1}^{m_n} |\mu_n(A_{nj}) - \mu(A_{nj})|.$$

The asymptotic normality of this statistic in case μ has a density was discussed in Beirlant, Györfi and Lugosi [4]. Moreover, Beirlant, Devroye, Györfi and Vajda [3], and Devroye and Györfi [8] proved that if

$$\lim_{n\to\infty} m_n = \infty, \qquad \lim_{n\to\infty} \frac{m_n}{n} = 0, \tag{1}$$

and

$$\lim_{n\to\infty} \max_{j=1,\ldots,m_n} \mu(A_{nj}) = 0, \tag{2}$$

then, for all $0 < \varepsilon < 2$,

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbf{P}\{L_n > \varepsilon\} = -g_L(\varepsilon),$$

where

$$g_L(\varepsilon) = \inf_{0<p<1-\varepsilon/2} D(p \,\|\, p + \varepsilon/2),$$

and

$$D(\alpha \,\|\, \beta) = \alpha \ln \frac{\alpha}{\beta} + (1-\alpha) \ln \frac{1-\alpha}{1-\beta}.$$

It means that

$$\mathbf{P}\{L_n > \varepsilon\} = e^{-n(g_L(\varepsilon)+o(1))} \quad \text{as } n \to \infty.$$

The following theorem extends the results of Beirlant, Devroye, Györfi and Vajda [3], and Devroye and Györfi [8] to the statistic T_n.

Theorem 1 *Assume that conditions (1) and (2) are satisfied. Then, under $\mathcal{H}_0$, for all $0 < \varepsilon < 2$,*

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbf{P}\{T_n > \varepsilon\} = -g_T(\varepsilon),$$

where

$$g_T(\varepsilon) = (1+\varepsilon/2)\ln(1+\varepsilon/2) + (1-\varepsilon/2)\ln(1-\varepsilon/2).$$

Observe that for small ε,

$$g_T(\varepsilon) \sim \varepsilon^2/4, \tag{3}$$

and that

$$\lim_{\varepsilon \uparrow 2} g_T(\varepsilon) = 2\ln 2.$$

According to Beirlant, Devroye, Györfi and Vajda [3], for small ε,

$$g_L(\varepsilon) \sim \varepsilon^2/2.$$

Moreover, in contrast to $g_T(\varepsilon)$, the rate function $g_L(\varepsilon)$ is unbounded as $\varepsilon \uparrow 2$, so T_n and L_n have different large deviation properties.

Proof. Introduce the generating function of the sequence $(T_n)_{n\geq 1}$:

$$\lambda_T(s) = \lim_{n\to\infty} \frac{1}{n} \ln \mathbf{E}\{e^{snT_n}\}, \quad s > 0.$$

By Scheffé's [18] theorem for partitions

$$T_n = \sum_{A\in\mathcal{P}_n} |\mu_n(A) - \mu'_n(A)| = 2 \max_{A\in\sigma(\mathcal{P}_n)} \left(\mu_n(A) - \mu'_n(A)\right),$$

where the class of sets $\sigma(\mathcal{P}_n)$ contains all sets obtained by unions of cells of $\mathcal{P}_n$. Therefore

$$\begin{aligned}
\mathbf{E}\{e^{snT_n}\} &= \mathbf{E}\{\max_{A\in\sigma(\mathcal{P}_n)} e^{2sn(\mu_n(A)-\mu'_n(A))}\} \\
&\leq \sum_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\{e^{2sn(\mu_n(A)-\mu'_n(A))}\} \\
&\leq 2^{m_n} \max_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\{e^{2sn(\mu_n(A)-\mu'_n(A))}\} \qquad (4) \\
&= 2^{m_n} \max_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\{e^{2sn\mu_n(A)}\}\mathbf{E}\{e^{-2sn\mu'_n(A)}\}.
\end{aligned}$$

Clearly,

$$\begin{aligned}
\mathbf{E}\{e^{2sn\mu_n(A)}\} &= \sum_{k=0}^{n} e^{2sk} \binom{n}{k} \mu(A)^k \left(1-\mu(A)\right)^{n-k} \\
&= \left(e^{2s}\mu(A) + 1 - \mu(A)\right)^n,
\end{aligned}$$

and, similarly, under $\mathcal{H}_0$,

$$\begin{aligned}
\mathbf{E}\{e^{-2sn\mu'_n(A)}\} &= \sum_{k=0}^{n} e^{-2sk} \binom{n}{k} \mu(A)^k \left(1-\mu(A)\right)^{n-k} \\
&= \left(e^{-2s}\mu(A) + 1 - \mu(A)\right)^n.
\end{aligned}$$

The remainder of the proof is under the null hypothesis $\mathcal{H}_0$. From above, we deduce that

$$
\begin{aligned}
&\mathbf{E}\{e^{snT_n}\} \\
&\le 2^{m_n} \max_{A\in\sigma(\mathcal{P}_n)} \left(e^{2s}\mu(A)+1-\mu(A)\right)^n \left(e^{-2s}\mu(A)+1-\mu(A)\right)^n \\
&= 2^{m_n} \max_{A\in\sigma(\mathcal{P}_n)} \left[\left(e^{2s}\mu(A)+1-\mu(A)\right) \left(e^{-2s}\mu(A)+1-\mu(A)\right)\right]^n \\
&= 2^{m_n} \max_{A\in\sigma(\mathcal{P}_n)} \left[1+\mu(A)\,(1-\mu(A))\,(e^{2s}+e^{-2s}-2)\right]^n \\
&\le 2^{m_n} \left[1+(e^{2s}+e^{-2s}-2)/4\right]^n \\
&= 2^{m_n} \left[1/2+(e^{2s}+e^{-2s})/4\right]^n . \qquad (5)
\end{aligned}
$$

This together with (1) implies that

$$\lambda_T(s) \le \ln(1/2+(e^{2s}+e^{-2s})/4). \qquad (6)$$

Similarly

$$
\begin{aligned}
\mathbf{E}\{e^{snT_n}\} &= \mathbf{E}\{\max_{A\in\sigma(\mathcal{P}_n)} e^{2sn(\mu_n(A)-\mu_n'(A))}\} \\
&\ge \max_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\{e^{2sn(\mu_n(A)-\mu_n'(A))}\} \\
&= \max_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\{e^{2sn\mu_n(A)}\}\mathbf{E}\{e^{-2sn\mu_n'(A)}\} \\
&= \max_{A\in\sigma(\mathcal{P}_n)} \left(e^{2s}\mu(A)+1-\mu(A)\right)^n \left(e^{-2s}\mu(A)+1-\mu(A)\right)^n \\
&= \max_{A\in\sigma(\mathcal{P}_n)} \left[\left(e^{2s}\mu(A)+1-\mu(A)\right) \left(e^{-2s}\mu(A)+1-\mu(A)\right)\right]^n \\
&= \max_{A\in\sigma(\mathcal{P}_n)} \left[1+\mu(A)\,(1-\mu(A))\,(e^{2s}+e^{-2s}-2)\right]^n .
\end{aligned}
$$

This together with (2) implies that

$$\lambda_T(s) \ge \ln(1/2+(e^{2s}+e^{-2s})/4). \qquad (7)$$

From (6) and (7) we deduce that

$$\lambda_T(s) = \ln(1/2+(e^{2s}+e^{-2s})/4).$$

The function $\lambda_T(s)$ is differentiable. Therefore, by the Gärtner-Ellis theorem (cf. Dembo and Zeitouni [7]), for all $0<\varepsilon<2$,

$$g_T(\varepsilon) = \max_{s>0}\left(s\varepsilon-\lambda_T(s)\right) = \max_{s>0}\left(s\varepsilon-\ln(1/2+(e^{2s}+e^{-2s})/4)\right).$$

One can verify that the maximum is achieved at

$$e^{2s} = \frac{1+\varepsilon/2}{1-\varepsilon/2}, \tag{8}$$

and then

$$\begin{aligned} g_T(\varepsilon) &= \varepsilon/2 \ln\left(\frac{1+\varepsilon/2}{1-\varepsilon/2}\right) + \ln\left(1-(\varepsilon/2)^2\right) \\ &= (1+\varepsilon/2)\ln(1+\varepsilon/2) + (1-\varepsilon/2)\ln(1-\varepsilon/2). \end{aligned}$$

■

The next proposition and Figure 1 clarify the respective positions of the rate functions $g_T(\varepsilon)$ and $g_L(\varepsilon)$.

Proposition 1 *For all* $0 < \varepsilon < 2$,

$$2g_L(\varepsilon/2) \le g_T(\varepsilon) \le g_L(\varepsilon). \tag{9}$$

Proof. We start with the right-hand side of inequality (9). By Jensen's inequality

$$\begin{aligned} \mathbf{E}\{e^{snT_n}\} &= \mathbf{E}\{\max_{A\in\sigma(\mathcal{P}_n)} e^{2sn(\mu_n(A)-\mu_n'(A))}\} \\ &\ge \max_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\{e^{2sn(\mu_n(A)-\mu_n'(A))}\} \\ &= \max_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\left\{\mathbf{E}\{e^{2sn(\mu_n(A)-\mu_n'(A))} \mid X_1,\ldots,X_n\}\right\} \\ &\ge \max_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\{e^{2sn(\mu_n(A)-\mathbf{E}\{\mu_n'(A)\mid X_1,\ldots,X_n\})}\} \\ &= \max_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\{e^{2sn(\mu_n(A)-\mu(A))}\}. \end{aligned}$$

The last term is also an upper bound on $2^{-m_n}\mathbf{E}\{e^{snL_n}\}$ (to see this, just adapt inequality (4) to the statistic $L_n = \sum_{j=1}^{m_n} |\mu_n(A_{nj}) - \mu(A_{nj})|$). Therefore

$$\lambda_T(s) \ge \lambda_L(s),$$

and so

$$g_T(\varepsilon) \le g_L(\varepsilon).$$

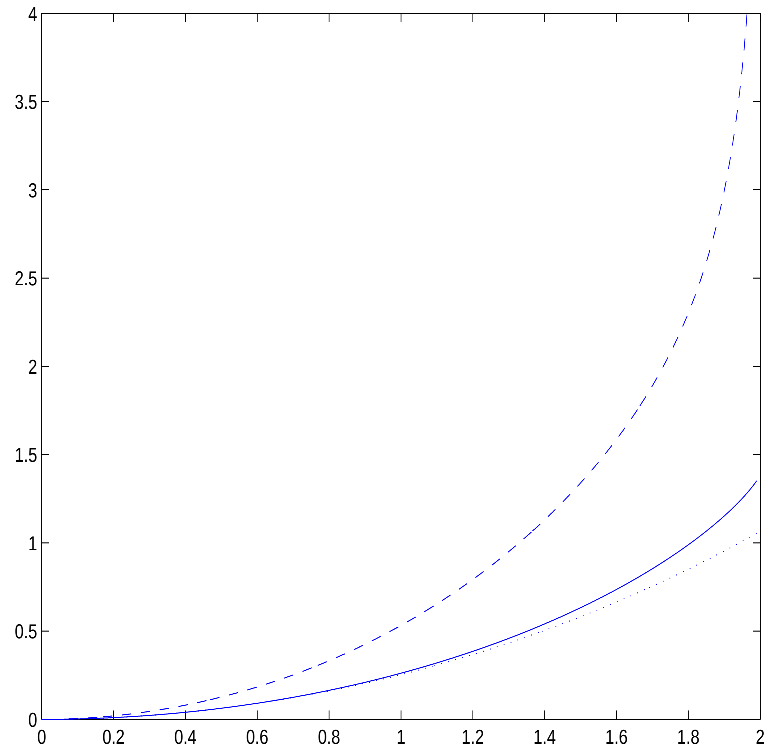

Figure 1: Rate functions $g_T(\varepsilon)$ (solid), $g_L(\varepsilon)$ (dashed) and $2g_L(\varepsilon/2)$ (dotted).

With respect to the lower bound, define

$$L'_n = \sum_{j=1}^{m_n} |\mu'_n(A_{nj}) - \mu(A_{nj})|.$$

Because of the triangle inequality

$$T_n \leq L_n + L'_n.$$

Consequently,

$$\begin{aligned}
\mathbf{E}\{e^{snT_n}\} &\leq \mathbf{E}\{e^{sn(L_n+L'_n)}\} \\
&= \mathbf{E}\{e^{snL_n}\}\mathbf{E}\{e^{snL'_n}\} \\
&= \mathbf{E}^2\{e^{snL_n}\}.
\end{aligned}$$

Thus

$$\lambda_T(s) \le 2\lambda_L(s),$$

and so

$$g_T(\varepsilon) = \max_{s>0} (s\varepsilon - \lambda_T(s)) \ge \max_{s>0} (s\varepsilon - 2\lambda_L(s)) = 2 \max_{s>0} (s\varepsilon/2 - \lambda_L(s)),$$

which implies that

$$g_T(\varepsilon) \ge 2g_L(\varepsilon/2).$$

■

Remark 1. Beirlant, Devroye, Györfi and Vajda [3] calculated $g_L(\varepsilon)$ using Sanov's theorem. The proof of Theorem 1 provides an alternative derivation of $g_L(\varepsilon)$. Indeed, by Scheffé's [18] theorem for partitions

$$L_n = \sum_{A\in\mathcal{P}_n} |\mu_n(A) - \mu(A)| = 2 \max_{A\in\sigma(\mathcal{P}_n)} (\mu_n(A) - \mu(A)).$$

Therefore

$$\begin{aligned}
\mathbf{E}\{e^{snL_n}\} &= \mathbf{E}\{\max_{A\in\sigma(\mathcal{P}_n)} e^{2sn(\mu_n(A)-\mu(A))}\} \\
&\le \sum_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\{e^{2sn(\mu_n(A)-\mu(A))}\} \\
&\le 2^{m_n} \max_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\{e^{2sn(\mu_n(A)-\mu(A))}\} \\
&= 2^{m_n} \max_{A\in\sigma(\mathcal{P}_n)} \mathbf{E}\{e^{2sn\mu_n(A)}\}e^{-2sn\mu(A)} \\
&= 2^{m_n} \max_{A\in\sigma(\mathcal{P}_n)} \left(e^{2s}\mu(A) + 1 - \mu(A)\right)^n \left(e^{-2s\mu(A)}\right)^n \\
&\le 2^{m_n} \left[\max_{x\in[0,1]} e^{-2sx} \left(e^{2s}x + 1 - x\right)\right]^n.
\end{aligned}$$

Condition (1) implies that

$$\lambda_L(s) \le \ln \left[\max_{x\in[0,1]} e^{-2sx} \left(e^{2s}x + 1 - x\right)\right].$$

We have to maximize on $[0,1]$ the function

$$x \mapsto e^{-2sx} \left(e^{2s}x + 1 - x\right).$$

The maximum is achieved at

$$x = \frac{e^{2s} - 1 - 2s}{2s(e^{2s} - 1)},$$

which results in

$$\lambda_L(s) \leq -1 + \frac{2s}{e^{2s} - 1} + \ln\left(\frac{e^{4s} - 2e^{2s} + 1}{2s(e^{2s} - 1)}\right), \tag{10}$$

and we may get, under (2), the same lower bound, so

$$g_L(\varepsilon) = \max_{s>0}\left(s\varepsilon - \lambda_L(s)\right) = \max_{s>0}\left[s\varepsilon + 1 - \frac{2s}{e^{2s} - 1} - \ln\left(\frac{e^{4s} - 2e^{2s} + 1}{2s(e^{2s} - 1)}\right)\right].$$

This is the same rate function as in Beirlant, Devroye, Györfi and Vajda [3], just in a different form. ■

Remark 2. By virtue of inequality (10), the sole condition (1) implies that

$$\mathbf{P}\{L_n > \varepsilon\} \leq e^{-n(g_L(\varepsilon)+o(1))} \quad \text{as } n \to \infty.$$

Therefore, by the Borel-Cantelli lemma, $L_n \to 0$ a.s. In other words, the goodness of fit test statistic L_n is strongly consistent, independently of the underlying distribution μ. ■

The technique of Theorem 1 yields a distribution-free strong consistent test of homogeneity, which rejects the null hypothesis if T_n becomes large. The concept of strong consistent test is quite unusual, it means that both on $\mathcal{H}_0$ and on its complement the test makes a.s. no error after a random sample size. In other words, denoting by $\mathbf{P}_0$ (*resp.* $\mathbf{P}_1$) the probability under the null hypothesis (*resp.* under the alternative), we have

$$\mathbf{P}_0\{\text{rejecting } \mathcal{H}_0 \text{ for only finitely many } n\} = 1$$

and

$$\mathbf{P}_1\{\text{accepting } \mathcal{H}_0 \text{ for only finitely many } n\} = 1.$$

In a real life problem, for example, when we get the data sequentially, one gets data just once, and should make good inference for these data. Strong consistency means that the single sequence of inference is a.s. perfect if the sample size is large enough. This concept is close to the definition of discernability introduced by Dembo and Peres [6]. For a discussion and references, we refer the reader to Devroye and Lugosi [10]. We insist on the fact that the test presented in Corollary 1 is entirely distribution-free, i.e., the measures μ and μ' are completely arbitrary.

Corollary 1 *Consider the test which rejects $\mathcal{H}_0$ when*

$$T_n > c_1\sqrt{\frac{m_n}{n}},$$

where

$$c_1 > 2\sqrt{\ln 2} \approx 1.6651.$$

Assume that condition (1) is satisfied and

$$\lim_{n\to\infty} \frac{m_n}{\ln n} = \infty.$$

Then, under $\mathcal{H}_0$, after a random sample size the test makes a.s. no error. Moreover, if

$$\mu \neq \mu',$$

and for any sphere S centered at the origin

$$\lim_{n\to\infty} \max_{A_{nj}\cap S\neq 0} \mathrm{diam}(A_{nj}) = 0, \tag{11}$$

then after a random sample size the test makes a.s. no error.

Proof. Under $\mathcal{H}_0$, we easily obtain from the proof of Theorem 1 (cf. (5) and (8)) a non-asymptotic bound for the tail of the distribution of T_n, namely

$$\mathbf{P}\{T_n > \varepsilon\} \leq \inf_{s>0} \frac{\mathbf{E}\{e^{snT_n}\}}{e^{sn\varepsilon}} \leq 2^{m_n} e^{-ng_T(\varepsilon)}.$$

Thus, by (3),

$$\begin{aligned}
\mathbf{P}\left\{T_n > c_1\sqrt{\frac{m_n}{n}}\right\} &\leq 2^{m_n} e^{-ng_T\left(c_1\sqrt{m_n/n}\right)} \\
&= 2^{m_n} e^{-nc_1^2(m_n/n)/4 + n\mathrm{o}(m_n/n)} \\
&= e^{-\left(c_1^2/4 - \ln 2 + \mathrm{o}(1)\right)m_n},
\end{aligned}$$

as $n \to \infty$. Therefore the condition $m_n / \ln n \to \infty$ implies that

$$\sum_{n=1}^{\infty} \mathbf{P}\left\{T_n > c_1 \sqrt{\frac{m_n}{n}}\right\} < \infty,$$

and by the Borel-Cantelli lemma we are ready with the first half of the corollary. Concerning the second half, apply the triangle inequality

$$T_n \geq \sum_{A \in \mathcal{P}_n} |\mu(A) - \mu'(A)| - L_n - L_n'.$$

According to Remark 2, $L_n \to 0$ and $L_n' \to 0$ a.s.. Moreover, we prove that by condition (11),

$$\sum_{A \in \mathcal{P}_n} |\mu(A) - \mu'(A)| \quad \to \quad 2 \sup_B |\mu(B) - \mu'(B)| > 0, \tag{12}$$

as $n \to \infty$, where the last supremum is taken over all Borel subsets of $\mathbb{R}^d$, and therefore

$$\liminf_{n \to \infty} T_n \geq 2 \sup_B |\mu(B) - \mu'(B)| > 0 \tag{13}$$

a.s. In order to show (12) we apply the technique from Barron, Györfi and van der Meulen [1]. Choose a measure λ which dominates μ and μ', for example, $\lambda = \mu + \mu'$, and denote by f the Radon-Nikodym derivative of $\mu - \mu'$ with respect to λ. Then, on the one hand,

$$\begin{aligned} \sum_{A \in \mathcal{P}_n} |\mu(A) - \mu'(A)| &= \sum_{A \in \mathcal{P}_n} \left| \int_A f \, \mathrm{d}\lambda \right| \\ &\leq \sum_{A \in \mathcal{P}_n} \int_A |f| \, \mathrm{d}\lambda \\ &= \int |f| \, \mathrm{d}\lambda \\ &= 2 \sup_B |\mu(B) - \mu'(B)|. \end{aligned}$$

On the other hand, for uniformly continuous f, using (11),

$$\sum_{A \in \mathcal{P}_n} \left| \int_A f \, \mathrm{d}\lambda \right| \to \int |f| \, \mathrm{d}\lambda.$$

If f is arbitrary then, for a given $\delta > 0$, choose a uniformly continuous $\tilde{f}$ such that

$$\int |f - \tilde{f}| \, \mathrm{d}\lambda < \delta.$$

Thus

$$\begin{aligned}
\sum_{A \in \mathcal{P}_n} \left| \int_A f \, \mathrm{d}\lambda \right| &\geq \sum_{A \in \mathcal{P}_n} \left| \int_A \tilde{f} \, \mathrm{d}\lambda \right| - \sum_{A \in \mathcal{P}_n} \left| \int_A (f - \tilde{f}) \, \mathrm{d}\lambda \right| \\
&\geq \sum_{A \in \mathcal{P}_n} \left| \int_A \tilde{f} \, \mathrm{d}\lambda \right| - \int |f - \tilde{f}| \, \mathrm{d}\lambda \\
&\geq \sum_{A \in \mathcal{P}_n} \left| \int_A \tilde{f} \, \mathrm{d}\lambda \right| - \delta \\
&\to \int |\tilde{f}| \, \mathrm{d}\lambda - \delta \\
&\geq \int |f| \, \mathrm{d}\lambda - 2\delta \\
&= 2 \sup_B |\mu(B) - \mu'(B)| - 2\delta.
\end{aligned}$$

The result follows since δ was arbitrary. ■

We conclude this section by mentioning that the problem of the good choice of the partition in T_n is a difficult one. The rectangle partition is 'good' if the cell's probabilities are approximately equal. A promising direction of further study would be to consider random partitions, where, for example, the cells are statistically equivalent blocks with respect to the sample $X_1, \ldots, X_n$ (see Devroye, Györfi and Lugosi [9], Chapter 21).

3 Asymptotic normality

Beirlant, Györfi and Lugosi [4] proved that, under conditions (1) and (2),

$$\sqrt{n} \left(L_n - \mathbf{E}\{L_n\} \right) / \sigma \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where $\xrightarrow{\mathcal{D}}$ stands for the convergence in distribution and $\sigma^2 = 1 - 2/\pi$. The technique of Beirlant, Györfi and Lugosi [4] involves a Poisson representation of the empirical process in conjunction with Bartlett's [2] idea of partial inversion for obtaining characteristic functions of conditional distributions. Using the method of these authors, we can prove the following:

Theorem 2 *Assume that conditions (1) and (2) are satisfied. Then, under $\mathcal{H}_0$, there exists a centering sequence $(C_n)_{n\geq 1}$ depending on μ such that*

$$\sqrt{n}\,(T_n - C_n)\,/\sigma \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where $\sigma^2 = 2(1-2/\pi)$. (The centering constant C_n will be defined at the beginning of the proof of Theorem 2).

The main difficulty in proving Theorem 2 is that it states asymptotic normality of T_n, which is a sum of *dependent* random variables. To overcome this problem, we use a 'Poissonization' argument originating from the fact that an empirical process is equal in distribution to the conditional distribution of a Poisson process given the sample size (for more on Poissonization techniques, we refer the reader to Beirlant, Györfi and Lugosi [4], Beirlant and Mason [5], and Giné, Mason and Zaitsev [11]).

To go straight to the point, for each $n \geq 1$, denote by N_n and N_n' two independent Poisson (n) random variables, defined on the same probability space as the sequences $(X_i)_{i\geq 1}$ and $(X_i')_{i\geq 1}$, and independent of these sequences. The Poissonized version $\tilde{T}_n$ of T_n is then defined by

$$\tilde{T}_n = \sum_{j=1}^{m_n} |\mu_{N_n}(A_{nj}) - \mu'_{N_n'}(A_{nj})|,$$

where, for any Borel subset A,

$$\mu_{N_n}(A) = \frac{\#\{i : X_i \in A, i = 1, \ldots, N_n\}}{n},$$

and, similarly,

$$\mu'_{N_n'}(A) = \frac{\#\{i : X_i' \in A, i = 1, \ldots, N_n'\}}{n}.$$

Clearly, for each $j \in \{1, \ldots, m_n\}$,

$$n\mu_{N_n}(A_{nj}) = \sum_{i=1}^{N_n} \mathbf{1}_{[X_i \in A_{nj}]}$$

and

$$n\mu'_{N'_n}(A_{nj}) = \sum_{i=1}^{N'_n} \mathbf{1}_{[X'_i \in A_{nj}]}.$$

Thus, setting

$$\mathbf{Y}_n = (n\mu_{N_n}(A_{n1}), \ldots, n\mu_{N_n}(A_{nm_n}))$$

and

$$\mathbf{Y}'_n = (n\mu'_{N'_n}(A_{n1}), \ldots, n\mu'_{N'_n}(A_{nm_n})),$$

it is a simple exercise to show that $\mathbf{Y}_n$ and $\mathbf{Y}'_n$ are independent vectors of independent random variables with

$$(n\mu_{N_n}(A_{nj})) \stackrel{\mathcal{D}}{=} (n\mu'_{N'_n}(A_{nj})) \stackrel{\mathcal{D}}{=} \text{Poisson}\,(n\mu(A_{nj}))\,.$$

Moreover,

$$(\mathbf{Y}_n | N_n = n) \stackrel{\mathcal{D}}{=} (\mathbf{Y}'_n | N'_n = n) \stackrel{\mathcal{D}}{=} \text{Multinomial}\,(n; \mu(A_{n1}), \ldots, \mu(A_{nm_n}))\,.$$

The key of the proof of Theorem 2 is the following property, which is a slight extension of the proposition page 311 in Beirlant, Györfi and Lugosi [4]. The notation $\mathcal{N}_3(\mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2)$ will stand for the trivariate normal distribution with means μ_1, μ_2, μ_3, variances $\sigma_1^2, \sigma_2^2, \sigma_3^3$, and independent components.

Proposition 2 *Let g_{nj} ($n \geq 1$, $j = 1, \ldots, m_n$) be real measurable functions, with*

$$\mathbf{E}\Big\{ g_{nj}\left(\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})\right) \Big\} = 0,$$

and let

$$M_n = \sum_{j=1}^{m_n} g_{nj}\left(\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})\right).$$

Assume that

$$\left(M_n, \frac{N_n - n}{\sqrt{n}}, \frac{N'_n - n}{\sqrt{n}}\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}_3(0, 0, 0, \sigma^2, 1, 1) \quad \text{as } n \to \infty,$$

where σ is a positive constant. Then

$$\frac{1}{\sigma} \sum_{j=1}^{m_n} g_{nj}\left(\mu_n(A_{nj}) - \mu'_n(A_{nj})\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0,1).$$

Sketch of proof of Proposition 2. Consider the characteristic functions

$$\Phi_n(t,u,u') \stackrel{\text{def}}{=} \mathbf{E}\left\{\exp\left(itM_n + iuN_n + iu'N_n'\right)\right\}$$

and

$$\Psi_n(t) \stackrel{\text{def}}{=} \mathbf{E}\left\{\exp\left(it\sum_{j=1}^{m_n} g_{nj}\left(\mu_n(A_{nj}) - \mu_n'(A_{nj})\right)\right)\right\}.$$

Clearly,

$$\Psi_n(t) = \mathbf{E}\left\{\exp\left(itM_n\right)|(N_n,N_n') = (n,n)\right\}, \tag{14}$$

and

$$\Phi_n(t,u,u') = \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} e^{iuk+iu'k'}\mathbf{E}\left\{\exp\left(itM_n\right)|(N_n,N_n') = (k,k')\right\}p_n(k,k'),$$

with

$$p_n(k,k') = \mathbf{P}\left\{(N_n,N_n') = (k,k')\right\} = e^{-2n}n^{k+k'}/(k!k'!)\,.$$

From this, by Fourier's inversion formula, it follows that

$$\begin{aligned}&\mathbf{E}\left\{\exp\left(itM_n\right)|(N_n,N_n') = (n,n)\right\}\\ &\quad= \frac{1}{(2\pi)^2 p_n(n,n)}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{-i(u+u')n}\Phi_n(t,u,u')\,\mathrm{d}u\mathrm{d}u'.\end{aligned}$$

Stirling's formula gives

$$(2\pi)^2 p_n(n,n) = (2\pi)^2 e^{-2n} n^{2(n-1)}/\left((n-1)!\right)^2 \sim 2\pi/n \text{ as } n\to\infty.$$

Hence, substituting v for $u\sqrt{n}$ and v' for $u'\sqrt{n}$, we get by (14) that

$$\Psi_n(t) = \frac{1}{2\pi}\left(1+\mathrm{o}(1)\right)\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}}\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} e^{-i(v+v')\sqrt{n}}\,\Phi_n(t,v/\sqrt{n},v'/\sqrt{n})\,\mathrm{d}v\mathrm{d}v'.$$

By the assumption of the proposition

$$e^{-i(v+v')\sqrt{n}}\,\Phi_n(t,v/\sqrt{n},v'/\sqrt{n}) \to e^{-t^2\sigma^2/2}e^{-(v^2+v'^2)/2} \quad \text{as } n\to\infty.$$

Therefore, by a variant of the dominated convergence theorem (see Rao [16], page 136), it follows that

$$\Psi_n(t) \to e^{-t^2\sigma^2/2}.$$

$\blacksquare$

We are now ready to prove Theorem 2.

Proof of Theorem 2. We will show the theorem with the centering constant

$$C_n = \sum_{j=1}^{m_n} \mathbf{E}|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})|.$$

We use Proposition 2 to prove that

$$\frac{\sqrt{n}}{\sigma}\sum_{j=1}^{m_n}\Big(|\mu_n(A_{nj}) - \mu'_n(A_{nj})| - \mathbf{E}|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})|\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad (15)$$

where we recall that $\sigma^2 = 2(1-2/\pi)$. To prove (15), we choose the functions g_{nj} as

$$g_{nj}(x) = \sqrt{n}\Big(|x| - \mathbf{E}\left|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})\right|\Big), \quad j = 1, \ldots, m_n,$$

and we check the conditions of Proposition 2. Introduce

$$\begin{aligned} S_n &= t\sqrt{n}\sum_{j=1}^{m_n}\Big(|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})| - \mathbf{E}|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})|\Big) \\ &\quad + v\frac{N_n - n}{\sqrt{n}} + v'\frac{N'_n - n}{\sqrt{n}}, \end{aligned}$$

for which a central limit result is to be shown to hold. Observe first that

$$\begin{aligned} \mathbf{Var}\{S_n\} &= n\sum_{j=1}^{m_n}\Big[t^2\mathbf{Var}|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})| \\ &\qquad + 2tv\mathbf{E}\left\{|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})|\left(\mu_{N_n}(A_{nj}) - \mu(A_{nj})\right)\right\} \\ &\qquad + 2tv'\mathbf{E}\left\{|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})|\Big(\mu'_{N'_n}(A_{nj}) - \mu(A_{nj})\Big)\right\}\Big] \\ &\quad + v^2 + v'^2, \end{aligned}$$

as $\mathbf{E}\{n\mu_{N_n}(A_{nj})\} = \mathbf{Var}\{n\mu_{N_n}(A_{nj})\} = n\mu(A_{nj})$, $j = 1, \ldots, m_n$. In Lemma 1 at the end of this section, we bring together some technical results on the Poisson distribution, which will be used here. From (c) in Lemma 1, for any $\varepsilon > 0$, one can choose s_0 such that for $s > s_0$

$$2(1-2/\pi) - \varepsilon \leq \frac{\mathbf{Var}|U - U'|}{s} \leq 2(1-2/\pi) + \varepsilon,$$

where U and U' are independent Poisson (s) random variables. Thus, on the one hand, we get that

$$\begin{aligned}
& n \sum_{j=1}^{m_n} \mathbf{Var}|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})| \\
& \leq \sum_{j=1}^{m_n} \mu(A_{nj}) \left\{ 2\mathbf{1}_{[n\mu(A_{nj}) \leq s_0]} + (2(1-2/\pi)+\varepsilon)\mathbf{1}_{[n\mu(A_{nj}) > s_0]} \right\} \\
& \leq 2(1-2/\pi) + \varepsilon + \sum_{j=1}^{m_n} 2\mu(A_{nj})\mathbf{1}_{[n\mu(A_{nj}) \leq s_0]} \\
& \leq 2(1-2/\pi) + \varepsilon + \sum_{j=1}^{m_n} \frac{2s_0}{n}\mathbf{1}_{[n\mu(A_{nj}) \leq s_0]} \\
& \leq 2(1-2/\pi) + \varepsilon + \frac{2m_n s_0}{n} \\
& \to 2(1-2/\pi) + \varepsilon \quad \text{as } n \to \infty.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
& n \sum_{j=1}^{m_n} \mathbf{Var}|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})| \\
& \geq (2(1-2/\pi) - \varepsilon) \sum_{j=1}^{m_n} \mu(A_{nj})\mathbf{1}_{[n\mu(A_{nj}) > s_0]} \\
& = (2(1-2/\pi) - \varepsilon) \left(1 - \sum_{j=1}^{m_n} \mu(A_{nj})\mathbf{1}_{[n\mu(A_{nj}) \leq s_0]} \right) \\
& \to 2(1-2/\pi) - \varepsilon \quad \text{as } n \to \infty.
\end{aligned}$$

Therefore, since ε was arbitrary,

$$n \sum_{j=1}^{m_n} \mathbf{Var}|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})| \to 2(1-2/\pi).$$

To complete the asymptotics of $\mathbf{Var}\{S_n\}$, it remains to show that

$$n\sum_{j=1}^{m_n}\mathbf{E}\left\{|\mu_{N_n}(A_{nj})-\mu'_{N'_n}(A_{nj})|\,(\mu_{N_n}(A_{nj})-\mu(A_{nj}))\right\}\to 0 \quad \text{as } n\to\infty.$$

This however follows from (d) in Lemma 1 using a similar argument as above. To finish the proof of

$$S_n \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, t^2\sigma^2+v^2+v'^2\right) \qquad \text{as } n\to\infty,$$

by Lyapunov's central limit theorem, it suffices to prove that

$$n^{3/2}\Bigg(\sum_{j=1}^{m_n}\mathbf{E}\Big\{\Big|\,t\left(|\mu_{N_n}(A_{nj})-\mu'_{N'_n}(A_{nj})|-\mathbf{E}|\mu_{N_n}(A_{nj})-\mu'_{N'_n}(A_{nj})|\right)$$
$$+v(\mu_{N_n}(A_{nj})-\mu(A_{nj}))+v'(\mu'_{N'_n}(A_{nj})-\mu(A_{nj}))\Big|^3\Big\}\Bigg)$$

goes to 0, or, by invoking the c_r-inequality, that

$$n^{3/2}\sum_{j=1}^{m_n}\mathbf{E}\left\{\Big||\mu_{N_n}(A_{nj})-\mu'_{N'_n}(A_{nj})|-\mathbf{E}|\mu_{N_n}(A_{nj})-\mu'_{N'_n}(A_{nj})|\Big|^3\right\}\to 0,$$

and

$$n^{3/2}\sum_{j=1}^{m_n}\mathbf{E}|\mu_{N_n}(A_{nj})-\mu(A_{nj})|^3\to 0,$$

as $n\to\infty$. From (b) in Lemma 1, we deduce that

$$\begin{aligned}
&n^{3/2}\sum_{j=1}^{m_n}\mathbf{E}\left\{\Big||\mu_{N_n}(A_{nj})-\mu'_{N'_n}(A_{nj})|-\mathbf{E}|\mu_{N_n}(A_{nj})-\mu'_{N'_n}(A_{nj})|\Big|^3\right\}\\
&\le Cn^{-3/2}\left(\sum_{j=1}^{m_n}n\mu(A_{nj})+\sum_{j=1}^{m_n}n^{3/2}\mu(A_{nj})^{3/2}\right)\\
&\quad\text{(where } C \text{ is a positive universal constant)}\\
&\le C\left(\frac{1}{\sqrt{n}}+\max_{j=1,\ldots,m_n}\mu(A_{nj})^{1/2}\right)\to 0,
\end{aligned}$$

as $n\to\infty$. Analogously, using (a) in Lemma 1, the second limit is shown to be zero. ■

Theorem 2 yields the asymptotic null distribution of a consistent homogeneity test, which rejects the null hypothesis if T_n becomes large. In contrast to Corollary 1, and because of condition (2), this new test is *not* distribution-free. In particular, the measures μ and μ' have to be nonatomic.

Corollary 2 *Put* $\alpha \in (0,1)$, *and let* $C^* \approx 0.7655$ *denote the universal constant in Lemma 2. Consider the test which rejects* $\mathcal{H}_0$ *when*

$$T_n > c_2 \sqrt{\frac{m_n}{n}} + C^* \frac{m_n}{n} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1-\alpha),$$

where

$$\sigma^2 = 2(1 - 2/\pi) \quad \text{and} \quad c_2 = \frac{2}{\sqrt{\pi}} \approx 1.1284,$$

and where Φ *denotes the standard normal distribution function. Then, under the conditions of Theorem 2, the test has asymptotic significance level* α. *Moreover, under the additional condition (11), the test is consistent.*

Proof. According to Theorem 2, under $\mathcal{H}_0$,

$$\mathbf{P}\{\sqrt{n}(T_n - C_n)/\sigma \leq x\} \approx \Phi(x),$$

therefore the error probability with threshold x is

$$\alpha = 1 - \Phi(x).$$

Thus the α-level test rejects the null hypothesis if

$$T_n > C_n + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1-\alpha).$$

However, C_n depends on the unknown distribution, thus we apply an upper bound on C_n, and so decrease the error probability. The following chain of inequalities is valid:

$$\begin{aligned} C_n &= \sum_{j=1}^{m_n} \mathbf{E}|\mu_{N_n}(A_{nj}) - \mu'_{N'_n}(A_{nj})| \\ &\leq \sum_{j=1}^{m_n} \sqrt{\frac{2\,\mu(A_{nj})}{n}} \mathbf{E}|Z_j| + C^* \frac{m_n}{n} \end{aligned}$$

$$
\begin{aligned}
& \quad \text{(by Lemma 2, where } Z_1, \ldots, Z_{m_n} \text{ are i.i.d. standard normal)} \\
= & \sum_{j=1}^{m_n} \sqrt{\frac{\mu(A_{nj})}{n}} \frac{2}{\sqrt{\pi}} + C^* \frac{m_n}{n} \\
\leq & \; c_2 \sqrt{\frac{m_n}{n}} + C^* \frac{m_n}{n} \\
& \quad \text{(by Jensen's inequality).}
\end{aligned}
$$

Thus

$$
\begin{aligned}
\alpha & \approx \mathbf{P}\left\{ T_n > C_n + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1-\alpha) \right\} \\
& \geq \mathbf{P}\left\{ T_n > c_2 \sqrt{\frac{m_n}{n}} + C^* \frac{m_n}{n} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1-\alpha) \right\}.
\end{aligned}
$$

This proves that the test has asymptotic error probability at most α.

Under $\mu \neq \mu'$, the consistency of the test follows from (13). ■

Note that, by condition (1),

$$
c_2 \sqrt{\frac{m_n}{n}} + C^* \frac{m_n}{n} + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1-\alpha) = c_2 \sqrt{\frac{m_n}{n}} \left(1 + \mathrm{o}(1)\right),
$$

therefore the order of the threshold does not depend on the level α.

In the sequel, the letter C will denote a positive universal constant, whose value may vary from line to line. The notation Z stands for the standard normal random variable.

Lemma 1 *Let U and U' be independent Poisson (s) random variables. Then*

$$
\begin{aligned}
(a) & \qquad \mathbf{E}|U - s|^3 \leq C\,(s + s^{3/2}). \\
(b) & \qquad \mathbf{E}\Big| |U - U'| - \mathbf{E}|U - U'| \Big|^3 \leq C\,(s + s^{3/2}). \\
(c) & \qquad \lim_{s \to \infty} \left| \mathbf{E} \left| \frac{U - U'}{\sqrt{2s}} \right| - \mathbf{E}|Z| \right| = 0,
\end{aligned}
$$

$$\textit{and} \qquad \lim_{s\to\infty}\left|\mathbf{E}\left|\frac{U-U'}{\sqrt{2s}}\right|^2-\mathbf{E}|Z|^2\right|=0.$$

$$\textit{Consequently,} \qquad \lim_{s\to\infty}\frac{\mathbf{Var}|U-U'|}{s}=2\mathbf{Var}|Z|=2(1-2/\pi).$$

$$(d) \qquad \lim_{s\to\infty}\frac{\mathbf{E}\{|U-U'|(U-s)\}}{s}=0.$$

Proof. (a) is a consequence of extension of Rosenthal's inequality to Poissonized sums of random variables (see for example Giné, Mason and Zaitsev [11], Lemma 2.3, page 730).

To prove (b), observe that by the c_r- and Jensen's inequalities, we have

$$\mathbf{E}\Big||U-U'|-\mathbf{E}|U-U'|\Big|^3\le C\mathbf{E}\left|U-s\right|^3,$$

and apply (a).

Let us now turn to the proof of (c) and (d). To this aim, observe first that

$$U\overset{\mathcal{D}}{=}\sum_{i=1}^n\zeta_i,$$

where the ζ_i's are independent Poisson (s/n) random variables. Similarly, set

$$U'\overset{\mathcal{D}}{=}\sum_{i=1}^n\zeta_i'.$$

The result is related to the classical Berry-Esseen theorem and will follow from Theorem 1 of Sweeting [20]. More precisely, according to Fact 6.1, inequality (6.18), page 746, in Giné, Mason and Zaitsev [11], and using (a), we can write

$$\left|\mathbf{E}\left|\frac{U-U'}{\sqrt{2s}}\right|-\mathbf{E}|Z|\right|\le\frac{C}{\sqrt{n}}\left(1+(s/n)^{-1/2}\right),$$

$$\left|\mathbf{E}\left|\frac{U-U'}{\sqrt{2s}}\right|^2-\mathbf{E}|Z|^2\right|\le\frac{C}{\sqrt{n}}\left(1+(s/n)^{-1/2}\right),$$

and

$$\left|\frac{\mathbf{E}\{|U-U'|(U-s)\}}{s}\right|\le\frac{C}{\sqrt{n}}\left(1+(s/n)^{-1/2}\right).$$

Taking for n the upper part of s leads to the desired result. ∎

Lemma 2 *Let B and B' be independent Binomial (n, p) random variables, $n \in \mathbb{N}^*$ and $p \in (0,1)$. Then*

$$\left| \mathbf{E} \left| \frac{B - B'}{\sqrt{2np(1-p)}} \right| - \mathbf{E}|Z| \right| \leq \frac{C^*}{\sqrt{2np(1-p)}},$$

where $C^* \approx 0.7655$.

Proof. Note that

$$B \stackrel{\mathcal{D}}{=} \sum_{i=1}^{n} w_i,$$

where the w_i's are independent Bernoulli (p) random variables, and apply Fact 6.1, inequality (6.18), page 746, in Giné, Mason and Zaitsev [11]. Note that C^* is the Berry-Essen constant. The best current estimate is 0.7655 (Shiganov [19]). ■

Acknowledgments. The authors thank two anonymous referees for their valuable comments and suggestions.

References

[1] Barron, A. R., Györfi, L. and van der Meulen, E. C. (1992). Distribution estimation consistent in total variation and in two types of information divergence, *IEEE Trans. Inform. Theory*, 38, pp. 1437-1454.

[2] Bartlett, M. S. (1938). The characteristic function of a conditional statistic, *J. London Math. Soc.*, 13, pp. 62-67.

[3] Beirlant, J., Devroye, L., Györfi, L. and Vajda, I. (2001). Large deviations of divergence measures on partitions, *J. Statist. Plann. Inference*, 93, pp. 1–16.

[4] Beirlant, J., Györfi, L. and Lugosi, G. (1994). On the asymptotic normality of the L_1- and L_2-errors in histogram density estimation, *Canad. J. Statist.*, 22, pp. 309-318.

[5] Beirlant, J. and Mason, D. M. (1995). On the asymptotic normality of L_p-norms of empirical functionals, *Math. Methods Statist.*, 4, pp. 1-19.

[6] Dembo, A. and Peres, Y. (1994). A topological criterion for hypothesis testing, *Ann. Statist.*, 22, pp. 106–117.

[7] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications, Second Edition*, Springer-Verlag, New York.

[8] Devroye, L. and Györfi, L. (2002). Distribution and density estimation, in *Principles of Nonparametric Learning*, (L. Györfi, Ed.), Springer-Verlag, Wien, pp. 223-286.

[9] Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*, Springer-Verlag, New York.

[10] Devroye, L. and Lugosi, G. (2002). Almost sure classification of densities, 14, *J. Nonparametr. Stat.*, pp. 675-698.

[11] Giné, E., Mason, D. M. and Zaitsev, A. Yu (2003). The L_1-norm density estimator process, *Ann. Probab.*, 31, pp. 719-768.

[12] Greenwood, P. E. and Nikulin, M. S. (1996). *A Guide to Chi-Squared Testing*, Wiley, New York.

[13] Györfi, L. and van der Meulen, E. C. (1990). A consistent goodness of fit test based on the total variation distance, in *Nonparametric Functional Estimation and Related Topics*, (G. Roussas, Ed.), Kluwer, Dordrecht, pp. 631-645.

[14] Pardo, M. C., Pardo, L. and Vajda, I. (2004). Testing homogeneity of independent samples from arbitrary models, *Research Report, No 2104, Institute of Automation and Information Theory of the Czech Academy of Sciences.*

[15] Pardo, L., Pardo, M. C. and Zografos, K. (1999). Homogeneity for multinomial populations based on ϕ-divergences, *J. Japan Statist. Soc.*, 29, pp. 213–228.

[16] Rao, C. R. (1973). *Statistical Inference and its Applications, Second Edition*, Wiley, New York.

[17] Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness-of-fit Statistics for Discrete Multivariate Data*, Springer-Verlag, Berlin.

[18] Scheffé, H. (1947). A useful convergence theorem for probability distributions, *Ann. Math. Statist.*, 18, pp. 434–438.

[19] Shiganov, I.S. (1986). Refinement of the upper bound of the constant in the central limit theorem (English), *J. Soviet Math.*, 35, 2545–2550.

[20] Sweeting, T. J. (1977). Speeds of convergence for the multidimensional central limit theorem, *Ann. Probab.*, 5, pp. 28-41.