



Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Poisson limit theorems for the Cressie–Read statistics

Grzegorz Rempala^{a,*}, Jacek Wesołowski^b^a The Ohio State University, The United States of America^b Politechnika Warszawska, Poland

ARTICLE INFO

Article history:

Received 21 June 2021

Received in revised form 3 July 2022

Accepted 22 July 2022

Available online 29 July 2022

MSC:

60F05

62G20

Keywords:

Power divergence

Multinomial model

Martingale limit theorem

ABSTRACT

We derive Poisson limit theorems for the multinomial Cressie–Read goodness-of-fit statistics and some of their modifications under the assumption that as the sample size n goes to infinity, the number of groups m_n increases with the sample size. Our results extend the Poisson limit theorems for the Pearson chi-square statistic reported in Rempala and Wesołowski (2016).

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

The “double-asymptotic” paradigm of data analysis in which the parameter space increases with the sample size seems increasingly important in the modern era of “big-data” statistics. Perhaps the most natural example is the high dimensional multinomial model where the number of categories can grow with, and possibly exceed, the sample size. In the context of statistical goodness-of-fit testing for survey responses, such high dimensional multinomial problem was considered already in Fienberg and Holland (1973) and then studied in more detail, for instance, in Morris (1975) and Cressie and Read (1984).

The double-asymptotic paradigm is also natural in many statistical problems of modern genomic research where often the number of observed outcomes explicitly depends upon the size of a collected sample. In the last two decades, a great deal of effort has been dedicated towards understanding how to extend some of the results of classical inferential statistical analysis to such settings. Several examples in molecular studies were given, for instance, in Browning et al. (2018) or Pietrzak et al. (2016) and both the goodness of fit and hypothesis testing in double-asymptotic models were recently considered in Chen and Qin (2010), Kim et al. (2020), Janková et al. (2020).

Motivated by these and other similar applications where the classical goodness-of-fit tests often fail (Haberman, 1988; Kim, 2020), we establish in this paper some general results for the class of the so-called multinomial Cressie–Read (CR) goodness-of-fit statistics (Cressie and Read, 1984) (and some of its close relatives) for triangular arrays of discrete random variables for which the number of classes m_n is either infinite or grows with the sample size n .

In statistical practice, it is often assumed that the CR goodness-of-fit statistics that are asymptotically (as $n \rightarrow \infty$) chi-square distributed (for instance, the Pearson chi-square, log-likelihood ratio, or the Freeman–Tukey statistic), must be also asymptotically normal when both $m_n, n \rightarrow \infty$. Although this assumption is indeed often warranted, see e.g. Tumanyan

* Corresponding author.

E-mail address: grem Pala@cph.osu.edu (G. Rempala).

(1956), Cressie and Read (1984), Menéndez et al. (1998), Inglot et al. (1991), Pérez and Pardo (2002), Ogata and Taniguchi (2009), Mirakhmedov (2016), there are also instances when it is clearly false. A recent example was given, for instance, in the paper Rempała and Wesolowski (2016) where it was shown that when m_n is of order n^2 the Pearson chi-square statistic is asymptotically Poisson (in the case of a uniform distribution this result was already noted much earlier in Steck (1957)). In the current paper we show that the Poisson limit result extends to a broad sub-class of the CR-statistics that includes, for example, a log-likelihood ratio statistic. We are currently unaware of any other similar asymptotic results for the CR goodness-of-fit statistics.

The paper is organized as follows. In the next section we present our main result on the Poisson limit of the CR-statistics when both the sample size and the number of multinomial classes increase to infinity. We show that our result extends in particular the Pearson chi-square double-asymptotic theorem from Rempała and Wesolowski (2016). We also give some additional examples and numerical illustrations. In the following section we consider a closely related class of CR-statistics (dubbed *uniform* CR-statistics) that arises naturally when analyzing power of the goodness of fit testing of the uniform distribution against certain class of alternatives — an important problem in high dimensional signal detection (Pour et al., 2020; Pietrzak and Rempała, 2017). This result is of particular interest since, in general, the power of the tests based on CR-statistics is poorly understood in high dimensional problems (see, for instance the recent discussion on the Pearson chi-square in Berrett and Samworth, 2021). The proofs are given in Section 3. Although relatively elementary, they are moderately cumbersome and technically involved. The techniques rely on martingale methods, in particular on the Poisson Principle of Conditioning established in Beška et al. (1982). The result is quoted for readers' convenience in the first part of Section 3. Finally, in a short Appendix we give a partial proof of the conjecture on the non-decreasing efficiency (in terms of the variance) of the modified CR-statistics vs its classical counterpart. At present, we are only able to show this result for positive integer values of the canonical parameter λ .

2. The classical CR-statistic

For any $n \geq 1$ consider iid r.v.'s $X_n, X_{n,1}, \dots, X_{n,n}$ with a discrete distribution

$$\mathbb{P}(X_n = j) = p_n(j), \quad j \in M_n.$$

Denote also

$$\hat{p}_n(j) = \frac{N_n(j)}{n},$$

where

$$N_n(j) = \sum_{k=1}^n I_{X_{n,k}=j}, \quad j \in M_n.$$

The CR-statistic (known also as the Cressie–Read goodness-of-fit statistic or the power divergence statistic), see Cressie and Read (1984), has the form

$$CR_\lambda(n) = n \sum_{j \in M_n} p_n(j) g_\lambda \left(\frac{\hat{p}_n(j)}{p_n(j)} \right), \quad \lambda > -1, \quad (1)$$

where $g_\lambda : [0, \infty) \rightarrow [0, \infty)$ is defined as

$$g_\lambda(x) = \begin{cases} 2x^{\frac{\lambda+1}{\lambda}}, & \lambda \neq 0, \\ 2x \log x, & \lambda = 0 \text{ (by continuity)}. \end{cases}$$

As already indicated in the introduction, we refer to (1) as the *CR-statistic*. Important examples are

- the chi-squared statistic

$$\chi^2(n) := CR_1(n) = n \sum_{j \in M_n} \frac{(\hat{p}_n(j) - p_n(j))^2}{p_n(j)},$$

- the log-likelihood ratio statistic

$$G^2(n) := CR_0(n) = 2n \sum_{j \in M_n} \hat{p}_n(j) \log \left(\frac{\hat{p}_n(j)}{p_n(j)} \right).$$

It is convenient to define the random variables $P_n := p_n(X_n)$, $n \geq 1$. Note that when $\#M_n =: m_n$ is finite we have $\mathbb{E} P_n^{-1} = m_n$, $n \geq 1$. In such case double asymptotics, i.e. the limit when both $n, m_n \rightarrow \infty$, for the chi-square statistics was considered in Rempała and Wesolowski (2016). In particular, for a special regime of $n, m_n \rightarrow \infty$ the limiting distribution turns out to be Poissonian as asserted by the following.

Theorem 2.1 (Rempała and Wesółowski, 2016). Assume that

$$\frac{n^2}{m_n} \rightarrow \eta > 0, \quad (2)$$

$$np_n^* \rightarrow 0 \quad (3)$$

where $p_n^* = \sup_{j \in M_n} p_n(j)$ and

$$\frac{1}{nm_n} \text{Var} P_n^{-1} \rightarrow 0. \quad (4)$$

Then $\chi^2(n)$ is asymptotically Poisson. More precisely,

$$\tilde{\chi}^2(n) := \frac{\chi^2(n) - m_n}{\sqrt{2m_n}} \xrightarrow{d} \sqrt{\frac{\eta}{2}} \text{Pois}\left(\frac{\eta}{2}\right) - \sqrt{\frac{\eta}{2}}. \quad (5)$$

Our main result of the current paper is the following extension of Theorem 2.1 to CR-statistics.

Theorem 2.2 (Poisson Asymptotics for CR-statistics). Assume that the condition (3) is satisfied. For $\lambda \geq 0$ let

$$\beta_n(r, \lambda) := n^2 \frac{\mathbb{E} P_n^{1-r\lambda}}{(\mathbb{E} P_n^{-\lambda})^r} \rightarrow \eta > 0, \quad \text{for } r = 1, 2, 3. \quad (6)$$

and

$$n \frac{\text{Var} P_n^{-\lambda}}{(\mathbb{E} P_n^{-\lambda})^2} \rightarrow 0, \quad \text{when } \lambda > 0, \quad n \text{Var} \log P_n \rightarrow 0, \quad \text{when } \lambda = 0. \quad (7)$$

Then

$$\tilde{\text{CR}}_\lambda(n) := n^\lambda \frac{\text{CR}_\lambda(n) - n^2 \mathbb{E} P_n g_\lambda\left(\frac{1}{nP_n}\right)}{\mathbb{E} P_n^{-\lambda}} \xrightarrow{d} g_\lambda(2) \text{Pois}\left(\frac{\eta}{2}\right). \quad (8)$$

Thus the above theorem states that the properly normalized CR statistic is asymptotically Poisson distributed. We note

Remark 2.1. Theorem 2.2 specializes to Theorem 2.1 when $\lambda = 1$. Indeed, observe first that (4) is equivalent to (7) for $\lambda = 1$. Then since $\mathbb{E} P_n^{-1} = m_n$ we see that (6) becomes

$$\beta_n(1, 1) = \beta_n(2, 1) = \frac{n^2}{m_n}, \quad \beta_n(3, 1) = \frac{n^2}{m_n^3} \mathbb{E} P_n^{-2}. \quad (9)$$

Thus conditions (6) for $r = 1, 2$ and (2) coincide. Since (4) can be written as

$$\frac{m_n}{n} \left(\frac{1}{m_n^2} \mathbb{E} P_n^{-2} - 1 \right) \rightarrow 0$$

and by (2) we have $\frac{m_n}{n} \rightarrow \infty$, thus

$$\frac{1}{m_n^2} \mathbb{E} P_n^{-2} \rightarrow 1. \quad (10)$$

Consequently, (9), we imply $\beta_n(3, 1) \rightarrow \eta$.

Note that for $\lambda = 1$ the normalized CR-statistic assumes the form

$$\tilde{\text{CR}}_1(n) = \frac{\sqrt{2n}}{\sqrt{m_n}} \tilde{\chi}^2(n) + \frac{n^2}{m_n}.$$

Therefore, in view of (2), the statements (8) and (5) are equivalent.

For the log-likelihood ratio statistics $G_2(n)$, inserting $\lambda = 0$ in Theorem 2.2 we get

Corollary 2.3. Assume (3) and

$$n^2 \mathbb{E} P_n \rightarrow \eta. \quad (11)$$

If

$$n \text{Var} \log P_n \rightarrow 0$$

then

$$\frac{G_2(n) + 2n \mathbb{E} \log n P_n}{4 \log 2} \xrightarrow{d} \text{Pois}\left(\frac{\eta}{2}\right).$$

An illustration of this result with the underlying distribution being uniform is given in Fig. 1.

The proof of Theorem 2.2 based on the Poissonian version (see Beška et al., 1982) of the principle of conditioning (see Jakubowski, 1986) is given in Section 4. In this proof we will use also the following decomposition of CR-statistics.

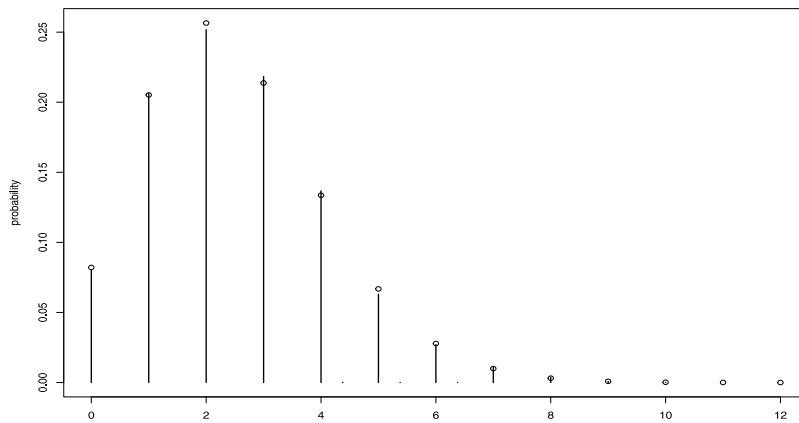


Fig. 1. Empirical probabilities (shown as bars) of $(G_2(n) + 2n\mathbb{E} \log nP_n)/(4 \log 2)$ based on $n = 10,000$ realizations and the uniform distribution on the set $\{1, \dots, n^2/5\}$ vs. the Poisson distribution probabilities (shown as circles) with the parameter $\mu_n = 2.509$ being the average of empirical means. The parameter of the asymptotic Poisson distribution is $\eta/2 = 2.5$.

Proposition 2.4. For any CR-statistic we have

$$CR_\lambda(n) = \frac{U_\lambda(n)}{n^\lambda} + S_\lambda(n) \quad (12)$$

where

$$U_\lambda(n) = \sum_{j \in M_n} \frac{g_\lambda(N_n(j))}{(p_n(j))^\lambda} \quad \text{and} \quad S_\lambda(n) = n \sum_{1 \leq k \leq n} p_n(X_{n,k}) g_\lambda\left(\frac{1}{np_n(X_{n,k})}\right). \quad (13)$$

Proof of Proposition 2.4. Note that g_λ satisfies

$$g_\lambda(ab) = b^{\lambda+1}g_\lambda(a) + ag_\lambda(b), \quad a, b \geq 0. \quad (14)$$

Therefore, taking $a = N_n(j)$ and $b = (np_n(j))^{-1}$ in (14) we see that CR-statistic can be rewritten as

$$CR_\lambda(n) = \frac{1}{n^\lambda} \sum_{j \in M_n} \frac{g_\lambda(N_n(j))}{(p_n(j))^\lambda} + \sum_{j \in M_n} N_n(j) g_\lambda((np_n(j))^{-1}).$$

Observe that

$$\sum_{j \in M_n} N_n(j) f(j) = \sum_{k=1}^n f(X_{n,k})$$

for arbitrary function f . This identity for $f(x) = g_\lambda((np_n(x))^{-1})$ applied to the second sum above yields (12). $\square$

In the following Remark we illustrate (12) for $\lambda = 1$, that is, for the important special case of the chi-square statistic.

Remark 2.2. When $\lambda = 1$ the decomposition (12) reads

$$\chi^2(n) = \frac{U_1(n) + nS_1(n)}{n}, \quad (15)$$

with

$$U_1(n) = \sum_{j \in M_n} \frac{N_n^2(j) - N_n(j)}{p_n(j)}.$$

Note that $N_n^2(j) - N_n(j) = \sum_{1 \leq k \neq \ell \leq n} I_{X_{n,k}=X_{n,\ell}=j}$ and thus, changing the order of summations, we get

$$U_1(n) = \sum_{1 \leq k \neq \ell \leq n} \frac{I_{X_{n,k}=X_{n,\ell}}}{p_n(X_{n,k})}.$$

Also

$$nS_1(n) = \sum_{1 \leq k \leq n} \frac{1}{p_n(X_{n,k})} - n^2.$$

Consequently, (15) agrees with the decomposition of the chi-square statistics from Rempala and Wesolowski (2016).

We note that (see Rempała and Wesółowski, 2016)

$$\mathbb{C}\text{ov}(U_1(n), S_1(n)) = 0 \quad (16)$$

and, consequently, that the variance of the $\chi^2(n)$ statistic is never smaller than that of $U_1(n)/n$ for any finite $n > 1$. In fact, we conjecture that in general, for any $\lambda \geq 0$

$$\mathbb{C}\text{ov}(U_\lambda(n), S_\lambda(n)) \geq 0, \quad (17)$$

and hence that the variance of $CR_\lambda(n)$ is never smaller than that of $U_\lambda(n)/n^\lambda$ for $n > 1$. However, at present we only know (17) to be true for positive integer values of λ . The proof is provided in Appendix. The inequality (17) suggests that to reduce the asymptotic variance of the goodness-of-fit statistics one could replace $CR_\lambda(n)$ with $U_\lambda(n)$.

Remark 2.3. Applying representation (12) to the normalized CR-statistics, as given in (8), after elementary algebra one gets

$$\tilde{CR}_\lambda(n) = \frac{U_\lambda(n) + n^\lambda \bar{S}_\lambda(n)}{\mathbb{E} P_n^{-\lambda}},$$

where $\bar{S}_\lambda(n) = S_\lambda(n) - \mathbb{E} S_\lambda(n)$. Since $\mathbb{V}\text{ar } n^\lambda S_\lambda(n) = n \mathbb{V}\text{ar } P_n^{-\lambda}$, we see that (7) combined with (6) for $r = 2$ implies

$$\lim_{n \rightarrow \infty} \mathbb{V}\text{ar} \frac{n^\lambda \bar{S}_\lambda(n)}{\mathbb{E} P_n^{-\lambda}} \rightarrow 0,$$

whence under assumption of Theorem 2.2

$$\tilde{CR}_\lambda(n) = \tilde{U}_\lambda(n) + o_{\mathbb{P}}(n), \quad \text{with } \tilde{U}_\lambda(n) = \frac{U_\lambda(n)}{\mathbb{E} P_n^{-\lambda}} \quad (18)$$

where $U_\lambda(n)$ is defined in the first part of (13).

Note that the definition of $\tilde{CR}_\lambda(n)$, see (8) (or $\tilde{U}_\lambda(n)$, see (18)), implicitly assumes that $\mathbb{E} P_n^{-\lambda} < \infty$. When $\lambda \geq 1$ this condition implies that M_n is a finite set, since

$$\mathbb{E} P_n^{-\lambda} \geq \mathbb{E} P_n^{-1} = \# M_n.$$

This is in contrast to the case $0 \leq \lambda < 1$ when distributions with infinite support M_n and finite $\mathbb{E} P_n^{-\lambda}$ do exist. Consider, for instance, the following

Example 2.1. Let $X_n \sim \text{geo}(\alpha_n)$, i.e. let X_n follow the geometric distribution. Then $\mathbb{P}(X_n = j) = \alpha_n(1 - \alpha_n)^j$ for $j \geq 0$. In this case $\mathbb{E} P_n^\gamma < \infty$ only if $\gamma > -1$ and

$$\mathbb{E} P_n^\gamma = \frac{\alpha_n^{1+\gamma}}{(1-\alpha_n)^{1+\gamma}}. \quad (19)$$

Consequently, $\tilde{CR}_\lambda(n)$ is well-defined for non-negative λ only when $\mathbb{E} P_n^{-\lambda} < \infty$, that is, only when $\lambda \in [0, 1)$.

Examples of $\tilde{U}_\lambda(n)$ statistics for $\lambda = 1, 0$ are

$$\tilde{U}_1(n) = \frac{1}{m_n} \sum_{j=1}^{m_n} \frac{N_n(j)(N_n(j)-1)}{p_n(j)} \quad \text{and} \quad \tilde{U}_0(n) = 2 \sum_{j \in M_n} N_n(j) \log N_n(j).$$

Example 2.2 (Example 2.1 Continued). As seen above for $X_n \sim \text{geo}(\alpha_n)$ the statistic $\tilde{U}_\lambda(n)$ is only properly defined for $\lambda \in [0, 1)$. By (19), for $\lambda \in (0, 1)$

$$\beta_n(r, \lambda) = n^2 \alpha_n^{2-r} \frac{(1-(1-\alpha_n)^{1-\lambda})^r}{1-(1-\alpha_n)^{2-r\lambda}}, \quad r = 1, 2, 3.$$

In this case (3) implies $n\alpha_n \rightarrow 0$ so that (6) cannot be satisfied for $\lambda > 0$. For $\lambda = 0$ the assumptions (3) and (11) of Corollary 2.3 are satisfied iff $n^2 \alpha_n \rightarrow \eta > 0$, see (19). However,

$$\begin{aligned} \mathbb{V}\text{ar} \log p_{X_n} &= \mathbb{V}\text{ar}(\log \alpha_n(1 - \alpha_n)^{X_n}) = \log^2(1 - \alpha_n) \mathbb{V}\text{ar} X_n \\ &= \frac{1-\alpha_n}{\alpha_n^2} \log^2(1 - \alpha_n) = (1 - \alpha_n) \left[\log(1 - \alpha_n)^{\frac{1}{\alpha_n}} \right]^2 \rightarrow 1, \end{aligned}$$

and we see that the third assumption of Corollary 2.3 is not satisfied.

As we have just seen in Example 2.2, the assumptions of Theorem 2.2 (in particular, of Corollary 2.3) are not satisfied for the geometric probabilities. In the next example we consider a modification of the geometric distribution for which the assumptions of Theorem 2.2 are satisfied (see Fig. 2).

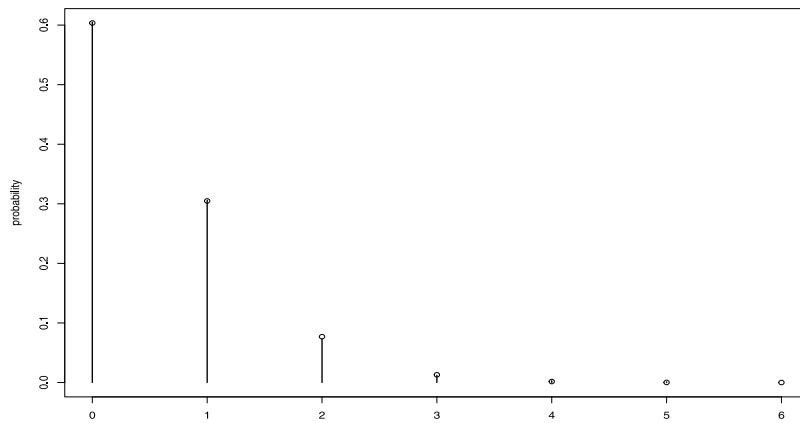


Fig. 2. Empirical probabilities (shown as bars) of $\tilde{U}_{0.5}(n)$ based on $n = 1000$ realizations and the $\text{geo}(1/n^2)$ distribution truncated from above at $n^2 - 1$ (for better comparison, the actual statistic values were slightly rounded to closest integers) vs. the Poisson distribution probabilities (shown as circles) with the parameter $\mu_n = 0.478$ being the average of empirical means. The parameter of asymptotic Poisson distribution is $\eta/2 = 0.5$.

Example 2.3. Let us take $\lambda = 1/2$ and consider $m_n = \#(M_n) < \infty$, $n \geq 1$, such that (2) holds true. Then (6) specializes to

$$n^2 \frac{\mathbb{E} P_n^{1/2}}{\mathbb{E} P_n^{-1/2}} \rightarrow \eta \quad \text{and} \quad \frac{n^2}{(\mathbb{E} P_n^{-1/2})^2} \rightarrow \eta, \quad (20)$$

and (7) becomes

$$\frac{1}{n} \left(m_n - (\mathbb{E} P_n^{-1/2})^2 \right) \rightarrow 0. \quad (21)$$

Thus, under (2) it suffices to assume the first part of (20) and

$$n \left(\frac{m_n}{(\mathbb{E} P_n^{-1/2})^2} - 1 \right) \rightarrow 0 \quad (22)$$

to ensure that the conditions of Theorem 2.2 are satisfied.

We now consider the geometric distribution truncated from above, i.e.

$$p_n(j) = \frac{1-c_n}{1-c_n^{m_n}} c_n^j, \quad j = 0, 1, \dots, m_n - 1,$$

for m_n such that (2) holds. We will show that then both (22) and the first relation of (20) are satisfied for some specific choice of the sequence c_n . Note that, since

$$\mathbb{E} P_n^{-1/2} = \sum_{j=0}^{m_n-1} (p_n(j))^{1/2} = \left(\frac{1-c_n}{1-c_n^{m_n}} \right)^{1/2} \frac{1-c_n^{m_n/2}}{1-c_n^{1/2}}$$

and

$$\mathbb{E} P_n^{1/2} = \sum_{j=0}^{m_n-1} (p_n(j))^{3/2} = \left(\frac{1-c_n}{1-c_n^{m_n}} \right)^{3/2} \frac{1-c_n^{3m_n/2}}{1-c_n^{3/2}},$$

then taking $c_n = 1 - \frac{1}{m_n^2}$ gives after some elementary algebra

$$n \left(\frac{m_n}{(\mathbb{E} P_n^{-1/2})^2} - 1 \right) = n \left(\left(\frac{1 + \left(1 - \frac{1}{m_n^2}\right)^{m_n/2}}{1 + \left(1 - \frac{1}{m_n^2}\right)^{1/2}} \right)^2 \frac{1}{m_n \left(1 - \left(1 - \frac{1}{m_n^2}\right)^{m_n}\right)} - 1 \right). \quad (23)$$

To conclude that (22) holds, we apply the double inequality $1 - mx \leq (1 - x)^m \leq 1 - mx + \frac{m(m-1)}{2}x^2$ true for $0 < x < 1$. (The proof for integer $m \geq 1$ follows by induction.) In particular, we get

$$\frac{1}{m_n \left(1 - \left(1 - \frac{1}{m_n^2}\right)^{m_n}\right)} \rightarrow 1, \quad (24)$$

and hence the whole expression (23) is of order $\frac{n}{m_n} \rightarrow 0$.

Similarly, elementary algebra gives

$$n^2 \frac{\mathbb{E} P_n^{1/2}}{\mathbb{E} P_n^{-1/2}} = \frac{1 + \left(1 - \frac{1}{m_n^2}\right)^{m_n/2} + \left(1 - \frac{1}{m_n^2}\right)^{m_n}}{2 + \left(1 - \frac{1}{m_n^2}\right)^{1/2}} \frac{1}{m_n \left(1 - \left(1 - \frac{1}{m_n^2}\right)^{m_n}\right)} \frac{n^2}{m_n}.$$

Note that the first factor converges to 1 and so does the second one, by (24). Since the third one converges to η by (2), we may therefore conclude that (20) holds.

Finally, we note that

$$np_n^* = n \frac{1 - c_n}{1 - c_n^{m_n}} = \frac{1}{n} \frac{n^2}{m_n} \frac{1}{m_n \left(1 - \left(1 - \frac{1}{m_n^2}\right)^{m_n}\right)},$$

and by (24) and (2) it follows that (3) is also satisfied. Consequently, the assertions of Theorem 2.2 hold. The illustration for the statistic $\tilde{U}_{0.5}(n)$ in this case is presented in Fig. 2.

Another simple family of distributions for which Theorem 2.2 holds true is given in the following example.

Example 2.4. Denote $p_n(1) = \dots = p_n(k_n) = a_n > 0$ and $p_n(k_n + 1), \dots, p_n(n^2) = b_n > 0$. Then $m_n = n^2$ and necessarily $k_n a_n + (n^2 - k_n) b_n = 1$. Moreover, for any real s

$$\mathbb{E} P_n^{-(s+1)} = \sum_{j=1}^{n^2} \frac{1}{(p_n(j))^s} = \frac{k_n}{a_n^s} + \frac{n^2 - k_n}{b_n^s}.$$

Therefore for $r = 1, 2$, and 3

$$\beta_n(r, \lambda) := n^2 \frac{\frac{k_n}{a_n^{r\lambda-2}} + \frac{n^2 - k_n}{b_n^{r\lambda-2}}}{\left(\frac{k_n}{a_n^{\lambda-1}} + \frac{n^2 - k_n}{b_n^{\lambda-1}}\right)^r}.$$

For $a_n = n^\alpha$ and $k_n = n^\beta$ we have $b_n = \frac{1 - n^{\alpha+\beta}}{n^2 - n^\beta}$. Then we need to have $\alpha + \beta < 0$ and $\beta \leq 2$. Moreover, in order to have (3) we need to have $\alpha < -1$.

Then

$$\beta_n(r, \lambda) = n^2 \frac{\frac{n^\beta}{n^{\alpha(r\lambda-2)}} + \frac{(n^2 - n^\beta)^{r\lambda-1}}{(1 - n^{\alpha+\beta})^{r\lambda-2}}}{\left(\frac{n^\beta}{n^{\alpha(\lambda-1)}} + \frac{(n^2 - n^\beta)^\lambda}{(1 - n^{\alpha+\beta})^{\lambda-1}}\right)^r}.$$

Note that when $\beta - \alpha r \lambda + 2\alpha < 2r\lambda - 2$ the leading term in the numerator is $n^{r\lambda}$ and it is also the leading term in the denominator if only $\beta - \alpha \lambda + \alpha < 2\lambda$. Thus if $(2 + \alpha)\lambda > \beta + 2\alpha + 2$ then both inequalities are obviously satisfied and thus (6) holds true with $\eta = 1$.

Note that condition (7) for $\lambda > 0$ assumes now the form

$$n \frac{\text{var } p_{X_n}^{-\lambda}}{(\mathbb{E} p_{X_n}^{-\lambda})^2} = n \left(\frac{\frac{n^\beta}{n^{\alpha(2\lambda-1)}} + \frac{(n^2 - n^\beta)^{2\lambda}}{(1 - n^{\alpha+\beta})^{2\lambda-1}}}{\left(\frac{n^\beta}{n^{\alpha(\lambda-1)}} + \frac{(n^2 - n^\beta)^\lambda}{(1 - n^{\alpha+\beta})^{\lambda-1}}\right)^2} - 1 \right) \rightarrow 0$$

and we see that it holds for certain ranges of λ , α and β values. Consequently, the assumptions of Theorem 2.2 hold, for instance, when

1. $\lambda = 1, \alpha = -3/2, \beta < 1/3$,
2. $\lambda = 2, \alpha = -9/5, \beta < 1/2$,
3. $\lambda = 1/2, \alpha < -7/5, \beta = 1/3$.

3. The uniform CR-statistic

Note that when $(p_n(j))_{1 \leq j \leq m_n}$ is the uniform distribution, that is $p_n(j) = \frac{1}{m_n}$, $1 \leq j \leq m_n$, we have

$$\tilde{CR}_\lambda(n) = \sum_{j=1}^{m_n} g_\lambda(N_n(j)) =: \mathcal{U}_\lambda(n) \quad (25)$$

Note that (4) is trivially satisfied, conditions (6) reduce to (2) and, consequently (3) follows. Therefore we have the following corollary of Theorem 2.2.

Corollary 3.1. Let $p_n(j) = \frac{1}{m_n}$, $1 \leq j \leq m_n$, $n \geq 1$. Assume (2). Then for every $\lambda \geq 0$

$$\mathcal{U}_\lambda(n) \xrightarrow{d} g_\lambda(2) \text{Pois}\left(\frac{\eta}{2}\right). \quad (26)$$

Even though $\mathcal{U}_\lambda(n)$ is defined here as the normalized CR-statistic for the uniform distribution, the statistic $\mathcal{U}_\lambda(n)$ may be also considered for non-uniform distributions $(p_n(j))_{j \in M_n}$, possibly with infinite countable size of M_n . As we show below, this extension is useful for instance in analyzing the power in the problem of hypothesis testing for the uniform distribution.

We first show that under some natural regularity conditions $\mathcal{U}_\lambda(n)$ is asymptotically Poisson, also for distributions other than uniform (and even when the support M_n is not necessarily a finite set).

Theorem 3.2. Assume (3) and (11) hold true for $(p_n(j))_{j \in M_n}$, $n \geq 1$.

Then for $\lambda \geq 0$ the convergence (26) holds true.

Proof. Observe that $g_\lambda(0) = g_\lambda(1) = 0$. Therefore,

$$\mathcal{U}_\lambda(n) = g_\lambda(2) \sum_{j \in M_n} I_{N_n(j)=2} + \sum_{j \in M_n} g_\lambda(N_n(j)) I_{N_n(j) \geq 3}.$$

Since $N_n(j) \sim \text{Bin}(n, p_n(j))$, $j \in M_n$,

$$\mathbb{P}(\exists j \in M_n : N_n(j) \geq 3) \leq \sum_{j \in M_n} \mathbb{P}(N_n(j) \geq 3) \leq \sum_{j \in M_n} \frac{(np_n(j))^3}{3!} \leq \frac{np_n^*}{3!} n^2 \mathbb{E} P_n \rightarrow 0.$$

Therefore,

$$\mathbb{P}\left(\sum_{j \in M_n} g_\lambda(N_n(j)) I_{N_n(j) \geq 3} > \varepsilon\right) \leq \mathbb{P}(\exists j \in M_n : N_n(j) \geq 3) \rightarrow 0.$$

Whence, it suffices to show that

$$K_{n,2} := \sum_{j \in M_n} I_{N_n(j)=2} \xrightarrow{d} \text{Pois}\left(\frac{\eta}{2}\right). \quad (27)$$

Since

$$\mathbb{E} K_{n,2} = \binom{n}{2} \mathbb{E} P_n (1 - P_n)^{n-2}$$

we have

$$(1 - p_n^*)^{n-2} \binom{n}{2} \mathbb{E} P_n \leq \mathbb{E} K_{n,2} \leq \binom{n}{2} \mathbb{E} P_n.$$

By (3)

$$(1 - p_n^*)^{n-2} = \left(1 - \frac{np_n^*}{n}\right)^{n-2} \rightarrow 1$$

and thus (11) implies $\mathbb{E} K_{n,2} \rightarrow \frac{\eta}{2}$. Consequently, convergence in (27) follows from Theorem III.3.1 of Kolchin et al. (1978). $\square$

Below we discuss some of its implications and relevant examples. First, we note that when $\lambda = 0$ or $\lambda = 1$ Theorem 3.2 gives the following.

Corollary 3.3. Assume (3) and (11). Then

$$\frac{\sum_{1 \leq k \leq n} \log N_n(X_{n,k})}{2 \log 2} \xrightarrow{d} \text{Pois}\left(\frac{\eta}{2}\right)$$

and

$$\frac{\sum_{1 \leq k \leq n} N_n(X_{n,k}) - n}{2} \xrightarrow{d} \text{Pois}\left(\frac{\eta}{2}\right).$$

Proof. For any function f we have

$$\sum_{j \in M_n} N_n(j) f(N_n(j)) = \sum_{1 \leq k \leq n} f(N_n(X_{n,k})).$$

Thus taking $f(x) = \log x$ and $f(x) = x$ we obtain

$$\mathcal{U}_0(n) = 2 \sum_{j \in M_n} N_n(j) \log N_n(j) = 2 \sum_{1 \leq k \leq n} \log N_n(X_{n,k})$$

and

$$\mathcal{U}_1(n) + n = \sum_{j \in M_n} (N_n(j))^2 = \sum_{1 \leq k \leq n} N_n(X_{n,k}).$$

Application of [Theorem 3.2](#) concludes the proof. $\square$

The following are simple examples when [Theorem 3.2](#) is applicable.

Example 3.1 (Hypothesis Testing). For $a \in [0, 1/2]$ let $\text{DPD}(a, m_n)$ be the discrete power distribution defined by

$$p_n(j) = \frac{j^{-a}}{C_n(a)}, \quad j \in \{1, \dots, m_n\}, \quad \text{where} \quad C_n(a) = \sum_{j=1}^{m_n} j^{-a} \simeq \frac{m_n^{1-a}}{1-a}.$$

Note that $\text{DPD}(0, m) = \text{U}(\{1, \dots, m\})$ the uniform distribution on $\{1, \dots, m\}$. Since

$$\mathbb{E} P_n = \sum_{j \in M_n} p_n^2(j) = \frac{1}{C_n^2(a)} \sum_{j=1}^{m_n} \frac{1}{j^{2a}} = \frac{C_n(2a)}{C_n^2(a)} \simeq \frac{m_n^{1-2a}}{1-2a} \left(\frac{1-a}{m_n^{1-a}} \right)^2 = \frac{(1-a)^2}{m_n(1-2a)},$$

for $m_n \simeq n^2/\gamma$ we see that [\(11\)](#) holds, as

$$n^2 \mathbb{E} P_n \rightarrow \eta_\gamma(a) = \frac{\gamma(1-a)^2}{1-2a}.$$

Further, noting that $p_n^* = C_n^{-1}$ we see that [\(3\)](#) holds also, since

$$np_n^* = \frac{n}{C_n} \simeq \frac{n(1-a)}{m_n^{1-a}} = (1-a) \frac{n}{\sqrt{m_n}} \frac{1}{m_n^{1/2-a}} \rightarrow 0 \quad \text{if only} \quad a < 1/2.$$

Consequently, by [Theorem 3.2](#) we get

$$\mathcal{U}_\lambda \rightarrow g_\lambda(2) \text{Pois}(\eta_\gamma(a)/2), \quad a \in [0, 1/2].$$

Let μ be the probability distribution given by $p_n(j)$, $j = 1, \dots, m$ and assume that one wants to test the hypothesis $H_0 : \mu = \text{U}(\{1, \dots, m\})$ against the alternative $H_1 : \mu \in \{\text{DPD}(a, m), a \in (0, 1/2)\}$, where the sample from μ is of size $n = \sqrt{\gamma m}$ for $\gamma > 0$. In view of the above discussion the test statistic $\mathcal{U}_\lambda/g_\lambda(2)$ is asymptotically $\text{Pois}(\gamma/2)$ under H_0 and asymptotically $\text{Pois}(\eta_\gamma(a)/2)$, $a \in (0, 1/2)$, under the alternatives.

Recall that for $N_b \sim \text{Pois}(b)$ we have

$$\mathbb{P}(N_b \geq j) = \frac{\Gamma(b, j)}{(j-1)!}, \quad \text{where} \quad \Gamma(b, j) = \int_0^b t^{j-1} e^{-t} dt \quad \text{is the incomplete Gamma function.}$$

The asymptotic critical region at the significance level α for testing H_0 has the form $K = [g_\lambda(2)k_{\gamma/2}, \infty)$, where

$$k_{\gamma/2} = \inf\{k \in \{1, 2, \dots\} : \mathbb{P}(N_{\gamma/2} \geq k) \leq \alpha\}.$$

Then the asymptotic power function of the test is

$$f_\gamma(a) = \mathbb{P}(N_{\eta_\gamma(a)/2} \geq k_{\gamma/2}) = \frac{\Gamma(\eta_\gamma(a)/2, k_{\gamma/2})}{(k_{\gamma/2}-1)!}, \quad a \in [0, 1/2].$$

Take $\alpha = 0.01$ and consider four choices of $\gamma \in \{1, 4, 10, 20\}$. Then $k_{1/2} = 3$, $k_2 = 6$, $k_5 = 11$ and $k_{10} = 18$. The graphs of f_1, f_4, f_{10}, f_{20} are presented in [Fig. 3](#).

The comparison of the Poisson power function for $\alpha \approx 0.05$, ($f_{2.732}$ and $k_{1.366} = 4$) with the empirical power for different values of n is presented in [Fig. 4](#). Since the rate of convergence of the empirical power seems to only marginally depend on the value of λ only a single asymptotic curve was plotted (solid line).

Example 3.2 (Poisson vs Gaussian Asymptotic Power). In this example we compare the Poisson and Gaussian asymptotics of the CR-statistics for testing the hypothesis that the observations are from the uniform distribution (see [Example 3.1](#)). To illustrate the general problem concisely, we focus on the most popular case of $\lambda = 1$, i.e. the chi-square test statistic.

Let $(p_n(j))_{1 \leq j \leq m_n}$ be the uniform distribution, that is $p_n(j) = \frac{1}{m_n}$, $1 \leq j \leq m_n$. Moreover, instead of [\(2\)](#) assume that

$$\frac{n^2}{m_n} = \eta > 0. \tag{28}$$

Since $\mathcal{U}_1(n) = \chi_n^2$, by [Corollary 3.3](#),

$$\chi_n^2 \xrightarrow{d} 2\text{Pois}\left(\frac{\eta}{2}\right). \tag{29}$$

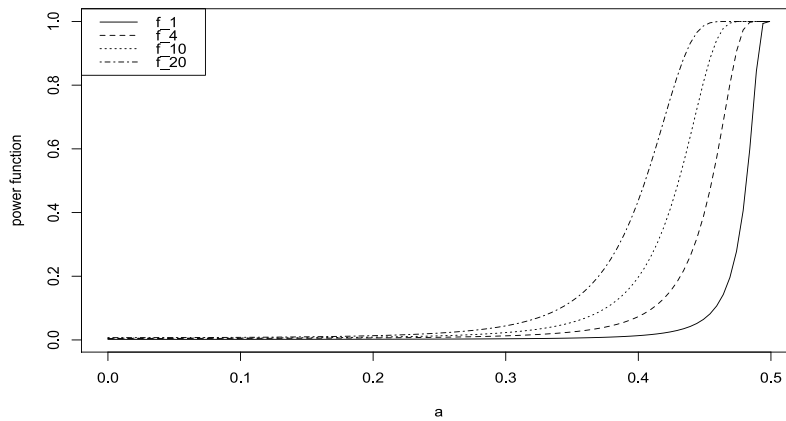


Fig. 3. Test power functions of Example 3.1.

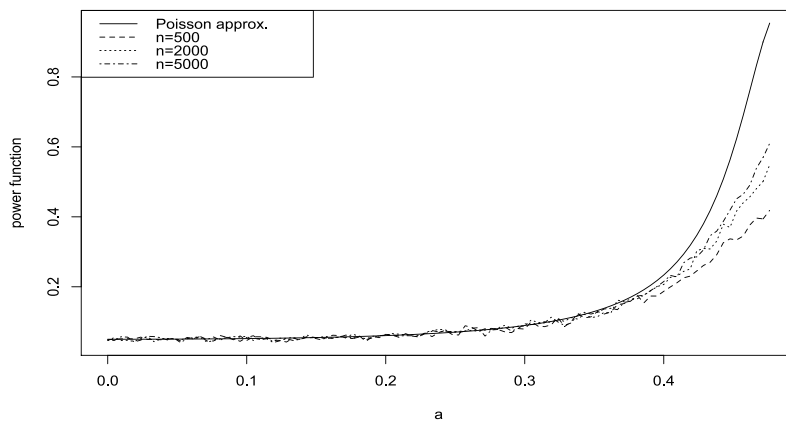


Fig. 4. Empirical power functions of Example 3.1 for different values of n and their Poisson approximation for $\alpha \approx 0.05$. The empirical power value is based on 2000 realizations.

On the other hand it is known (see Cor. 2.2 in Cressie and Read (1984) or Cor. 3.2 in Rempała and Wesolowski (2016)) that in case $n/m_n \rightarrow a > 0$

$$\frac{\chi_n^2 - m_n}{\sqrt{2m_n}} \xrightarrow{d} N(0, 1). \quad (30)$$

In practice it means that for large m_n and n and the value of η in (28) much smaller than m_n one should use the Poisson asymptotics (29). Note that if $\eta \approx am_n$, with relatively small $a > 0$, the correct asymptotics is (30). Let us see what happens when (28) holds true but instead of the Poissonian asymptotics (29) one uses incorrect Gaussian asymptotics (30) to construct the critical region at the given significance level α . Such asymptotic critical region has the form

$$K_\alpha = \left[m_n + z_{1-\alpha} \sqrt{2m_n}, \infty \right).$$

where z_q denotes the q th quantile of the standard normal law. By construction for $Z \sim N(0, 1)$ we have therefore $\mathbb{P}(Z \geq m_n + z_{1-\alpha} \sqrt{2m_n}) \approx \alpha$. However, under the Poissonian asymptotics

$$\alpha(n, m_n) := \mathbb{P}(K_\alpha) = \mathbb{P}\left(N_{\eta/2} \geq \left\lceil \frac{m_n + z_{1-\alpha} \sqrt{2m_n}}{2} \right\rceil\right),$$

where η is defined in (28) and $N_{\eta/2} \sim \text{Pois}(\eta/2)$. For illustration, the plot of $\alpha(n, m_n)$ corresponding to $\alpha = 0.1$ is presented in Fig. 5. As we see from the plot, $\alpha(n, m_n)$ is close to the nominal value $\alpha = 0.1$ only for $\eta \approx m_n$ even for relatively small sample size n . Of course, when $\eta \approx m_n$ and m_n is large we have $2N_{m_n/2} \stackrel{d}{\approx} N(m_n, 2m_n)$. Note that the calculation of the critical region using Gaussian approximation results in a highly conservative test as measured by the (correct) Poisson asymptotics: the graph is well below the level 0.1.

In Fig. 6 we compared the power function of the test based on the Poisson asymptotics with the power function based on the Gaussian ones. We consider, as in Example 3.1, the uniform H_0 and the alternatives from the DPD class. The size of both tests is $\alpha = 0.1$, while measured respectively with respect to the Poisson or Gaussian asymptotics. However in

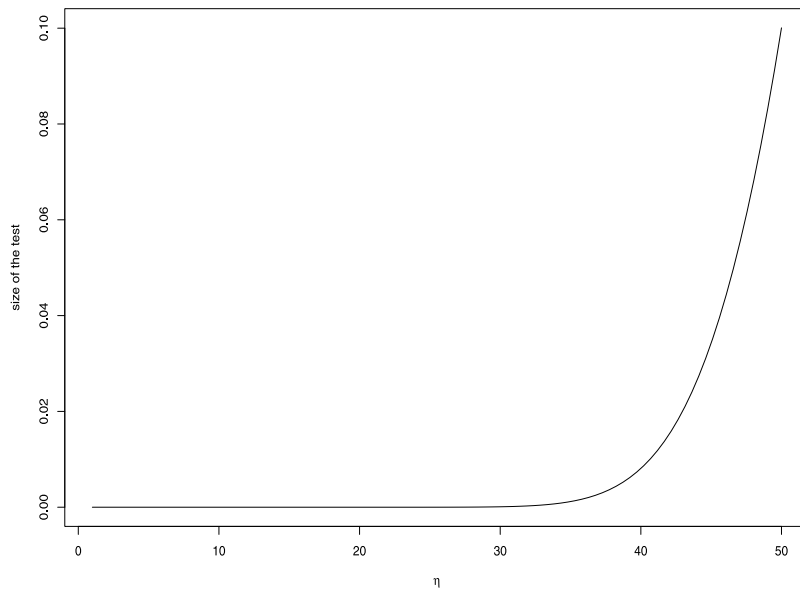


Fig. 5. The size $\alpha(n, m_n)$ of the Gaussian based test of size $\alpha = 0.1$ under Poissonian asymptotics for different values of $\eta = \frac{n^2}{m_n}$ and $m_n = 50$.

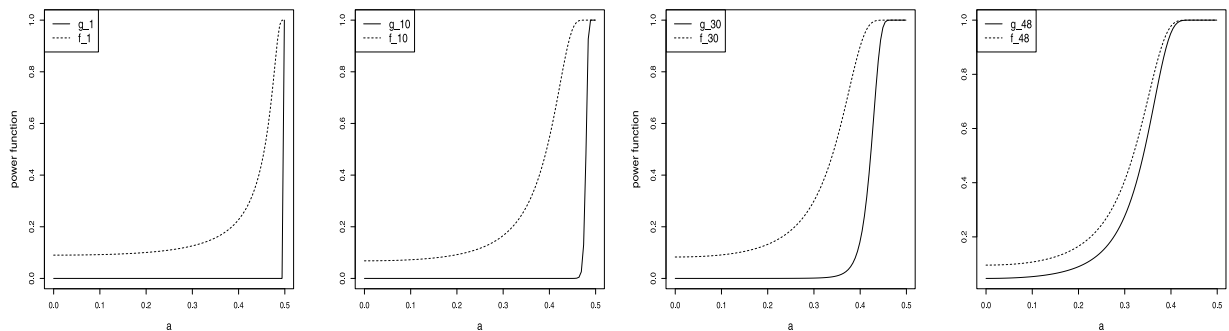


Fig. 6. The power function based on the Poisson asymptotics for the critical region for $\alpha = 0.1$ based on: (1) the Poisson asymptotics (dotted line, f_η) vs. (2) the Gaussian asymptotics (solid line, g_η). Each panel is for different values of $\eta = \frac{n^2}{m_n} \in \{1, 10, 30, 48\}$. The value of $m_n = 50$ is common for all panels.

both cases we use suitable Poisson asymptotics (see again [Example 3.1](#)), to construct power functions. [Fig. 6](#) shows that for η even relatively large ($\eta = 1, 10, 30$) when compared with $m_n = 50$ the power function for the Gaussian based test behaves much worse (it is close to zero on almost all alternatives) than the power function for the Poisson based test. On the other hand, when η approaches m_n ($\eta = 48$) these two power functions become closer and closer and for $\eta = 50$ they are indistinguishable (of course, such values of η yield rather Gaussian than Poisson asymptotics).

Example 3.3 (Continuation of [Example 2.1](#)). For X_n from the geometric distribution $\text{geo}(\alpha_n)$ we have $p_n^* = \alpha_n$ and, by (19),

$$\mathbb{E} P_n = \frac{\alpha_n^2}{1 - (1 - \alpha_n)^2} = \frac{\alpha_n}{2 - \alpha_n}.$$

Therefore (3) and (11) are satisfied under the assumption $n^2\alpha_n \rightarrow 2\eta$ and thus the assertion of [Theorem 3.2](#) follows. This is illustrated in [Fig. 7](#) for $\lambda = 0$ and in [Fig. 8](#) for $\lambda = 2$.

Example 3.4. Let X_n be of the Poisson $\text{Pois}(n^4)$ distribution, $n \geq 1$. Then

$$p_n^* = e^{-n^4} \frac{n^{4n^4}}{(n^4)!}, \quad n \geq 1,$$

and, by the Stirling approximation,

$$np_n^* \simeq \frac{1}{n\sqrt{2\pi}} \rightarrow 0.$$

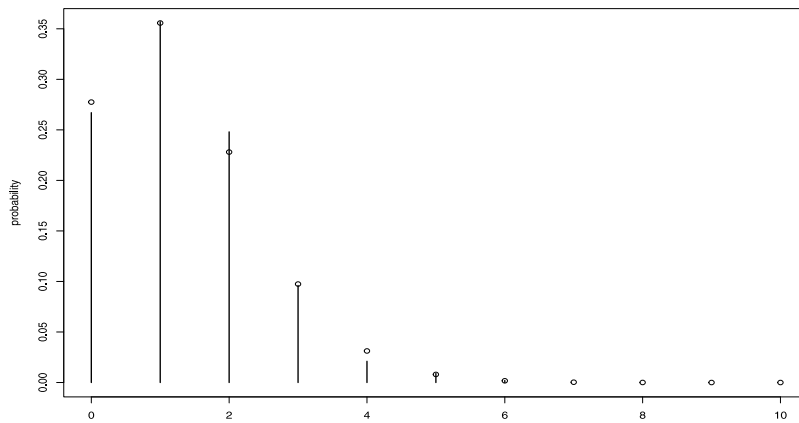


Fig. 7. Empirical probabilities (bars) of $\mathcal{U}_0(n)$ based on 1000 realizations for $n = 10,000$ from the geometric distribution $\text{geo}(5/n^2)$ vs. the probabilities (circles) of Poisson distribution with its parameter $\mu_n = 1.282$ being the average of empirical means. The parameter of asymptotic Poisson distribution is $\eta/2 = 1.25$.

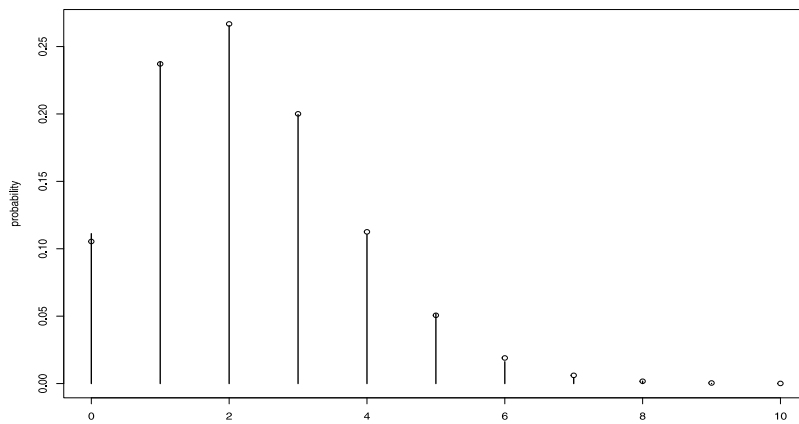


Fig. 8. Empirical probabilities (bars) of $\mathcal{U}_2(n)$ based on 10,000 realizations for $n = 10,000$ from the geometric distribution $\text{geo}(9/n^2)$ vs. the probabilities (circles) of Poisson distribution with its parameter $\mu_n = 2.214$ being the average of empirical means. The parameter of asymptotic Poisson distribution is $\eta/2 = 2.25$.

Moreover,

$$\mathbb{E} P_n = \sum_{j \geq 0} \left(e^{-n^4} \frac{n^{4j}}{j!} \right)^2 = e^{-2n^4} \sum_{j \geq 0} \frac{(n^8)^j}{(j!)^2} = e^{-2n^4} I_0(2n^4), \quad n \geq 1,$$

where I_0 is the modified Bessel function of the first kind. Since $I_0(s) \simeq \frac{e^s}{\sqrt{2\pi s}}$ (see e.g. 9.7.1 in Abramovitz and Stegun, 1964), we have

$$n^2 \mathbb{E} P_n \rightarrow \frac{1}{2\sqrt{\pi}}.$$

Thus, by Theorem 3.2 we conclude that $\frac{\mathcal{U}_\lambda}{g_\lambda(2)} \xrightarrow{d} \text{Pois}\left(\frac{1}{2\sqrt{2\pi}}\right)$.

4. Proofs

4.1. Poisson principle of conditioning

The proof relies on a martingale technique known as *the principle of conditioning* discussed in general in Jakubowski (1986). We apply its Poisson version from Beška et al. (1982) which is cited here for readers' convenience.

Theorem 4.1. Let $(Z_{n,k}, k = 1, \dots, n)_{n \geq 1}$ be a double sequence of non-negative random variables adapted to a row-wise increasing double sequence of σ -fields $(\mathcal{G}_{n,k}, k = 0, 1, \dots, n)_{n \geq 0}$ (with $\mathcal{G}_{n,0}$ being a trivial σ -field, $n \geq 1$). Assume that for

$n \rightarrow \infty$

$$\max_{1 \leq k \leq n} \mathbb{E}(Z_{n,k} | \mathcal{G}_{n,k-1}) \xrightarrow{\mathbb{P}} 0, \quad (31)$$

$$\sum_{k=1}^n \mathbb{E}(Z_{n,k} | \mathcal{G}_{n,k-1}) \xrightarrow{\mathbb{P}} \zeta > 0, \quad (32)$$

and for any $\varepsilon > 0$

$$\sum_{k=1}^n \mathbb{E}(Z_{n,k} I_{|Z_{n,k-1}| > \varepsilon} | \mathcal{G}_{n,k-1}) \xrightarrow{\mathbb{P}} 0. \quad (33)$$

Then

$$\sum_{k=1}^n Z_{n,k} \xrightarrow{d} \text{Pois}(\zeta).$$

4.2. Proof of Theorem 2.2

In view of Remark 2.3 and formula (18) it suffices to focus only on the asymptotic behavior of $\tilde{U}_\lambda(n)$. Note that $g_\lambda(0) = g_\lambda(1) = 0$ and that the function $\tilde{g}_\lambda := g_\lambda/g_\lambda(2)$ is non-decreasing on the set $T = \{0\} \cup [1, \infty)$. Let us introduce an auxiliary function

$$h_\lambda(x) = \tilde{g}_\lambda(x+1) - \tilde{g}_\lambda(x), \quad x \in T \quad (34)$$

and note that $0 \leq h_\lambda(x) \leq Cb^x$ on T with some constants $C > 0$ and $b > 1$ (possibly depending on $\lambda \geq 0$). Moreover, since for $x > 0$

$$h'_\lambda(x) = \frac{2}{g_\lambda(2)} \begin{cases} \frac{(x+1)^\lambda - x^\lambda}{\lambda} > 0, & \lambda > 0, \\ \log(x^{-1} + 1) > 0, & \lambda = 0 \end{cases}$$

the function h_λ is nondecreasing on $[1, \infty)$ for any $\lambda \geq 0$ (actually, it is non-decreasing on T).

Recall that for any $n \geq 1$ the distribution of X_n is given by $\mathbb{P}(X_n = j) = p_n(j)$, $j \in M_n$ and that we define $P_n := p_n(X_n)$. Note that for any function f and a random vector $\mathbf{Y}$, such that $\mathbf{Y}$ and X_n are independent

$$\sum_{j \in M_n} f(\mathbf{Y}, j) = \mathbb{E} \left(\frac{f(\mathbf{Y}, X_n)}{P_n} | \mathbf{Y} \right),$$

where the sum is assumed to be finite. Therefore, (see (18)) we may write

$$\tilde{U}_\lambda(n) = \frac{1}{\mathbb{E} P_n^{-\lambda}} \mathbb{E} \left(\frac{g_\lambda(N_n(X_n))}{P_n^{\lambda+1}} | \mathbf{X}_n \right)$$

where $\mathbf{X}_n = (X_{n,1}, \dots, X_{n,n})$. Denote

$$L_{n,k} = \frac{1}{\mathbb{E} P_n^{-\lambda}} \mathbb{E} \left(\frac{\tilde{g}_\lambda(N_{n,k}(X_n)) - \tilde{g}_\lambda(N_{n,k-1}(X_n))}{P_n^{\lambda+1}} | \mathbf{X}_n \right), \quad k = 1, \dots, n$$

with $N_{n,k}(j) = \sum_{i=1}^k I_{X_{n,i}=j}$, $k = 1, \dots, n$, and $N_{n,0}(j) = 0$. In particular, $N_{n,n} = N_n$, which was considered earlier. Then

$$\tilde{U}_\lambda(n) = g_\lambda(2) \sum_{k=1}^n L_{n,k}.$$

We will prove that $\sum_{k=1}^n L_{n,k}$ converges in law to a Poisson distribution by verifying the conditions of Theorem 4.1.

First, we observe that

$$L_{n,k} = \mathbb{E} \left(G_{n,k}(X_n) \frac{I_{X_{n,k}=X_n}}{P_n} | \mathbf{X}_n \right), \quad (35)$$

where $G_{n,k}(j) = \frac{h_\lambda(N_{n,k-1}(j))}{(p_n(j))^\lambda} \frac{1}{\mathbb{E} P_n^{-\lambda}}$ and

$$\mathbb{E}(L_{n,k} | \mathbf{X}_{n,k-1}) = \mathbb{E}(G_{n,k}(X_n) | \mathbf{X}_n). \quad (36)$$

Second, we observe that

$$\mathbb{E} L_{n,k}^r = \mathbb{E} G_{n,k}^r(X_n). \quad (37)$$

Indeed, let $Z_{n,1}, \dots, Z_{n,r}, X_{n,1}, \dots, X_{n,n}, X_n$ be iid rv's. Then due to (35) we can write

$$\mathbb{E} L_{n,k}^r = \mathbb{E} \prod_{i=1}^r \mathbb{E} \left(G_{n,k}(Z_{n,i})^{\frac{I_{X_{n,k}=Z_{n,i}}}{p_n(Z_{n,i})}} | \mathbf{X}_n \right)$$

and by the conditional independence of $G_{n,k}(Z_{n,i})^{\frac{I_{X_{n,k}=Z_{n,i}}}{p_n(Z_{n,i})}}, i = 1, \dots, r$, given $\mathbf{X}_n$

$$\mathbb{E} L_{n,k}^r = \mathbb{E} \prod_{i=1}^r G_{n,k}^r(Z_{n,i})^{\frac{I_{X_{n,k}=Z_{n,i}}}{p_n(Z_{n,i})}} = \mathbb{E} \frac{G_{n,k}^r(Z_{n,1})}{(p_n(Z_{n,1}))^r} I_{Z_{n,1}=\dots=Z_{n,r}=X_{n,k}}$$

and thus (37) follows due to $\mathbb{E}(I_{Z_{n,1}=\dots=Z_{n,r}=X_{n,k}} | \mathbf{X}_{n,k-1}, Z_{n,1}) = (p_n(Z_{n,1}))^r$.

Third, we observe that

$$\sum_{k=1}^n \mathbb{E} L_{n,k}^r \rightarrow \frac{\eta}{2}, \quad r = 1, 2, 3. \quad (38)$$

Indeed, since the conditional distribution of $N_{n,k-1}(X_n)$ given $\mathbf{X}_n$ is binomial $\text{Bin}(k-1, P_n)$, we have

$$\begin{aligned} \mathbb{E}(h_{\lambda}^r(N_{n,k-1}(X_n)) | \mathbf{X}_n) &= \sum_{k=1}^n \sum_{i=1}^{k-1} h_{\lambda}^r(i) \binom{k-1}{i} P_n^i (1-P_n)^{k-1-i} = \sum_{i=1}^{n-1} h_{\lambda}^r(i) P_n^i \sum_{k=i+1}^n \binom{k-1}{i} (1-P_n)^{k-1-i} \\ &= P_n \sum_{k=2}^n (k-1)(1-P_n)^{k-2} + \sum_{i=2}^{n-1} h_{\lambda}^r(i) P_n^i \sum_{k=i+1}^n \binom{k-1}{i} (1-P_n)^{k-1-i}. \end{aligned}$$

Consequently, (37) together with the definition of $G_{n,k}$, yield

$$\sum_{k=1}^n \mathbb{E} L_{n,k}^r = \mathcal{I}_1(n, r) + \mathcal{I}_2(n, r),$$

where

$$\mathcal{I}_1(n, r) = \frac{\mathbb{E} P_n^{1-r\lambda} \sum_{k=2}^n (k-1)(1-P_n)^{k-2}}{(\mathbb{E} P_n^{-\lambda})^r}$$

and

$$\mathcal{I}_2(n, r) = \frac{\mathbb{E} P_n^{-r\lambda} \sum_{i=2}^{n-1} h_{\lambda}^r(i) P_n^i \sum_{k=i+1}^n \binom{k-1}{i} (1-P_n)^{k-1-i}}{(\mathbb{E} P_n^{-\lambda})^r}.$$

Now, $\mathcal{I}_1(n, r) \rightarrow \frac{\eta}{2}$ for $r = 1, 2, 3$ which follows from (6) in view of the double inequality

$$(1 - p_n^*)^{\frac{n-1}{2n}} \beta_n(r, \lambda) \leq \mathcal{I}_1(n, r) \leq \frac{n-1}{2n} \beta_n(r, \lambda).$$

On the other hand, since

$$\sum_{i=2}^{n-1} h_{\lambda}^r(i) P_n^i \sum_{k=i+1}^n \binom{k-1}{i} (1-P_n)^{k-1-i} \leq C P_n n^2 \sum_{i=2}^{n-1} (b^r p_n^*)^{i-1}$$

for sufficiently large n , i.e. such that $b^r n p_n^* \in (0, 1)$, we get

$$0 \leq \mathcal{I}_2(n, r) \leq C \beta_n(r, \lambda) \sum_{i=1}^{\infty} (b^r n p_n^*)^i = C \beta_n(r, \lambda) \frac{b^r n p_n^*}{1 - b^r n p_n^*} \rightarrow 0.$$

Thus (38) is established.

Now we are ready to verify conditions (31), (32) and (33).

Ad (31). Since h_{λ} is non-decreasing we get

$$\max_{1 \leq k \leq n} \mathbb{E}(L_{n,k} | \mathbf{X}_{n,k-1}) = \mathbb{E}(G_{n,n}(X_n) | \mathbf{X}_n).$$

Note that

$$\mathbb{E} \frac{h_{\lambda}(N_{n,n-1}(X_n))}{P_n^{\lambda}} = \mathbb{E} \frac{\mathbb{E}(h_{\lambda}(N_{n,n-1}(X_n)) | \mathbf{X}_n)}{P_n^{\lambda}} = \mathbb{E} \frac{\sum_{i=1}^{n-1} h_{\lambda}(i) \binom{n-1}{i} P_n^i (1-P_n)^{n-1-i}}{P_n^{\lambda}} \leq C \frac{b n p_n^*}{1 - b n p_n^*} \mathbb{E} P_n^{-\lambda}$$

for n sufficiently large. Consequently,

$$0 \leq \mathbb{E} \max_{1 \leq k \leq n} \mathbb{E}(L_{n,k} | \mathbf{X}_{n,k-1}) \leq \frac{1}{\mathbb{E} P_n^{-\lambda}} C \frac{bnp_n^*}{1-bnp_n^*} \mathbb{E} P_n^{-\lambda} = C \frac{bnp_n^*}{1-bnp_n^*} \rightarrow 0$$

and (31) follows.

Ad (32). Due to (38) for $r = 1$, in order for (32) to hold it suffices to show that

$$\text{Var} \sum_{k=1}^n \mathbb{E}(L_{n,k} | \mathbf{X}_{n,k-1}) \rightarrow 0.$$

Consider $Y_n \stackrel{d}{=} X_n$ such that $Y_n, X_n, \mathbf{X}_n$ are independent. Then by (36) it follows that for any $1 \leq k, \ell \leq n$

$$\mathbb{E}(L_{n,k} | \mathbf{X}_{n,k-1}) \mathbb{E}(L_{n,\ell} | \mathbf{X}_{n,\ell-1}) = \mathbb{E}(G_{n,k}(X_n) | \mathbf{X}_n) \mathbb{E}(G_{n,\ell}(Y_n) | \mathbf{X}_n),$$

Consequently,

$$\begin{aligned} \text{Var} \sum_{k=1}^n \mathbb{E}(L_{n,k} | \mathbf{X}_{n,k-1}) &= \sum_{1 \leq k, \ell \leq n} \{ \mathbb{E}[\mathbb{E}(G_{n,k}(X_n) | \mathbf{X}_n) \mathbb{E}(G_{n,\ell}(Y_n) | \mathbf{X}_n)] - \mathbb{E} G_{n,k}(X_n) \mathbb{E} G_{n,\ell}(Y_n) \}. \end{aligned}$$

Thus, conditional independence of $G_{n,k}(X_n)$ and $G_{n,\ell}(Y_n)$ given $\mathbf{X}_n$, yields

$$\text{Var} \sum_{k=1}^n \mathbb{E}(L_{n,k} | \mathbf{X}_{n,k-1}) = \sum_{k, \ell=1}^n \text{Cov}(G_{n,k}(X_n), G_{n,\ell}(Y_n)). \quad (39)$$

For $k = \ell$ we have

$$\begin{aligned} \text{Cov}(G_{n,k}(X_n), G_{n,k}(Y_n)) &= \mathbb{E} P_n \text{Var}(G_{n,k}(X_n) | X_n) + \mathbb{E} I_{X_n \neq Y_n} \text{Cov}(G_{n,k}(X_n), G_{n,k}(Y_n) | X_n, Y_n). \end{aligned} \quad (40)$$

By definition of $G_{n,k}(X_n)$ we can write

$$\begin{aligned} \mathbb{E} P_n \text{Var}(G_{n,k}(X_n) | X_n) &\leq \mathbb{E} P_n \mathbb{E}(G_{n,k}^2(X_n) | X_n) \\ &= \frac{1}{(\mathbb{E} P_n^{-\lambda})^2} \mathbb{E} \frac{\mathbb{E}(h_\lambda^2(N_{n,k-1}(X_n)) | X_n)}{P_n^{2\lambda-1}} \leq \mathbb{E} \frac{C \sum_{i=1}^{k-1} b^{2i} \binom{k-1}{i} P_n^i}{P_n^{2\lambda-1} (\mathbb{E} P_n^{-\lambda})^2}. \end{aligned}$$

Therefore, changing the order of summation, we get

$$\begin{aligned} \sum_{k=1}^n \mathbb{E} P_n \text{Var}(G_{n,k}(X_n) | X_n) &\leq \frac{C \sum_{i=1}^{n-1} b^{2i} \mathbb{E} P_n^{i+1-2\lambda} n^{i+1}}{(\mathbb{E} P_n^{-\lambda})^2} \\ &\leq \frac{C \beta_n(2, \lambda)}{n} \sum_{i=1}^{n-1} (b^2 np_n^*)^i \leq \frac{C \beta_n(2, \lambda)}{n} \frac{b^2 np_n^*}{1-b^2 np_n^*} \rightarrow 0. \end{aligned}$$

Now, consider the second part of (40). Note that on the set $\{X_n \neq Y_n\}$ we have

$$\begin{aligned} \text{Cov}(G_{n,k}(X_n), G_{n,k}(Y_n) | X_n, Y_n) &= \frac{1}{(\mathbb{E} P_n^{-\lambda})^2} \frac{1}{P_n P'_n} \text{Cov}(h_\lambda(N_{n,k-1}(X_n)), h_\lambda(N_{n,k-1}(Y_n)) | X_n, Y_n), \end{aligned}$$

where $P'_n = p_n(Y_n)$, i.e. $P'_n \stackrel{d}{=} P_n$ and P_n and P'_n are independent.

Since the vector $(N_{n,k-1}(i), N_{n,k-1}(j))$ for $i \neq j$ has a multinomial distribution, it follows that its two components are negatively associated (NA). As any non-decreasing (integrable) functions of NA random variables are negatively correlated¹ the above right-hand side is non-positive, and it thus follows that the second summand in (40) is also non-positive.

For $k \neq \ell$ we have

$$\text{Cov}(G_{n,k}(X_n), G_{n,\ell}(Y_n)) = \mathbb{E} P_n \text{Cov}(G_{n,k}(X_n), G_{n,\ell}(X_n) | X_n) + \mathbb{E} I_{X_n \neq Y_n} \text{Cov}(G_{n,k}(X_n), G_{n,\ell}(Y_n) | X_n, Y_n). \quad (41)$$

We conclude that the second term is non-positive due to the NA property described above. For $k < \ell$ in view of the Schwartz inequality and monotonicity of h_λ the first term can be bounded as follows

$$\mathbb{E} P_n \text{Cov}(G_{n,k}(X_n), G_{n,\ell}(X_n) | X_n) \leq \mathbb{E} P_n \mathbb{E}(G_{n,\ell}^2(X_n) | X_n).$$

¹ See, e.g., Joag-Dev and Proschan (1983) for some relevant facts of the theory of NA random variables.

Consequently,

$$2 \sum_{1 \leq k < \ell \leq n} \mathbb{E} P_n \text{Cov}(G_{n,k}(X_n), G_{n,\ell}(X_n) | X_n) \leq 2 \sum_{\ell=2}^n (\ell-1) \mathbb{E} P_n \mathbb{E}(G_{n,\ell}^2(X_n) | X_n)$$

Thus, similarly as before we get

$$2 \sum_{1 \leq k < \ell \leq n} \mathbb{E} P_n \text{Cov}(G_{n,k}(X_n), G_{n,\ell}(X_n) | X_n) \leq \frac{C \sum_{i=1}^{n-1} b^{2i} \mathbb{E} P_n^{i+1-2\lambda} n^{i+2}}{(\mathbb{E} P_n^{-\lambda})^2} \leq C \beta_n(2, \lambda) \frac{b^2 n p_n^*}{1-b^2 n p_n^*} \rightarrow 0.$$

Collecting the above results

$$\begin{aligned} & \text{Var} \sum_{k=1}^n \mathbb{E}(L_{n,k} | \mathbf{X}_{n,k-1}) \\ & \leq \sum_{k=1}^n \mathbb{E} P_n \text{Cov}(G_{n,k}(X_n), G_{n,k}(Y_n) | X_n) + \sum_{1 \leq k \neq \ell \leq n} \mathbb{E} P_n \text{Cov}(G_{n,k}(X_n), G_{n,\ell}(Y_n) | X_n, Y_n) \rightarrow 0, \end{aligned}$$

and thus condition (32) is verified.

Ad (33). Due to (38) we have

$$\begin{aligned} \mathbb{E} \sum_{k=1}^n \mathbb{E}(L_{n,k} I_{|L_{n,k}-1| > \varepsilon} | \mathbf{X}_{n,k-1}) &= \sum_{k=1}^n \mathbb{E} L_{n,k} I_{|L_{n,k}-1| > \varepsilon} \\ &\leq \varepsilon^{-2} \left(\sum_{k=1}^n \mathbb{E} L_{n,k} - 2 \sum_{k=1}^n \mathbb{E} L_{n,k}^2 + \sum_{k=1}^n \mathbb{E} L_{n,k}^3 \right) \rightarrow 0 \end{aligned}$$

and (33) also holds.

By Theorem 4.1 we conclude that $\sum_{k=1}^n L_{n,k}$ converges in law to the required Poisson distribution. $\square$

Acknowledgments

This research has been partially supported by the National Science Center, Poland, Grant 2016/21/B/ST1/00005 and US National Science Foundation under grant DMS 1853587. This research was initiated during second author's visit to Mathematical Biosciences Institute (MBI) at The Ohio State University during Spring 2018. MBI receives major funding from US National Science Foundation under grant DMS 1440386.

Appendix. The covariance conjecture

We consider here the conjecture (17) from the main body of the paper. First, we prove the equivalent representation of $\text{Cov}(U_\lambda(n), S_\lambda(n))$ valid for any $\lambda \geq 0$. With the help of this representation, we show that (17) holds for positive integer values of λ . Unfortunately, we are presently unable to extend this result to arbitrary nonnegative λ .

A.1. Covariance representation

Lemma A.1. For any $\lambda \geq 0$

$$\text{Cov}(U_\lambda(n), S_\lambda(n)) = \mathbb{E} \left\{ P_n^{-\lambda} (H_\lambda(P_n^{-1}) - \mathbb{E} H_\lambda(P_n^{-1})) (g_\lambda(B_{n-1} + 1) - g_\lambda(B_{n-1})) \right\}, \quad (42)$$

where $H_\lambda(x) = \frac{g_\lambda(x)}{x}$, $x > 0$, and the conditional distribution $\mathbb{P}_{B_{n-1}|X_n} \sim \text{Bin}(n-1, P_n)$.

Proof. By symmetry and bi-linearity of the covariance operator we get

$$\begin{aligned} & \text{Cov}(U_\lambda(n), S_\lambda(n)) \\ &= n^{1-\lambda} \sum_{j \in M_n} \text{Cov}(p_n^{-\lambda}(j) g_\lambda(N_n(j)), H_\lambda(p_n^{-1}(X_{n,n})) \\ &= n^{1-\lambda} \left(\sum_{j \in M_n} \mathbb{E} p_n^{-\lambda}(j) g_\lambda(N_n(j)) H_\lambda(p_n^{-1}(X_{n,n})) - \mathbb{E} H_\lambda(p_n^{-1}(X_{n,n})) \sum_{j \in M_n} \mathbb{E} p_n^{-\lambda}(j) g_\lambda(N_n(j)) \right) \\ &= n^{1-\lambda} (\mathbb{E} P_n^{-1-\lambda} g_\lambda(N_n(X_n)) H_\lambda(p_n^{-1}(X_{n,n})) - \mathbb{E} H_\lambda(p_n^{-1}(X_{n,n})) \mathbb{E} P_n^{-1-\lambda} g_\lambda(X_n)) \\ &= n^{1-\lambda} \text{Cov}(P_n^{-1-\lambda} g_\lambda(N_n(X_n)), H_\lambda(p_n^{-1}(X_{n,n}))) \\ &= n^{1-\lambda} \mathbb{E} \left\{ P_n^{-1-\lambda} \text{Cov}(g_\lambda(N_n(X_n)), H_\lambda(p_n^{-1}(X_{n,n})) | X_n) \right\}, \end{aligned}$$

where the last equality follows from the fact that $H_\lambda(p_n^{-1}(X_{n,n}))$ and X_n are independent.

But

$$\begin{aligned} & \mathbb{E} (g_\lambda(N_n(X_n)) H_\lambda(p_n^{-1}(X_{n,n})) | X_n) \\ &= \mathbb{E} \left([g_\lambda(N_{n-1}(X_n) + 1) I_{X_{n,n}=X_n} + g_\lambda(N_{n-1}(X_n)) I_{X_{n,n} \neq X_n}] H_\lambda(p_n^{-1}(X_{n,n})) | X_n \right) \\ &= \mathbb{E}(g_\lambda(N_{n-1}(X_n) + 1) | X_n) P_n H_\lambda(p_n^{-1}) + \mathbb{E}(g_\lambda(N_{n-1}(X_n)) | X_n) (\mathbb{E} H_\lambda(p_n^{-1}) - P_n H_\lambda(p_n^{-1})). \end{aligned}$$

Since

$$\mathbb{E}(g_\lambda(N_n(X_n)) | X_n) = \mathbb{E}(g_\lambda(N_{n-1}(X_n) + 1) | X_n) P_n + \mathbb{E}(g_\lambda(N_{n-1}(X_n)) | X_n) (1 - P_n),$$

we get

$$\begin{aligned} & \text{Cov} (g_\lambda(N_n(X_n)), H_\lambda(p_n^{-1}(X_{n,n})) | X_n) \\ &= [\mathbb{E}(g_\lambda(N_{n-1}(X_n) + 1) | X_n) - \mathbb{E}(g_\lambda(N_{n-1}(X_n)) | X_n)] P_n H_\lambda(p_n^{-1}) \\ &\quad - [\mathbb{E}(g_\lambda(N_{n-1}(X_n) + 1) | X_n) - \mathbb{E}(g_\lambda(N_{n-1}(X_n)) | X_n)] P_n \mathbb{E} H_\lambda(p_n^{-1}), \end{aligned}$$

and thus (42) follows. $\square$

A.2. Case of the positive integer λ

Proposition A.2. Assume that $\lambda \in \{1, 2, \dots\}$. Then $\text{Cov}(U_\lambda(n), S_\lambda(n)) \geq 0$.

Proof. Note that in view of the representation (42) the inequality (17) for any $\lambda > 0$ is equivalent to

$$\mathbb{E} P_n^{-\lambda} (P_n^{-\lambda} - \mathbb{E} P_n^{-\lambda}) [(B_{n-1} + 1)^{\lambda+1} - B_{n-1}^{\lambda+1} - 1] \geq 0. \quad (43)$$

Recall that for positive integer K

$$x^K = \sum_{i=1}^K S(K, i)(x)_i,$$

where $S(K, i)$, $i = 1, \dots, K$, are Stirling numbers of the second kind and $(x)_i = x(x-1)\dots(x-i+1)$, $i = 1, 2, \dots$, are descending Pochhammer symbols. We note that

$$(x+1)_i - (x)_i = i(x)_{i-1}$$

and thus

$$(x+1)^K - x^K - 1 = \sum_{i=1}^K S(K, i)((x+1)_i - (x)_i) - 1 = \sum_{i=2}^K iS(K, i)(x)_{i-1}.$$

The left hand side of (43) for $K := \lambda + 1$ takes on the form

$$\sum_{i=1}^K iS(K, i) \mathbb{E} \{P_n^{-\lambda} (P_n^{-\lambda} - \mathbb{E} P_n^{-\lambda}) \mathbb{E} ((B_{n-1})_{i-1} | X_n)\}. \quad (44)$$

Since

$$\mathbb{E} [(B_{n-1})_{i-1} | X_n] = (n-1)_{i-1} P_n^{i-1}$$

we get

$$\mathbb{E} \{P_n^{-\lambda} (P_n^{-\lambda} - \mathbb{E} P_n^{-\lambda}) \mathbb{E} ((B_{n-1})_{i-1} | X_n)\} = (n-1)_{i-1} (\mathbb{E} P_n^{-(2K-i-1)} - \mathbb{E} P_n^{-(K-i)} \mathbb{E} P_n^{-(K-1)})$$

Observe that for non-negative random variable Y and $r, s \geq 0$, such that $\mathbb{E} Y^{r+s} < \infty$ we have

$$\mathbb{E} Y^{r+s} \geq \mathbb{E} Y^r \mathbb{E} Y^s. \quad (45)$$

Since each of $2K - i - 1$, $K - i$, $K - 1$ is a nonnegative number, by (45) applied to $Y = P_n^{-1}$ we conclude that (44) is non-negative. $\square$

References

- Berrett, Thomas B, Samworth, Richard J, 2021. USP: an independence test that improves on Pearson's chi-squared and the G-test. arXiv preprint [arXiv:2101.10880](https://arxiv.org/abs/2101.10880).
- Beška, Marek, Kłopotowski, Andrzej, Słomiński, Leszek, 1982. Limit theorems for random sums of dependent d-dimensional random vectors. *Probab. Theory Related Fields* 61 (1), 43–57.

- Browning, Lauren M., Pietrzak, Maciej, Kuczma, Michal, Simms, Colin P., Kurczewska, Agnieszka, Refugia, Justin M., Lowery, Dustin J., Rempala, Grzegorz, Gutkin, Dmitriy, Ignatowicz, Leszek, 2018. TGF- β -mediated enhancement of TH17 cell generation is inhibited by bone morphogenetic protein receptor 1 α signaling. *Sci. Signal.* 11 (545).
- Chen, Son Xi, Qin, Ying-Li, 2010. A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Stat.* 38 (2), 808–835.
- Cressie, Noel, Read, Timothy R.C., 1984. Multinomial goodness-of-fit tests. *J. R. Stat. Soc. Ser. B (Methodological)* 1, 440–464.
- Fienberg, Stephen E., Holland, Paul W., 1973. Simultaneous estimation of multinomial cell probabilities. *J. Am. Stat. Assoc.* 68 (343), 683–691.
- Haberman, Shelby J., 1988. A warning on the use of chi-squared statistics with frequency tables with small expected cell counts. *J. Am. Stat. Assoc.* 83 (402), 555–560.
- Inglot, Tadeusz, Jurlewicz, Teresa, Ledwina, Teresa, 1991. Asymptotics for multinomial goodness of fit tests for a simple hypothesis. *Theory Probab. Appl.* 35 (4), 771–777.
- Jakubowski, Adam, 1986. Principle of conditioning in limit theorems for sums of random variables. *Ann. Probab.* 5, 902–915.
- Janková, Jana, Shah, Rajen D., Bühlmann, Peter, Samworth, Richard J., 2020. Goodness-of-fit testing in high dimensional generalized linear models. *J. R. Stat. Soc. Ser. B (Statistical Methodology)* 82 (3), 773–795.
- Joag-Dev, Kumar, Proschan, Frank, 1983. Negative association of random variables with applications. *Ann. Stat.* 1, 286–295.
- Kim, Ilmun, 2020. Multinomial goodness-of-fit based on U-statistics: High-dimensional asymptotic and minimax optimality. *J. Statist. Plann. Inference* 205, 74–91.
- Kim, Ilmun, Balakrishnan, Sivaraman, Wasserman, Larry, 2020. Robust multivariate nonparametric tests via projection averaging. *Ann. Stat.* 48 (6), 3417–3441.
- Menéndez, Mari Luisa, Morales, Domingo, Pardo, Leandro, Vajda, Igor, 1998. Asymptotic distributions of φ -divergences of hypothetical and observed frequencies on refined partitions. *Stat. Neerlandica* 52 (1), 71–89.
- Mirakhmedov, Sherzod M., 2016. Asymptotic intermediate efficiency of the chi-square and likelihood ratio goodness of fit tests. *arXiv preprint arXiv:1610.04135*.
- Morris, Carl, 1975. Central limit theorems for multinomial sums. *Ann. Statist.* 165–188.
- Ogata, Hiroaki, Taniguchi, Masanobu, 2009. Cressie-read power-divergence statistics for non-gaussian vector stationary processes. *Scandinavian J. Stat.* 36 (1), 141–156.
- Pérez, T., Pardo, J.A., 2002. Asymptotic normality for the K_ϕ -divergence goodness-of-fit tests. *J. Comput. Appl. Math.* 145, 301–317.
- Pietrzak, Maciej, Rempala, Grzegorz A., 2017. Asymptotic approaches to discovering cancer genomic signatures. In: *Handbook of Statistics*, vol. 37. Elsevier, pp. 23–36.
- Pietrzak, Maciej, Rempala, Grzegorz A., Seweryn, Michał, Wesołowski, Jacek, 2016. Limit theorems for empirical Rényi entropy and divergence with applications to molecular diversity analysis. *TEST* 4, 1–20.
- Pour, Ali Foroughi, Pietrzak, Maciej, Dalton, Lori A., Rempala, Grzegorz A., 2020. High dimensional model representation of log-likelihood ratio: binary classification with expression data. *BMC Bioinf.* 21, 1–27.
- Rempala, Grzegorz A., Wesołowski, Jacek, 2016. Double asymptotics for the chi-square statistic. *Statistics and Probability Letters* 119, 317–325.
- Steck, George P., 1957. Limit theorems for conditional distributions. *Univ. California Publ. Statist.* 2 (12), 237–284.
- Tumanyan, S. Kh., 1956. Asymptotic distribution of the chi-square criterion when the number of observations and number of groups increase simultaneously. *Teor. Veroyat. Yeyo Primen.* 1 (1), 131–145.