Journal of Statistical Planning and Inference 242 (2026) 106350

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Assessing goodness-of-fit for sparse categories using Rényi divergence

Raul Matsushita [a,b,*], Gabriel Gomes [a,c], Regina Da Fonseca [d], Eduardo Nakano [a], Roberto Vila [a]

[a] University of Brasilia, Graduate Program in Statistics, Brasília, 70910-900, DF, Brazil
[b] University of Brasilia, Graduate Program in Management, Brasília, 70910-900, DF, Brazil
[c] Federal Police of Brazil Identification Unit of Federal Police Station, Foz do Iguaçu, 85863-720, PR, Brazil
[d] Federal Institute of Goias, Department of Mathematics, Goiânia, 74055-110, GO, Brazil

ARTICLE INFO

Keywords:
Rényi divergence
Small counts
Asymptotic distribution
Test of goodness-of-fit
Chi-square test
Sparse frequency table

ABSTRACT

We present the Rényi divergence as a statistic for assessing goodness-of-fit in sparse frequency tables, where small expected counts can undermine the reliability of the traditional chi-square test. The Rényi divergence with index in (0,1) is a natural choice because it circumvents division-related issues by small frequencies. Our main result demonstrates that the Rényi statistic asymptotically follows a chi-square distribution. Through theoretical insights and Monte Carlo simulations, we evaluate the performance of the Rényi statistic across various values of the divergence index. We find that smaller index values improve the alignment of the Rényi statistic with the chi-square distribution and enhance its performance in sparse data settings. Additionally, the Rényi statistic exhibits good power properties in detecting deviations from the null hypothesis under these conditions. To illustrate its practical applicability, we present two real-world data analyses, highlighting the robustness of the Rényi divergence in scenarios involving sparse categories.

1. Introduction

The chi-square (χ^2) test is a widely used statistical method for comparing observed and expected frequencies in categorical or discrete datasets. It is closely related to the generalized likelihood ratio (GLR) test, which possesses desirable asymptotic properties: under suitable conditions, the GLR statistic follows a χ^2 distribution as the sample size increases. This large-sample property makes the χ^2 test a convenient and straightforward tool in statistical analysis.

However, the validity of the χ^2 test may be compromised in sparse frequency tables, where expected counts are small and asymptotic approximations no longer hold (Maiste and Weir, 2004). This limitation has been well-documented in the literature (Haberman, 1988; Baglivo et al., 1992; Maiste and Weir, 2004). Although collapsing categories or using resampling techniques such as Monte Carlo or bootstrap can help mitigate the problem (Gautam and Kimeldorf, 1999; Mudholkar and Hutson, 1997; Vexler et al., 2014), these solutions either involve a loss of information or add computational overhead. As such, the development of alternative test statistics that remain robust under sparsity remains an important line of research.

Divergence-based measures have long been proposed as alternatives to the χ^2 test for goodness-of-fit testing. In addition to the Kullback–Leibler (KL) divergence (Yang and Chen, 2019), which underpins the GLR test (Lehmann, 1986), various families have

* Corresponding author.
E-mail address: raulmta@unb.br (R. Matsushita).

https://doi.org/10.1016/j.jspi.2025.106350
Received 14 March 2025; Received in revised form 19 July 2025; Accepted 12 September 2025
Available online 19 September 2025
0378-3758/

been considered, including the power divergence (Read and Cressie, 1988), the ϕ-divergence (Pardo, 2005; Jager and Wellner, 2007; Noughabi and Balakrishnan, 2016), the (h,ϕ)-divergence (Morales et al., 1994; Salicru et al., 1994), and the (ϕ,a)-power divergence (Vonta et al., 2012). Among these, Rényi divergence has been studied in the context of hypothesis testing in exponential families (Morales et al., 2000, 2004), where some of its asymptotic properties have been established.

The objective of this paper is to develop and evaluate a goodness-of-fit test statistic based on Rényi divergence, with special attention to sparse frequency tables. Specifically, we investigate the range $\alpha \in (0,1)$, where the Rényi divergence is known to be well-behaved even when small expected counts are present. Our central research question is: *Can small values of the Rényi index α produce test statistics with asymptotic chi-square behavior and improved performance compared to the classical χ^2 test in sparse settings?*

To answer this, we derive a theoretical approximation showing the asymptotic equivalence between the Rényi-based statistic and the χ^2 distribution as $\alpha \to 0$. The Rényi divergence emerges as a natural candidate for this purpose, as it avoids divisions by small frequencies when $\alpha \in (0,1)$, which is particularly relevant in sparse settings (Rathie, 1970).

Our work extends the application of the Rényi divergence to goodness-of-fit testing, offering a feasible alternative to the traditional χ^2 test for analyzing sparse frequency tables. Unlike previous methods based on Rényi divergence with $\alpha > 0$ (Morales et al., 2000, 2004), we focus exclusively on the interval $\alpha \in (0,1)$. Specifically, we investigate suitable values of α within this range that ensure the *chi-squareness* of the proposed approach.

We then conduct Monte Carlo simulations to evaluate the empirical significance level, power, and mean squared error of the test statistic across a range of sparsity scenarios. Finally, we illustrate the method using real-world data from traffic crashes and fingerprint pattern distributions.

The remainder of the paper is organized as follows. Section 2 presents the asymptotic expansion of the Rényi statistic and its relationship to the χ^2 distribution. Section 3 reports a Monte Carlo simulation study comparing the Rényi and χ^2 statistics under various conditions, with emphasis on sparse tables. Section 4 presents two applications: one using motor vehicle crash data from New York State, and the other involving fingerprint data from a Brazilian police database. Section 5 concludes.

2. The Rényi divergence goodness-of-fit statistic

Let $\mathbf{p} = (p_1,\dots,p_m)'$ denote the unknown discrete probability vector of the underlying distribution, and let $\mathbf{p}_0 = (p_{0,1},\dots,p_{0,m})'$ denote a fully specified probability vector under the null hypothesis. We consider the classical goodness-of-fit testing problem

$$H_0 : \mathbf{p} = \mathbf{p}_0 \quad \text{versus} \quad H_1 : \mathbf{p} \neq \mathbf{p}_0.$$

Given a random sample of size n, let $\hat{\mathbf{p}} = (\hat{p}_1,\dots,\hat{p}_m)'$ denote the empirical probability vector obtained from the observed frequencies. To assess the divergence between $\hat{\mathbf{p}}$ and $\mathbf{p}_0$, we consider the Rényi divergence of order $\alpha \in (0,1)$, stating our main result as follows:

Theorem. *Let $(N_1,\dots,N_m)'$ be a multinomially distributed vector with m category probabilities $\mathbf{p} = (p_1,\dots,p_m)'$ in a frequency table, where $\sum_{j=1}^m N_j = n$ is the sample size. Denoting the Rényi divergence of order $0 < \alpha < 1$ of the null distribution $\mathbf{p}_0$ from the empirical distribution $\hat{\mathbf{p}}$ as*

$$D_\alpha \equiv D_\alpha(\mathbf{p}_0 \parallel \hat{\mathbf{p}}) = \frac{1}{\alpha - 1} \ln \sum_{j=1}^m p_{0,j}^\alpha \hat{p}_j^{1-\alpha},$$

for a sufficiently large sample size n, the test statistic

$$R_{\alpha,n,m} = \frac{2n}{\alpha} D_\alpha(\mathbf{p}_0 \parallel \hat{\mathbf{p}})$$

is asymptotically chi-square with $m-1$ degrees of freedom.

Proof.

As $(N_1,\dots,N_m)' \sim \text{multinomial}(n,\mathbf{p}_0)$, the empirical distribution $\hat{\mathbf{p}} = (\hat{p}_1,\dots,\hat{p}_m)'$ consists of the maximum likelihood estimators of the category probabilities $p_{0,j}$. Specifically, for each category $j = 1,\dots,m$, we have $\hat{p}_j = N_j/n$, $\forall j \in \{1,\dots,m\}$. Thus, for a sufficiently large n, we get $(\hat{p}_1,\dots,\hat{p}_m)' \sim \text{multinormal}$, so that we can write

$$\hat{p}_j = p_{0,j} + \varepsilon_j, \tag{1}$$

where $\varepsilon_j \sim \mathrm{N}(0, p_{0,j}(1-p_{0,j})/n)$, $\sum_{j=1}^m \varepsilon_j = 0$, and $\mathrm{Cov}(\varepsilon_j,\varepsilon_{j'}) = -p_{0,j}p_{0,j'}/n$, for $j \neq j'$ (Casella and Berger, 2002). Thus, under $H_0 : (N_1,\dots,N_m)' \sim \text{multinomial}(n,\mathbf{p}_0)$, using (1), we can rewrite the Pearson's χ^2 statistic as

$$\chi^2 \equiv \chi^2(\mathbf{p}_0 \parallel \hat{\mathbf{p}}) = n \sum_{j=1}^m \frac{(\hat{p}_j - p_{0,j})^2}{p_{0,j}} = n \sum_{j=1}^m \frac{\varepsilon_j^2}{p_{0,j}}, \tag{2}$$

having approximately the chi-square distribution with $m-1$ degrees of freedom (χ^2_{m-1}) for a sufficiently large n.

Now, the Rényi divergence of order α of a H_0 distribution p from its empirical estimate $\hat{p}$ is

$$D_\alpha \equiv D_\alpha(\mathbf{p}_0 \parallel \hat{\mathbf{p}}) = \frac{1}{\alpha - 1} \ln A_\alpha(\mathbf{p}_0 \parallel \hat{\mathbf{p}}), \tag{3}$$

Table 1
Expected absolute frequencies under H_0, with $n = 2000$ and $m \in \{4, 6, 8\}$.

(a) Example 3.1: Truncated geometric (9)

($m = 4$)

π_0	x = 0	1	2	3
0.90	1800.18	180.02	18.00	1.80
0.95	1900.01	95.00	4.75	0.24

($m = 6$)

π_0	x = 0	1	2	3	4	5
0.7	1401.02	420.31	126.09	37.83	11.35	3.40
0.8	1800.00	180.00	18.00	1.80	0.18	0.02

($m = 8$)

π_0	x = 0	1	2	3	4	5	6	7
0.5	1003.92	501.96	250.98	125.49	62.75	31.37	15.69	7.84
0.7	1400.09	420.03	126.01	37.80	11.34	3.40	1.02	0.31

(b) Example 3.2: Right-censored Poisson (10)

($m = 4$)

λ_0	x = 0	1	2	3
0.1	1809.67	180.97	9.05	0.31
0.2	1637.46	327.49	32.75	2.30

($m = 6$)

λ_0	x = 0	1	2	3	4	5
0.7	993.17	695.22	243.33	56.78	9.94	1.57
0.8	898.66	718.93	287.57	76.69	15.34	2.82

($m = 8$)

λ_0	x = 0	1	2	3	4	5	6	7
1.0	735.76	735.76	367.88	122.63	30.66	6.13	1.02	0.17
1.1	665.74	732.32	402.77	147.68	40.61	8.93	1.64	0.30

where

$$A_\alpha \equiv A_\alpha(\mathbf{p}_0 \parallel \hat{\mathbf{p}}) = \sum_{j=1}^{m} p_{0,j}^{\alpha} \hat{p}_j^{1-\alpha}, \tag{4}$$

for $0 < \alpha < 1$, avoiding the division by a rarely observed frequency (Rathie, 1970). The Rényi coefficient (4) remains computable even when $p_{0,j} = 0$ for some j, unlike the χ^2 divergence in (2).

Rewriting (4) in terms of ε_j, we find

$$A_\alpha = \sum_{j=1}^{m} p_{0,j} \left(1 + \frac{\varepsilon_j}{p_{0,j}}\right)^{1-\alpha}, \tag{5}$$

with $p_{0,j} > 0$.

Next, we establish a large sample relation between D_α and χ^2 for $0 < \alpha < 1$ with the help of Slutsky's Theorem. Under H_0, as $\hat{p}_j$ is a consistent estimator of $p_{0,j}$, for sufficiently large n, we can expand $\left(1 + \varepsilon_j/p_{0,j}\right)^{1-\alpha}$ into its power series, which is valid for $|\varepsilon_j/p_{0,j}| < 1$, provided that $\sum_{j=1}^{m} p_{0,j} = 1$ and $\sum_{j=1}^{m} \varepsilon_j = 0$. From here, we proceed with (5) as follows.

$$\begin{aligned} A_\alpha &= \sum_{j=1}^{m} p_{0,j} \left[1 + (1-\alpha)\frac{\varepsilon_j}{p_{0,j}} - \frac{\alpha(1-\alpha)}{2!}\frac{\varepsilon_j^2}{p_{0,j}^2} + \sum_{k\geq 3} \binom{1-\alpha}{k} \left(\frac{\varepsilon_j}{p_{0,j}}\right)^k\right] \\ &= \sum_{j=1}^{m} \left[p_{0,j} + (1-\alpha)\varepsilon_j - \frac{\alpha(1-\alpha)}{2!}\frac{\varepsilon_j^2}{p_{0,j}} + \sum_{k\geq 3} \binom{1-\alpha}{k} \frac{\varepsilon_j^k}{p_{0,j}^{k-1}}\right] \end{aligned}$$

Table 2
Example 3.1 (Truncated geometric, $m = 4$): Empirical significance levels from 5000 samples ($n = 2000$) using the asymptotic approximation $R_{\alpha,n,m} \sim \chi^2_3$.

		nominal significance level (%)					
	α	0.005	0.01	0.02	0.03	0.04	0.05
$\pi_0 = 0.90$	10^{-7}	0.004	0.009	0.022	0.033	0.042	0.054
	10^{-6}	0.004	0.009	0.022	0.033	0.042	0.054
	10^{-5}	0.004	0.009	0.022	0.033	0.042	0.054
	10^{-4}	0.004	0.009	0.022	0.033	0.042	0.054
	10^{-3}	0.004	0.009	0.022	0.033	0.042	0.054
	10^{-2}	0.004	0.009	0.022	0.033	0.043	0.054
	10^{-1}	0.005	0.010	0.023	0.034	0.045	0.057
	0.20	0.005	0.011	0.025	0.040	0.048	0.064
	0.30	0.006	0.014	0.029	0.044	0.057	0.073
	0.40	0.009	0.018	0.037	0.057	0.077	0.096
	0.50	0.014	0.027	0.056	0.087	0.117	0.148
	0.60	0.027	0.057	0.120	0.186	0.190	0.195
	0.70	0.110	0.173	0.179	0.185	0.190	0.194
	0.80	0.169	0.173	0.179	0.186	0.190	0.194
	0.90	0.169	0.173	0.179	0.186	0.190	0.194
	0.95	0.169	0.173	0.179	0.185	0.190	0.194
	1.01	0.169	0.174	0.179	0.185	0.190	0.194
	1.10	0.170	0.174	0.179	0.185	0.190	0.194
	2.00	0.173	0.176	0.182	0.187	0.193	0.198
	χ^2	0.008	0.015	0.025	0.033	0.041	0.049
$\pi_0 = 0.95$	10^{-7}	0.003	0.007	0.022	0.029	0.037	0.042
	10^{-6}	0.003	0.007	0.022	0.029	0.037	0.042
	10^{-5}	0.003	0.007	0.022	0.029	0.037	0.042
	10^{-4}	0.003	0.007	0.022	0.029	0.037	0.042
	10^{-3}	0.003	0.007	0.022	0.029	0.037	0.042
	10^{-2}	0.003	0.008	0.022	0.029	0.036	0.042
	10^{-1}	0.004	0.012	0.022	0.028	0.035	0.041
	0.20	0.009	0.014	0.021	0.028	0.035	0.041
	0.30	0.012	0.014	0.021	0.028	0.034	0.042
	0.40	0.012	0.014	0.019	0.028	0.035	0.044
	0.50	0.012	0.015	0.021	0.028	0.039	0.047
	0.60	0.012	0.016	0.022	0.032	0.042	0.054
	0.70	0.014	0.017	0.027	0.040	0.053	0.073
	0.80	0.015	0.021	0.039	0.064	0.077	0.088
	0.90	0.030	0.065	0.090	0.123	0.156	0.197
	0.95	0.161	0.317	0.649	0.789	0.792	0.796
	1.01	0.784	0.785	0.786	0.790	0.793	0.796
	1.10	0.784	0.785	0.787	0.791	0.794	0.796
	2.00	0.792	0.792	0.793	0.793	0.794	0.796
	χ^2	0.033	0.034	0.040	0.046	0.056	0.061

$$= 1 + \frac{\alpha(\alpha-1)}{2n}\chi^2 + r(\alpha,\varepsilon), \tag{6}$$

where χ^2 is defined in (2) and $r(\alpha,\varepsilon)$ is the remainder.

Taking the linear approximation $\ln(1+x) \approx x$ for x small enough, from (3) and (6), we find

$$\frac{2n}{\alpha}D_\alpha \approx \chi^2 + \frac{2n}{\alpha(\alpha-1)}r(\alpha,\varepsilon) = \chi^2 + \mathcal{O}_p(1). \tag{7}$$

The remainder $nr(\alpha,\varepsilon_j) \to 0$ in probability, because $n\varepsilon_j \to 0$ and

$$P\left(\left|n\varepsilon_j^k\right| \geq \delta\right) \leq \frac{\mathrm{E}\left|n\varepsilon_j^k\right|^{2/k}}{\delta^{2/k}} = \frac{p_{0,j}(1-p_{0,j})}{n^{1-2/k}\delta^{2/k}} \to 0,$$

for any $\delta > 0$ and $k \geq 3$ as $n \uparrow \infty$. Therefore, since $\chi^2 \xrightarrow{D} \chi^2_{m-1}$ as $n \uparrow \infty$, it follows by Slutsky's theorem that

$$R_{\alpha,n,m} = \frac{2n}{\alpha}D_\alpha \xrightarrow{D} \chi^2_{m-1}, \tag{8}$$

as $n \to \infty$. □

The approximation (7) generally improves as α decreases, making the choice of a small α beneficial for better accuracy. Analyzing the remainder term $r(\varepsilon_j)$ relative to α in (7), it depends on the binomial coefficients asymptotically affected by the choice of α. For

Table 3
Example 3.1 (Truncated geometric, $m = 6$): Empirical significance levels from 5000 samples ($n = 2000$) using the asymptotic approximation $R_{\alpha,n,m} \sim \chi^2_5$.

		nominal significance level (%)					
	α	0.005	0.01	0.02	0.03	0.04	0.05
$\pi_0 = 0.70$	10^{-7}	0.006	0.010	0.022	0.030	0.043	0.055
	10^{-6}	0.006	0.010	0.022	0.030	0.043	0.055
	10^{-5}	0.006	0.010	0.022	0.030	0.043	0.055
	10^{-4}	0.006	0.010	0.022	0.030	0.043	0.055
	10^{-3}	0.006	0.010	0.022	0.030	0.043	0.055
	10^{-2}	0.006	0.010	0.022	0.030	0.043	0.056
	10^{-1}	0.006	0.011	0.024	0.034	0.046	0.061
	0.20	0.008	0.013	0.028	0.040	0.052	0.067
	0.30	0.010	0.017	0.033	0.047	0.060	0.074
	0.40	0.015	0.024	0.042	0.055	0.070	0.079
	0.50	0.025	0.037	0.051	0.059	0.071	0.080
	0.60	0.040	0.044	0.052	0.060	0.072	0.081
	0.70	0.040	0.045	0.053	0.061	0.073	0.082
	0.80	0.041	0.045	0.054	0.063	0.073	0.083
	0.90	0.041	0.046	0.054	0.064	0.074	0.084
	0.95	0.041	0.047	0.055	0.065	0.075	0.084
	1.01	0.042	0.047	0.056	0.065	0.076	0.085
	1.10	0.043	0.048	0.058	0.067	0.077	0.086
	2.00	0.051	0.060	0.074	0.086	0.099	0.108
	χ^2	0.006	0.011	0.020	0.030	0.042	0.053
$\pi_0 = 0.80$	10^{-7}	0.004	0.008	0.017	0.025	0.035	0.045
	10^{-6}	0.004	0.008	0.017	0.025	0.035	0.045
	10^{-5}	0.004	0.008	0.017	0.025	0.035	0.045
	10^{-4}	0.004	0.008	0.017	0.025	0.035	0.045
	10^{-3}	0.004	0.008	0.017	0.025	0.035	0.045
	10^{-2}	0.004	0.007	0.017	0.025	0.035	0.045
	10^{-1}	0.004	0.008	0.018	0.028	0.039	0.048
	0.20	0.005	0.010	0.021	0.032	0.045	0.055
	0.30	0.007	0.012	0.026	0.039	0.053	0.069
	0.40	0.010	0.019	0.035	0.056	0.073	0.091
	0.50	0.018	0.032	0.062	0.089	0.101	0.115
	0.60	0.045	0.073	0.092	0.103	0.113	0.122
	0.70	0.082	0.087	0.100	0.114	0.124	0.134
	0.80	0.086	0.098	0.118	0.140	0.161	0.188
	0.90	0.152	0.235	0.371	0.479	0.552	0.604
	0.95	0.633	0.635	0.638	0.640	0.643	0.644
	1.01	0.633	0.635	0.638	0.640	0.642	0.644
	1.10	0.634	0.635	0.638	0.640	0.643	0.644
	2.00	0.635	0.637	0.641	0.643	0.644	0.646
	χ^2	0.010	0.017	0.028	0.036	0.043	0.052

each $k \geq 3$, we can write

$$\frac{1}{\alpha(\alpha-1)} \cdot \binom{1-\alpha}{k} = \frac{(1-\alpha)(-\alpha)(-\alpha-1)\cdots(1-\alpha-k+1)}{k!\alpha(\alpha-1)}$$
$$= \frac{(-\alpha-1)\cdots(1-\alpha-k+1)}{k!}.$$

As α becomes smaller, this term diminishes significantly for each $k \geq 3$ because each factor in the numerator that includes α scales down with α, leading to a suppression of higher-order terms in $r(\epsilon_j)$. Thus, considering the dominant term in $r(\epsilon_j)$ for small α for the kth term in the summation, we find

$$\frac{\epsilon_j^k}{p_{0,j}^{k-1}} \cdot \frac{1}{\alpha(\alpha-1)} \cdot \binom{1-\alpha}{k} \sim \frac{\alpha^{k-2}}{k!} \frac{\epsilon_j^k}{p_{0,j}^{k-1}}.$$

Hence, each term in $r(\epsilon_j)$ scales approximately as α^{k-2} for $k \geq 3$. This scaling shows that the terms in $r(\epsilon_j)$ decrease rapidly as α decreases.

Indeed, the use of small values of α in the Rényi divergence is not only supported by asymptotic considerations but also reinforced by theoretical properties. As $\alpha \to 0$, $D_\alpha(P \parallel Q)$ becomes increasingly sensitive to the support of P and converges to $-\log Q(\mathrm{supp}(P))$, which vanishes when P is absolutely continuous with respect to Q (*e.g.*, van Erven and Harremos, 2014). This limiting behavior connects the Rényi divergence to the notion of absolute continuity, and, in an asymptotic framework, relates to contiguity, where sequences of probability measures remain mutually close in a likelihood-ratio sense. These properties help justify the stability of the Rényi-based statistic in sparse-data regimes.

Table 4
Example 3.1 (Truncated geometric, $m = 8$): Empirical significance levels from 5000 samples ($n = 2000$) using the asymptotic approximation $R_{\alpha,n,m} \sim \chi^2_7$.

		nominal significance level (%)					
	α	0.005	0.01	0.02	0.03	0.04	0.05
$\pi_0 = 0.50$	10^{-7}	0.006	0.011	0.022	0.032	0.043	0.055
	10^{-6}	0.006	0.011	0.022	0.032	0.043	0.055
	10^{-5}	0.006	0.011	0.022	0.032	0.043	0.055
	10^{-4}	0.006	0.011	0.022	0.032	0.043	0.055
	10^{-3}	0.006	0.011	0.022	0.032	0.043	0.055
	10^{-2}	0.006	0.011	0.022	0.032	0.043	0.055
	10^{-1}	0.006	0.011	0.023	0.032	0.043	0.055
	0.20	0.006	0.012	0.023	0.033	0.044	0.056
	0.30	0.006	0.013	0.023	0.034	0.045	0.057
	0.40	0.007	0.013	0.024	0.035	0.046	0.057
	0.50	0.007	0.014	0.025	0.036	0.046	0.058
	0.60	0.008	0.014	0.026	0.037	0.046	0.059
	0.70	0.008	0.014	0.027	0.037	0.048	0.061
	0.80	0.009	0.015	0.028	0.037	0.049	0.061
	0.90	0.009	0.016	0.028	0.040	0.050	0.064
	0.95	0.010	0.017	0.029	0.041	0.051	0.065
	1.01	0.010	0.017	0.029	0.042	0.052	0.066
	1.10	0.011	0.018	0.030	0.043	0.054	0.069
	2.00	0.023	0.034	0.051	0.063	0.076	0.087
	χ^2	0.006	0.010	0.022	0.032	0.043	0.053
$\pi_0 = 0.70$	10^{-7}	0.004	0.008	0.018	0.028	0.038	0.047
	10^{-6}	0.004	0.008	0.018	0.028	0.038	0.047
	10^{-5}	0.004	0.008	0.018	0.028	0.038	0.047
	10^{-4}	0.004	0.008	0.018	0.028	0.038	0.047
	10^{-3}	0.004	0.008	0.018	0.028	0.038	0.047
	10^{-2}	0.004	0.009	0.018	0.028	0.038	0.048
	10^{-1}	0.004	0.008	0.019	0.031	0.039	0.050
	0.20	0.005	0.011	0.024	0.035	0.045	0.057
	0.30	0.007	0.014	0.030	0.043	0.057	0.068
	0.40	0.011	0.020	0.042	0.055	0.068	0.080
	0.50	0.021	0.036	0.055	0.068	0.081	0.092
	0.60	0.038	0.050	0.067	0.085	0.100	0.117
	0.70	0.049	0.068	0.098	0.128	0.152	0.180
	0.80	0.106	0.157	0.249	0.311	0.357	0.388
	0.90	0.393	0.402	0.424	0.443	0.463	0.480
	0.95	0.473	0.545	0.650	0.717	0.767	0.808
	1.01	0.844	0.844	0.845	0.847	0.847	0.848
	1.10	0.844	0.844	0.846	0.847	0.847	0.848
	2.00	0.845	0.846	0.847	0.848	0.850	0.851
	χ^2	0.014	0.022	0.036	0.044	0.057	0.065

3. Validation and power assessment

In this section, we investigate via Monte Carlo simulations how the choice of a small α contributes to the *chi-squareness* of the proposed Rényi statistic. To evaluate this, we design three examples in which the expected absolute frequencies include values below five. Under the null hypothesis that the data originate from a specified distribution with $m \in \{4, 6, 8\}$ categories, we generate 5000 random samples of size $n = 2000$ for each case. The divergence between observed and expected frequencies is measured using the Rényi statistic $R_{\alpha,n,m}$, focusing on small values of α (from 10^{-7} to 10^{-1}), while also including values slightly above one and $\alpha = 2$ for comparison. We use the Pearson's χ^2 statistic (2) as a benchmark throughout the analysis. These examples employ parametric forms to facilitate the study of the test's power function.

Example 3.1. Suppose we wish to test whether a simple random sample arises from a truncated geometric distribution with probability mass function

$$p_j = \frac{\pi(1-\pi)^j}{1-(1-\pi)^m}, \tag{9}$$

where $j \in \{0, 1, \ldots, m-1\}$, m is a fixed positive integer defining the truncation point, and $\pi \in (0, 1)$ is the success probability parameter. We consider testing the simple null hypothesis

$$H_0 : \pi = \pi_0 \quad \text{versus} \quad H_1 : \pi \neq \pi_0,$$

where $\pi_0 \in (0, 1)$ is a specified value under H_0. □

Table 5
Example 3.2 (Right-censored Poisson, $m = 4$): Empirical significance levels from 5000 samples ($n = 2000$) using the asymptotic approximation $R_{\alpha,n,m} \sim \chi^2_3$.

		nominal significance level (%)					
	α	0.005	0.01	0.02	0.03	0.04	0.05
$\lambda_0 = 0.1$	10^{-7}	0.004	0.008	0.017	0.027	0.033	0.040
	10^{-6}	0.004	0.008	0.017	0.027	0.033	0.040
	10^{-5}	0.004	0.008	0.017	0.027	0.033	0.040
	10^{-4}	0.004	0.008	0.017	0.027	0.033	0.040
	10^{-3}	0.004	0.008	0.017	0.027	0.033	0.040
	10^{-2}	0.004	0.008	0.016	0.027	0.033	0.040
	10^{-1}	0.004	0.008	0.017	0.028	0.032	0.039
	0.20	0.004	0.008	0.018	0.026	0.032	0.040
	0.30	0.004	0.008	0.019	0.025	0.032	0.040
	0.40	0.005	0.008	0.019	0.025	0.033	0.042
	0.50	0.005	0.011	0.020	0.026	0.034	0.046
	0.60	0.005	0.012	0.021	0.030	0.042	0.049
	0.70	0.008	0.014	0.026	0.038	0.046	0.057
	0.80	0.011	0.019	0.039	0.050	0.064	0.086
	0.90	0.036	0.062	0.129	0.195	0.270	0.340
	0.95	0.595	0.736	0.739	0.741	0.743	0.745
	1.01	0.735	0.737	0.739	0.740	0.742	0.745
	1.10	0.735	0.737	0.739	0.740	0.742	0.745
	2.00	0.737	0.739	0.740	0.742	0.743	0.745
	χ^2	0.014	0.023	0.041	0.057	0.060	0.066
$\lambda_0 = 0.2$	10^{-7}	0.002	0.008	0.022	0.034	0.044	0.057
	10^{-6}	0.002	0.008	0.022	0.034	0.044	0.057
	10^{-5}	0.002	0.008	0.022	0.034	0.044	0.057
	10^{-4}	0.002	0.008	0.022	0.034	0.044	0.057
	10^{-3}	0.002	0.008	0.022	0.034	0.044	0.057
	10^{-2}	0.002	0.008	0.022	0.034	0.044	0.057
	10^{-1}	0.003	0.010	0.024	0.037	0.050	0.063
	0.20	0.003	0.013	0.027	0.043	0.058	0.074
	0.30	0.005	0.015	0.035	0.053	0.074	0.091
	0.40	0.009	0.022	0.049	0.077	0.107	0.134
	0.50	0.018	0.042	0.091	0.127	0.136	0.142
	0.60	0.056	0.111	0.119	0.127	0.135	0.142
	0.70	0.108	0.112	0.119	0.128	0.134	0.142
	0.80	0.108	0.112	0.119	0.128	0.134	0.142
	0.90	0.109	0.113	0.119	0.128	0.134	0.142
	0.95	0.109	0.113	0.119	0.127	0.135	0.141
	1.01	0.109	0.113	0.120	0.128	0.135	0.141
	1.10	0.109	0.113	0.121	0.128	0.135	0.142
	2.00	0.112	0.115	0.124	0.131	0.140	0.144
	χ^2	0.007	0.011	0.022	0.031	0.041	0.049

Example 3.2. Let the null model be a right-censored Poisson distribution with rate parameter $\lambda > 0$ and support of length m. Its probability mass function is given by

$$p_j = \begin{cases} e^{-\lambda}\lambda^j/j!, & \text{if } j \in \{0, 1, \dots, m-2\}, \\ \displaystyle\sum_{k=m-1}^{\infty} e^{-\lambda}\lambda^k/k!, & \text{if } j = m-1. \end{cases} \tag{10}$$

We test the null hypothesis that the data follow this distribution with a prespecified rate λ_0:

$$H_0 : \lambda = \lambda_0 \quad \text{versus} \quad H_1 : \lambda \neq \lambda_0. \quad \square$$

Table 1 presents the expected absolute frequencies under H_0 for $n = 2000$. For $\pi_0 = 0.5$ in Example 3.1, the lowest expected frequency exceeds five, suggesting that the χ^2 statistic should perform well, in line with the conventional rule of thumb. Conversely, when $\pi_0 = 0.7$, three expected frequencies fall below five (3.40, 1.02, and 0.31), which may compromise the validity of the χ^2 approximation. Similarly, in Example 3.2, we also identify small expected frequencies, raising concerns about the applicability of the χ^2 test. These examples highlight the importance of evaluating alternative test statistics, such as the Rényi divergence, particularly under sparse frequency scenarios.

Tables 2–7 present the empirical significance levels obtained using critical values from the asymptotic approximation $R_{\alpha,n,m} \sim \chi^2_{m-1}$ at the corresponding nominal significance levels. For each configuration, we generated 5000 random samples of size $n = 2000$.

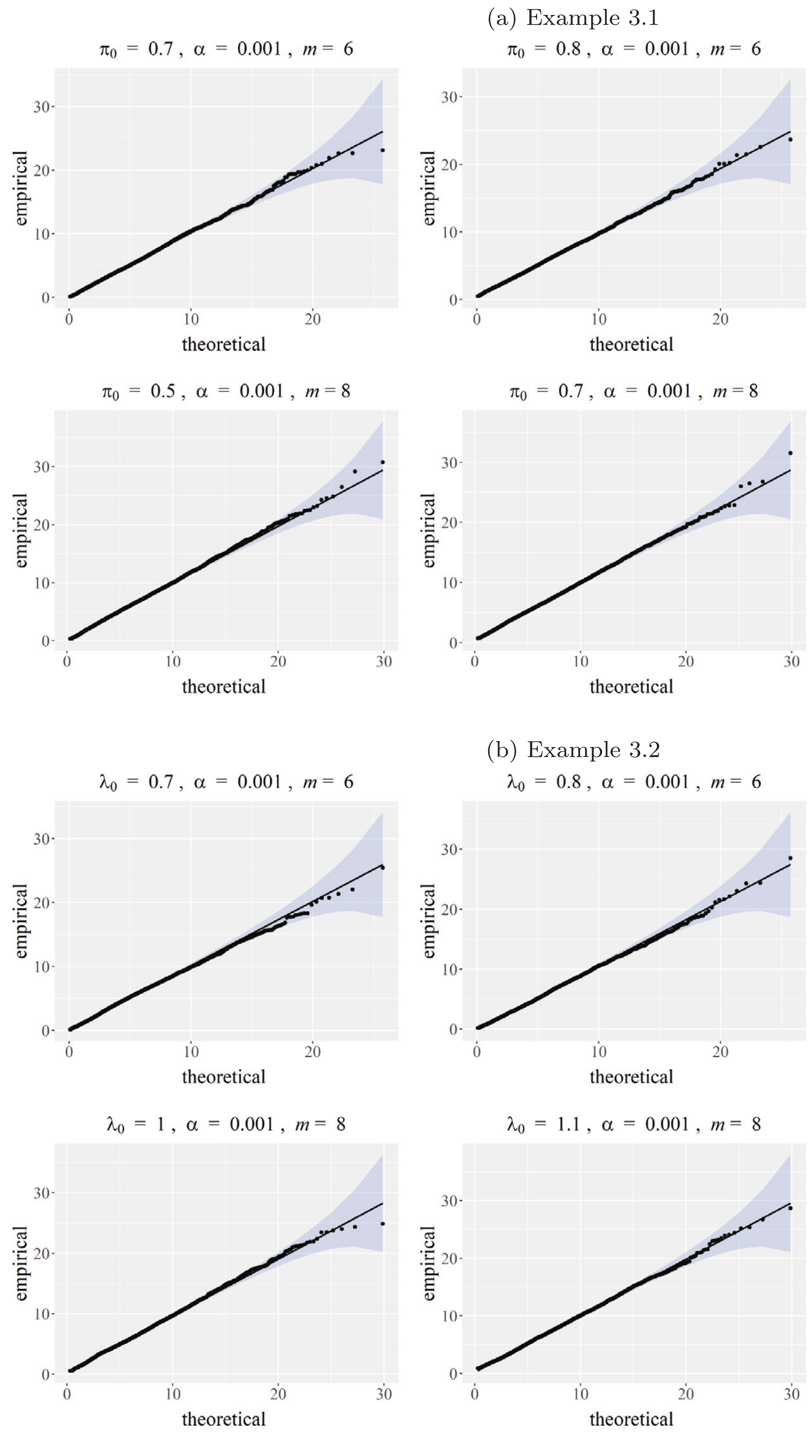

Fig. 1. Quantile-quantile plots comparing the distribution of $R_{\alpha,n,m}$ ($\alpha = 0.001$; 5000 replications; $n = 2000$) under H_0 from Examples 3.1 and 3.2 with the theoretical χ^2 distribution: χ^2_5 for $m = 6$ and χ^2_7 for $m = 8$.

Table 6
Example 3.2 (Right-censored Poisson, $m = 6$): Empirical significance levels from 5000 samples ($n = 2000$) using the asymptotic approximation $R_{\alpha,n,m} \sim \chi^2_5$.

		nominal significance level (%)					
	α	0.005	0.01	0.02	0.03	0.04	0.05
$\lambda_0 = 0.7$	10^{-7}	0.004	0.008	0.016	0.028	0.040	0.052
	10^{-6}	0.004	0.008	0.016	0.028	0.040	0.052
	10^{-5}	0.004	0.008	0.016	0.028	0.040	0.052
	10^{-4}	0.004	0.008	0.016	0.028	0.040	0.052
	10^{-3}	0.004	0.008	0.016	0.028	0.040	0.052
	10^{-2}	0.004	0.009	0.016	0.028	0.040	0.053
	10^{-1}	0.004	0.009	0.017	0.030	0.043	0.056
	0.20	0.004	0.009	0.019	0.032	0.047	0.061
	0.30	0.005	0.010	0.022	0.037	0.053	0.070
	0.40	0.007	0.013	0.028	0.046	0.063	0.081
	0.50	0.008	0.018	0.039	0.061	0.080	0.102
	0.60	0.015	0.031	0.063	0.094	0.120	0.145
	0.70	0.043	0.078	0.135	0.180	0.208	0.234
	0.80	0.191	0.214	0.221	0.229	0.234	0.243
	0.90	0.211	0.215	0.222	0.229	0.235	0.243
	0.95	0.211	0.215	0.222	0.229	0.235	0.243
	1.01	0.212	0.216	0.223	0.229	0.236	0.244
	1.10	0.212	0.216	0.223	0.230	0.237	0.244
	2.00	0.219	0.223	0.229	0.236	0.243	0.251
	χ^2	0.006	0.012	0.022	0.029	0.042	0.051
$\lambda_0 = 0.8$	10^{-7}	0.006	0.011	0.023	0.035	0.048	0.058
	10^{-6}	0.006	0.011	0.023	0.035	0.048	0.058
	10^{-5}	0.006	0.011	0.023	0.035	0.048	0.058
	10^{-4}	0.006	0.011	0.023	0.035	0.048	0.058
	10^{-3}	0.006	0.011	0.023	0.035	0.048	0.058
	10^{-2}	0.006	0.011	0.023	0.035	0.048	0.059
	10^{-1}	0.007	0.012	0.025	0.038	0.050	0.063
	0.20	0.008	0.013	0.030	0.042	0.056	0.067
	0.30	0.009	0.017	0.034	0.049	0.063	0.079
	0.40	0.013	0.023	0.043	0.061	0.079	0.094
	0.50	0.021	0.035	0.064	0.084	0.097	0.106
	0.60	0.044	0.068	0.081	0.089	0.099	0.106
	0.70	0.068	0.071	0.081	0.089	0.099	0.107
	0.80	0.068	0.071	0.081	0.090	0.099	0.107
	0.90	0.068	0.072	0.081	0.090	0.099	0.108
	0.95	0.068	0.072	0.081	0.090	0.099	0.108
	1.01	0.068	0.073	0.082	0.091	0.100	0.109
	1.10	0.068	0.074	0.083	0.092	0.100	0.109
	2.00	0.073	0.080	0.090	0.102	0.111	0.119
	χ^2	0.007	0.012	0.023	0.032	0.045	0.054

Compared to the chi-square statistic, $R_{\alpha,n,m}$ tends to maintain empirical significance levels closer to the nominal values, particularly for small values of α, even in scenarios with very low expected frequencies, such as $H_0 : \pi_0 = 0.7$ in Example 3.1 and the null hypotheses in Example 3.2. As α increases, the asymptotic approximation becomes less accurate, and the discrepancy between nominal and empirical levels may grow. These findings confirm that the approximation (7) improves as α decreases, thereby supporting the theoretical justification for using small divergence indices in practice.

Figs. 1(a) and 1(b) display quantile–quantile (QQ) plots comparing the distribution of $R_{\alpha,n,m}$ with $\alpha = 0.001$ to the theoretical χ^2_{m-1} distribution, illustrating the chi-squareness of the proposed statistic under the null hypothesis in both examples.

The empirical power functions of the proposed statistic, evaluated at a nominal significance level of 5% using the asymptotic approximation, are illustrated for $\pi_0 = 0.7$ (Fig. 2(a)) and $\lambda_0 = 1.0$ (Fig. 2(b)). Variations across small values of α have little impact on the shape of the power function. For the $R_{\alpha,n,m}$ statistic, the presence of small expected frequencies does not distort the symmetry of the power curve around the null, in contrast to the behavior observed for the χ^2 statistic.

To assess robustness, we evaluated the performance of the $R_{\alpha,n,m}$ and χ^2 statistics in terms of mean square error (MSE) under H_0, varying the sample size from 50 to 3000. Figs. 3(a) and 3(b) illustrate how the MSEs of $R_{0.001,n,8}$ and the χ^2 statistic evolve as n increases. As expected, low expected counts induce large fluctuations in the χ^2 statistic. In contrast, the Rényi statistic yields MSE values that remain stable across different sample sizes. This suggests that $R_{\alpha,n,m}$ exhibits more efficient behavior than the χ^2 statistic in sparse settings.

Table 7
Example 3.2 (Right-censored Poisson, $m = 8$): Empirical significance levels from 5000 samples ($n = 2000$) using the asymptotic approximation $R_{\alpha,n,m} \sim \chi^2_7$.

		nominal significance level (%)					
	α	0.005	0.01	0.02	0.03	0.04	0.05
$\lambda_0 = 1.0$	10^{-7}	0.004	0.008	0.018	0.026	0.037	0.045
	10^{-6}	0.004	0.008	0.018	0.026	0.037	0.045
	10^{-5}	0.004	0.008	0.018	0.026	0.037	0.045
	10^{-4}	0.004	0.008	0.018	0.026	0.037	0.045
	10^{-3}	0.004	0.008	0.018	0.026	0.037	0.045
	10^{-2}	0.004	0.008	0.018	0.026	0.037	0.045
	10^{-1}	0.004	0.008	0.019	0.027	0.038	0.045
	0.20	0.004	0.010	0.020	0.030	0.039	0.049
	0.30	0.006	0.011	0.021	0.031	0.042	0.052
	0.40	0.007	0.013	0.024	0.036	0.049	0.057
	0.50	0.009	0.015	0.031	0.044	0.055	0.066
	0.60	0.011	0.020	0.039	0.057	0.070	0.084
	0.70	0.020	0.036	0.062	0.090	0.114	0.139
	0.80	0.058	0.103	0.176	0.231	0.280	0.318
	0.90	0.364	0.370	0.378	0.388	0.400	0.408
	0.95	0.375	0.390	0.419	0.446	0.470	0.497
	1.01	0.897	0.898	0.899	0.900	0.901	0.902
	1.10	0.898	0.898	0.899	0.901	0.901	0.902
	2.00	0.898	0.899	0.900	0.901	0.902	0.903
	χ^2	0.020	0.027	0.037	0.044	0.055	0.066
$\lambda_0 = 1.1$	10^{-7}	0.004	0.008	0.020	0.030	0.041	0.051
	10^{-6}	0.004	0.008	0.020	0.030	0.041	0.051
	10^{-5}	0.004	0.008	0.020	0.030	0.041	0.051
	10^{-4}	0.004	0.008	0.020	0.030	0.041	0.051
	10^{-3}	0.004	0.008	0.020	0.030	0.041	0.051
	10^{-2}	0.004	0.008	0.020	0.030	0.041	0.050
	10^{-1}	0.005	0.009	0.020	0.032	0.042	0.053
	0.20	0.005	0.009	0.022	0.034	0.044	0.054
	0.30	0.006	0.012	0.024	0.038	0.048	0.059
	0.40	0.007	0.014	0.028	0.044	0.057	0.071
	0.50	0.011	0.020	0.038	0.058	0.074	0.092
	0.60	0.019	0.030	0.063	0.089	0.108	0.130
	0.70	0.045	0.077	0.126	0.163	0.195	0.215
	0.80	0.180	0.202	0.216	0.227	0.237	0.247
	0.90	0.212	0.224	0.249	0.276	0.300	0.324
	0.95	0.309	0.393	0.521	0.613	0.685	0.736
	1.01	0.796	0.796	0.798	0.799	0.800	0.803
	1.10	0.796	0.796	0.798	0.799	0.800	0.803
	2.00	0.797	0.797	0.799	0.800	0.802	0.804
	χ^2	0.012	0.020	0.033	0.047	0.059	0.072

Example 3.3. Let a random sample of size n be drawn from a binomial distribution with parameters $m - 1$ (number of Bernoulli trials) and success probability $\pi \in (0, 1)$. We consider testing

$$H_0 : \pi = \pi_0 \quad \text{versus} \quad H_1 : \pi \neq \pi_0,$$

where π_0 is a specified value under the null hypothesis. □

In Example 3.3, we study the sensitivity of the MSE to changes in π_0, considering different ranges for each value of $m \in \{4, 6, 8\}$. Fig. 4 shows that the Rényi statistic with $\alpha = 0.001$ exhibits more stable behavior than the χ^2 statistic in the presence of low expected frequencies, consistently across all considered values of m. Table 8 illustrates how small expected counts arise as π_0 decreases, inflating the MSE of the χ^2 statistic.

4. Applications

4.1. Vehicle crash data

We apply our proposed method to real-world motor vehicle crash data from the https://data.ny.gov (New York State open data portal). The dataset includes daily crash counts reported to the New York State Department of Motor Vehicles by motorists and police

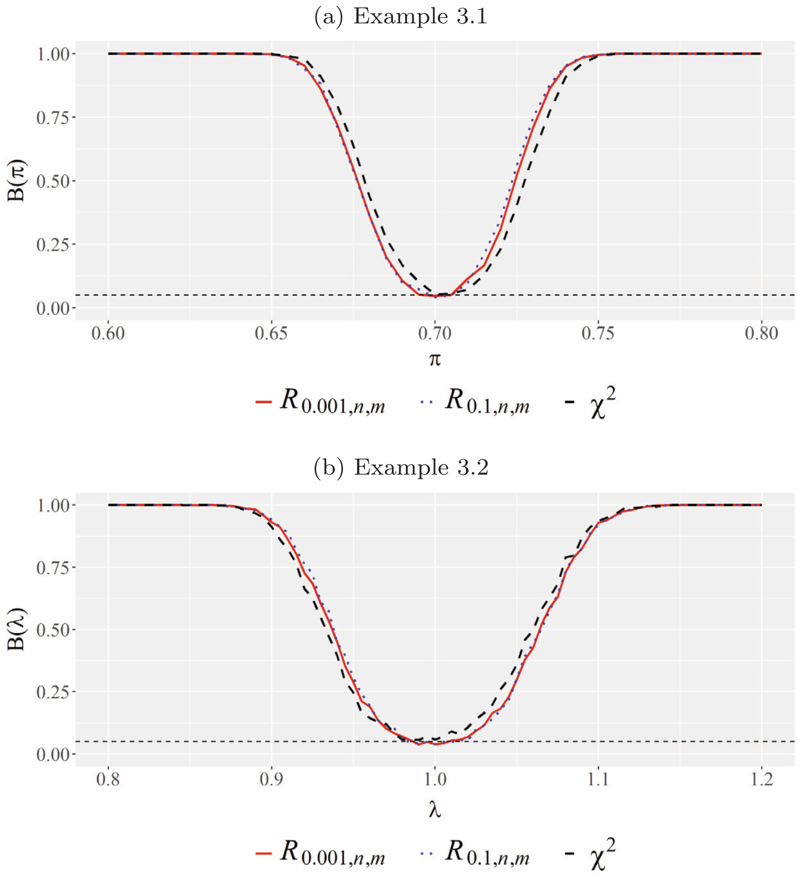

Fig. 2. Empirical power functions of the proposed statistic from Examples 3.1 and 3.2 at a nominal significance level of 5%, using critical values from the asymptotic approximation (χ^2_7).

agencies between 2018 and 2022. Our illustrative analysis focuses on crashes between 6:00 AM and 6:25 AM in six municipalities: Buffalo, New York City, Richmond, Islip, Kings, and Queens counties. The total sample comprises 1826 days.

The goal is to test whether the observed daily crash counts in these municipalities can be adequately described as a Touchard random variable X defined by the probability mass function

$$p_j = \frac{\lambda^j (j+1)^\delta}{j!\,\tau(\lambda,\delta)}, \tag{11}$$

where $\lambda > 0$ and $\delta \in \mathbb{R}$ are parameters, $j \in \mathbb{N}$, and $\tau(\lambda,\delta) = \sum_{k\in\mathbb{N}} \lambda^k (k+1)^\delta / k!$ is the normalizing function (Matsushita et al., 2019). The Touchard distribution is particularly relevant for count data where overdispersion or clustering effects may be present (De Andrade et al., 2021), making it suitable for analyzing crash frequency data.

Table 9 compares the observed daily crash counts with the expected values under the Touchard model for each municipality. The parameter values λ_0 and δ_0 under H_0 are reported in Table 10.

Despite small expected counts in some categories, the observed data align reasonably well with the model's predictions for most municipalities. Although the Touchard distribution has unbounded support, categories with negligible expected frequencies were treated as zero. Accordingly, the test was performed using the m observed categories shown in Table 9 for each municipality.

To formally assess the goodness of fit, we employed both the Rényi divergence with $\alpha = 0.001$ and the traditional chi-square test statistics. Table 10 summarizes the test results, including the Rényi divergence statistic $R_{\alpha,n,m}$, chi-square values, and the corresponding p-values.

The following observations emerged: In Buffalo, New York City, Richmond, and Queens, both the Rényi divergence statistic and the chi-square test indicate an adequate fit, with p-values exceeding typical significance levels. However, in Islip, the chi-square test identifies a poor fit (p-value = 0.003), whereas the Rényi statistic remains robust, suggesting that the model adequately describes

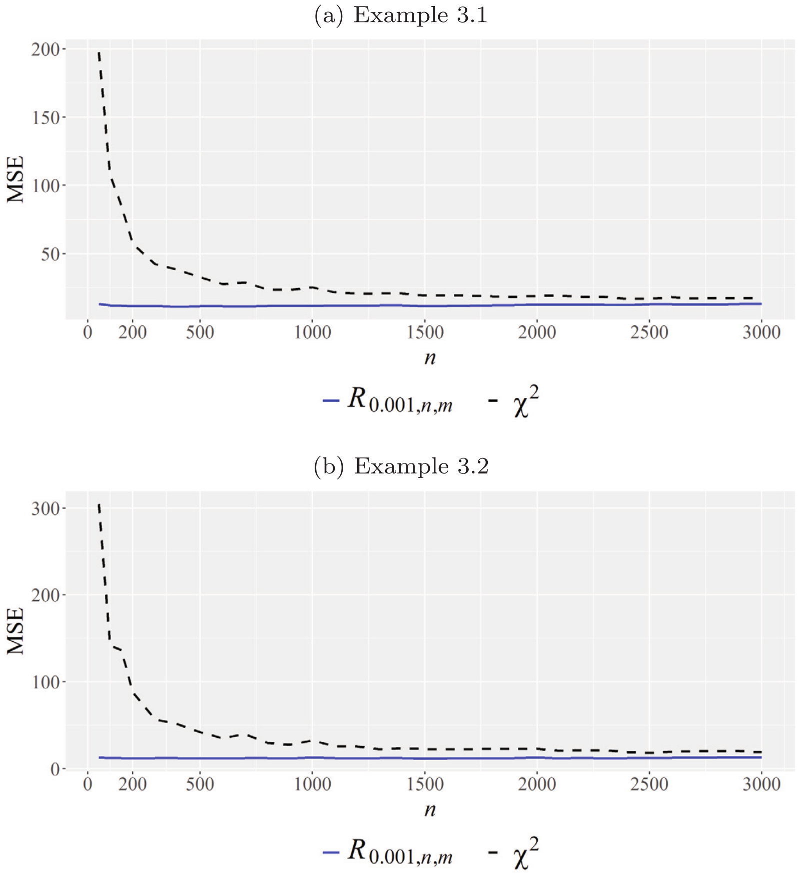

Fig. 3. Empirical MSE of $R_{\alpha,n,m}$ ($\alpha = 0.001$, $m = 8$) and the χ^2 statistic under H_0 from Examples 3.1 and 3.2, based on 5000 replications with sample sizes ranging from $n = 50$ to 3000.

the data (p-value = 0.184). Grouping small-count categories as $j \geq 3$ for the chi-square test yields $\chi^2 = 2.95$ with a p-value of 0.399, supporting a better fit.

For Buffalo, we also performed the chi-square test after combining categories $x = 2, 3$, and 4, as recommended in the literature. In this case, the test yields $\chi^2 = 0.74$ with a p-value of 0.689, further confirming Touchard as a possible model.

These results highlight the robustness of the Rényi divergence statistic in scenarios involving small expected counts, where the chi-square test may fail due to its sensitivity to sparse data. The Rényi statistic demonstrates stability and consistency across all municipalities, even when the observed crash frequencies involve rare events.

4.2. Fingerprint minutiae data

Fingerprint identification is a crucial component of forensic science, relying on the analysis of minutiae—distinct ridge patterns such as bifurcations and ridge endings. These features remain stable throughout an individual's lifetime and are sufficiently complex to distinguish between individuals. We analyzed a sample of 2027 minutiae extracted from a Brazilian police record to compare with the expected distribution reported by Gomes et al. (2024). Table 11 presents the expected and observed counts. Given the forensic importance of rare minutiae, we identified ten categories with expected frequencies below five.

Because this sample originates from the same population studied by Gomes et al. (2024), the null hypothesis should be retained. Using the asymptotic result in Eq. (8), we obtain $R_{\alpha,n,m} = 40.53$ (p-value = 0.075) with $\alpha = 0.0001$, $n = 2027$, and $m = 30$, supporting a good fit. In contrast, the chi-square test yields $\chi^2_{n,m} = 44.74$ (p-value = 0.031), indicating rejection at conventional levels.

To address the issue of low expected frequencies, we consolidated the ten least frequent minutiae types into a single category and recalculated the statistics. After this adjustment, we find $R_{\alpha,n,m} = 24.07$ (p-value = 0.193) and $\chi^2_{n,m} = 26.63$ (p-value = 0.113), both of which support the null. These results illustrate the robustness of the Rényi divergence in sparse data settings and reinforce

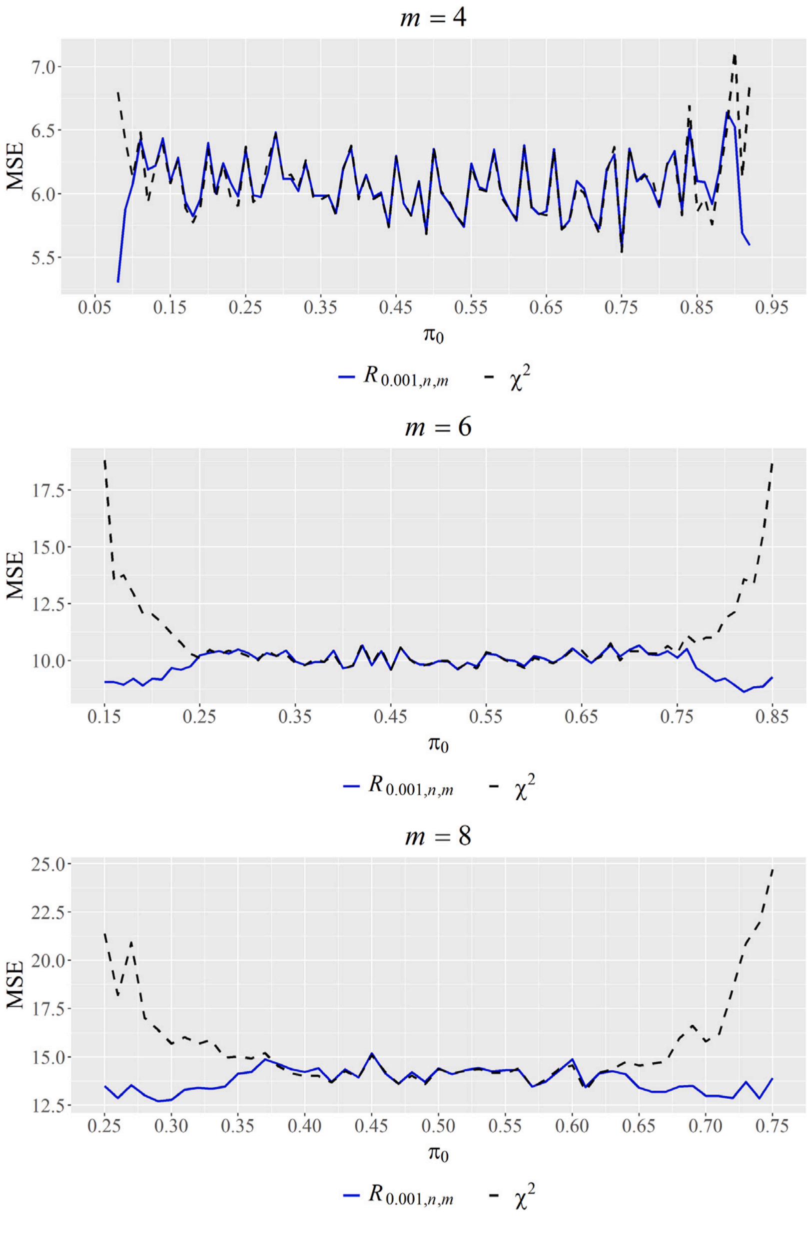

Fig. 4. Empirical MSE comparison between $R_{\alpha,n,m}$ ($\alpha = 0.001$) and χ^2 statistics under H_0 in Example 3.3, for different values of $m \in \{4, 6, 8\}$. Results are based on 5000 replications with a sample size of 2000.

Table 8
Example 3.3: Expected absolute frequencies under H_0, with $n = 2000$ and $m \in \{4, 6, 8\}$.

($m = 4$)

π_0	x = 0	1	2	3
0.10	1458	486	54	2
0.15	1228.25	650.25	114.75	6.75

($m = 6$)

π_0	x = 0	1	2	3	4	5
0.15	887.41	783.01	276.36	48.77	4.30	0.15
0.20	655.36	819.20	409.60	102.40	12.80	0.64
0.25	474.61	791.02	527.34	175.78	29.30	1.95
0.30	336.14	720.30	617.40	264.60	56.70	4.86
0.35	232.06	624.77	672.83	362.29	97.54	10.50

($m = 8$)

π_0	x = 0	1	2	3	4	5	6	7
0.25	266.97	622.92	622.92	346.07	115.36	23.07	2.56	0.12
0.30	164.71	494.13	635.30	453.79	194.48	50.01	7.14	0.44
0.35	98.04	369.55	596.97	535.74	288.48	93.20	16.73	1.29
0.40	55.99	261.27	522.55	580.61	387.07	154.83	34.41	3.28
0.45	30.45	174.39	428.04	583.70	477.57	234.44	63.94	7.47
0.50	15.63	109.38	328.12	546.87	546.87	328.12	109.38	15.63

its value as a practical alternative to the chi-square test, particularly in forensic contexts where rare minutiae play a critical role in identification.

5. Conclusion

This work explores Rényi divergence as a practical substitute for the conventional χ^2 test in evaluating goodness-of-fit for sparse frequency tables. Theoretical analyses, supported by Monte Carlo simulations, demonstrate that selecting a small Rényi index α improves the statistic's adherence to the chi-square distribution and enhances its precision relative to the traditional test. As α decreases, higher-order terms in the asymptotic expansion become negligible, improving the robustness of the test in sparse data scenarios. This behavior is especially relevant in applications where sparse data compromise the performance of the χ^2 statistic. Furthermore, the empirical power functions suggested that the Rényi divergence consistently detects deviations from the null hypothesis, maintaining symmetry and reliability even when expected counts are very small.

It is important to note that our results hold asymptotically, requiring a sufficiently large sample size to ensure the validity of the distributional approximation. For small sample sizes, alternative resampling techniques, such as the bootstrap or Monte Carlo methods, should be employed to obtain more accurate critical values and assess the goodness of fit.

The practical utility of Rényi divergence was further demonstrated in Section 4 through two illustrative analyses of real-world motor vehicle crash and fingerprint data. The results showed that while the traditional chi-square test struggled with small counts in some cases, Rényi divergence remained robust and provided consistent goodness-of-fit assessments.

While this study focuses on simple hypotheses with fully specified parameters under H_0, an important direction for future work involves extending the methodology to composite hypotheses involving unknown parameters. In such cases, parameter estimation would introduce additional variability, potentially altering the limiting distribution of the test statistic. Prior studies have shown that Rényi-type statistics can still perform well in such contexts, particularly within exponential family models (*e.g.*, Morales et al., 2004). Exploring this extension would enhance the applicability of the proposed approach in broader practical settings.

These findings highlight the potential value of information-theoretic measures, such as Rényi divergence, in modern statistical inference, particularly in cases where data sparsity poses significant challenges. Beyond these insights, the practical applicability of the Rényi statistic in fields such as rare event analysis, genetic studies, and financial risk modeling warrants further investigation. A natural extension of this work is to conduct a comprehensive empirical comparison between the Rényi-based test and other goodness-of-fit procedures specifically designed for sparse frequency tables, including those based on power divergence and resampling methods. By broadening the scope of goodness-of-fit testing, this study contributes to the ongoing adaptation of statistical methodologies to meet contemporary analytical challenges.

Table 9
Comparison of observed and expected daily crash frequencies: Observed crash counts and corresponding expected values under the Touchard model for six municipalities in New York State, 2018–2022.

(a) Buffalo

	x				
	0	1	2	3	4
observed	1,587	218	17	3	1
expected	1,588.3	212.7	22.8	2.0	0.1

(b) New York

	x				
	0	1	2	3	4
observed	1,113	533	139	37	4
expected	1,113.8	528.0	148.8	30.0	4.7

(c) Richmond

	x					
	0	1	2	3	4	5
observed	1,571	219	30	5	0	1
expected	1,572.1	215.1	33.3	4.8	0.6	0.1

(d) Islip

	x					
	0	1	2	3	4	5
observed	1,449	325	42	8	1	1
expected	1,451.4	315.4	51.7	6.7	0.7	0.1

(e) Kings

	x						
	0	1	2	3	4	5	6
observed	709	684	297	102	23	10	1
expected	715.7	660.3	316.3	102.6	25.2	5.0	0.8

(f) Queens

	x							
	0	1	2	3	4	5	6	7
observed	573	625	379	157	57	27	6	2
expected	581.5	602.1	380.9	174.4	62.6	18.5	4.7	1.0

Table 10
Goodness-of-fit test results for crash data: Rényi divergence ($\alpha = 0.001$) and chi-square statistics with associated p-values assessing the fit of the Touchard model to motor vehicle crash data.

	H_0 Touchard model		statistics			
municipality	λ_0	δ_0	$R_{\alpha,n,m}$	p-value	χ^2	p-value
Buffalo	0.41	−1.64	4.406	0.221	7.256	0.064
New York	0.72	−0.60	3.703	0.295	2.434	0.487
Richmond	0.98	−2.84	6.536	0.257	7.428	0.191
Islip	0.58	−1.43	6.215	0.184	16.080	0.003
Kings	1.01	−0.13	6.536	0.257	7.428	0.191
Queens	1.68	−0.70	8.284	0.218	8.477	0.205

Table 11
Comparison of expected and observed minutiae frequencies in a forensic fingerprint sample (n = 2027). The expected distribution is taken from Gomes et al. (2024). Due to data sparsity, the ten least frequent categories (listed below the dashed line) were grouped in a second analysis.

Minutiae type	Expected distribution	Counts: Expected	Counts: Observed
Ridge ending	0.44005	891.99	912
Bifurcation	0.19308	391.37	366
Convergence	0.15064	305.35	290
Break	0.02751	55.76	70
Point between the ridges	0.02635	53.42	50
Small fragment or island	0.02392	48.49	44
Overlap	0.01758	35.64	30
Conjugation	0.01507	30.55	30
Big fragment	0.01477	29.93	35
Crossbar	0.01414	28.66	30
Bridge	0.01197	24.27	24
Small enclosure	0.01094	22.17	25
Big enclosure	0.00920	18.64	20
Tripod	0.00835	16.92	22
Opposite bifurcations	0.00567	11.50	10
Others	0.00510	10.35	12
Dock	0.00486	9.85	15
Appendage	0.00393	7.97	8
Point in the ridge	0.00340	6.90	15
Angular	0.00324	6.57	3
Double bifurcation	0.00227	4.60	1
Emboque	0.00172	3.49	1
Double convergence	0.00140	2.83	1
Return	0.00136	2.75	3
M-B type	0.00124	2.50	2
M-C type	0.00093	1.89	3
Trifurcation of type B	0.00049	0.99	3
Needle	0.00043	0.86	0
Numerical	0.00022	0.45	0
Trifurcation of type C	0.00018	0.37	2
Total	1.00	2027	2027

Acknowledgments

Financial support from FAPDF (Grant number 00193-00001860/2023-17), CNPq (Grant number 311548/2022-9), UnB (Edital DPI/DPG 04/2025) and Capes (Finance Code 001) is acknowledged.

References

Baglivo, J., Olivier, D., Pagano, M., 1992. Methods for exact goodness-of-fit tests. J. Amer. Statist. Assoc. 87 (418), 464–469, URL http://www.jstor.org/stable/2290278.

Casella, G., Berger, R.L., 2002. Statistical Inference, second ed. Duxbury Press, Pacific Grove.

De Andrade, B.B., Matsushita, R.M., Rathie, P.N., Ozelim, L., De Oliveira, S.B., 2021. On a weighted Poisson distribution and its associated regression model. Chil. J. Stat. 12, 229–252.

van Erven, T., Harremos, P., 2014. Rényi divergence and Kullback-Leibler divergence. IEEE Trans. Inform. Theory 60 (7), 3797–3820. http://dx.doi.org/10.1109/TIT.2014.2320500.

Gautam, S., Kimeldorf, G., 1999. Some results on the maximal correlation in $2 \times k$ contingency tables. Amer. Statist. 53 (4), 336–341. http://dx.doi.org/10.1080/00031305.1999.10474486.

Gomes, G.A.S., de Oliveira, L.P.M., da Silva Carvalho, D., de Araújo Brito, F.C., Matsushita, R.Y., 2024. Standardizing fingerprint minutiae: A comprehensive inventory and statistical analysis based on Brazilian data. Forensic Sci. Int. 364, 112233. http://dx.doi.org/10.1016/j.forsciint.2024.112233.

Haberman, S.J., 1988. A warning on the use of chi-squared statistics with frequency tables with small expected cell counts. J. Amer. Statist. Assoc. 83 (402), 555–560, URL http://www.jstor.org/stable/2288877.

Jager, L., Wellner, J.A., 2007. Goodness-of-fit tests via phi-divergences. Ann. Statist. 35 (5), 2018–2053. http://dx.doi.org/10.1214/0009053607000000244.

Lehmann, E., 1986. Testing Statistical Hypotheses, second ed. John Wiley & Sons, New York.

Maiste, P.J., Weir, B.S., 2004. Optimal testing strategies for large, sparse multinomial models. Comput. Statist. Data Anal. 46 (3), 605–620. http://dx.doi.org/10.1016/j.csda.2003.08.002.

Matsushita, R., Pianto, D., De Andrade, B.B., Cançado, A., Da Silva, S., 2019. The Touchard distribution. Commun. Stat. Theory Method 48, 2049–2059.

Morales, D., Pardo, L., Pardo, M.C., Vajda, I., 2004. Rényi statistics for testing composite hypotheses in general exponential models. Statistics 38 (2), 133–147. http://dx.doi.org/10.1080/02331880310001634647.

Morales, D., Pardo, L., Salicrú, M., Menéndez, M., 1994. Asymptotic properties of divergence statistics in a stratified random sampling and its applications to test statistical hypotheses. J. Statist. Plann. Inference 38 (2), 201–221. http://dx.doi.org/10.1016/0378-3758(94)90035-3.

Morales, D., Pardo, L., Vajda, I., 2000. Rényi statistics in directed families of exponential experiments. Statistics 34 (2), 151–174. http://dx.doi.org/10.1080/02331880008802324.

Mudholkar, G.S., Hutson, A.D., 1997. Continuity corrected approximations for and 'exact' inference with pearson's X2. J. Statist. Plann. Inference 59 (1), 61–78. http://dx.doi.org/10.1016/S0378-3758(96)00091-2.

Noughabi, H.A., Balakrishnan, N., 2016. Tests of goodness of fit based on phi-divergence. J. Appl. Stat. 43 (3), 412–429. http://dx.doi.org/10.1080/02664763.2015.1063116.

Pardo, L., 2005. Statistical inference based on divergence measures. In: Statistics: A Series of Textbooks and Monographs, Chapman & Hall/CRC, Boca Raton, FL, http://dx.doi.org/10.1201/9781420034813.

Rathie, P.N., 1970. On a generalized entropy and a coding theorem. J. Appl. Probab. 7 (1), 124–133.

Read, T.R., Cressie, N.A., 1988. Goodness-of-fit statistics for discrete multivariate data. In: Springer Series in Statistics, Springer-Verlag, New York, http://dx.doi.org/10.1007/978-1-4612-4578-0.

Salicru, M., Morales, D., Menendez, M., Pardo, L., 1994. On the applications of divergence type measures in testing statistical hypotheses. J. Multivariate Anal. 51 (2), 372–391. http://dx.doi.org/10.1006/jmva.1994.1068.

Vexler, A., Kim, Y.M., Yu, J., Lazar, N.A., Hutson, A.D., 2014. Computing critical values of exact tests by incorporating Monte Carlo simulations combined with statistical tables. Scand. J. Stat. 41 (4), 1013–1030. http://dx.doi.org/10.1111/sjos.12079.

Vonta, F., Mattheou, K., Karagrigoriou, A., 2012. On properties of the (Φ, α)-power divergence family with applications in goodness of fit tests. Methodol. Comput. Appl. Probab. 14, 335–356. http://dx.doi.org/10.1007/s11009-010-9205-8.

Yang, P., Chen, B., 2019. Robust Kullback-Leibler divergence and universal hypothesis testing for continuous distributions. IEEE Trans. Inform. Theory 65 (4), 2360–2373. http://dx.doi.org/10.1109/TIT.2018.2879057.