Asymptotic properties of the normalised discrete associated-kernel estimator for probability mass function

Youssef Esstafta, Célestin C. Kokonendji, and Sobom M. Somé

1. Introduction

The modern notion of a discrete associated kernel for smoothing or estimating discrete functions \( f \), defined on the discrete set \( \mathbb{T} \subseteq \mathbb{R} \), requires the development of new properties of convergences of the corresponding estimator. The support \( \mathbb{T} \) of \( f \) is not subject to any restrictive condition; it can be bounded, unbounded, finite or infinite. In this sense, Abdous and Kokonendji (2009) presented some asymptotic properties for unnormalised discrete associated-kernel estimators of a probability mass function (pmf). Several authors pointed out the use of a discrete associated kernel from Dirac and discrete triangular kernels (Kokonendji et al. 2007; Kokonendji and Zocchi 2010) and also from extensions of Dirac kernels proposed by Aitchison and Aitken (1976) for categorial data; see, e.g. Wang and Van Ryzin (1981), Li and Racine (2003), Chu et al. (2017), Racine et al. (2020). Furthermore, we have count kernels as the binomial of Kokonendji and Senga Kiessé (2011) and, recently, the CoM-Poisson of Huang et al. (2022) which are both underdispersed (i.e. variance less than mean). See also Senga Kiessé (2017) and Harfouch et al. (2018) for
other properties. Notice that one can use them to estimate, instead of the pmf, discrete regression or weighted functions; see, e.g. Senga Kiessé and Cuny (2014), Senga Kiessé and Ventura (2016) and, Kokonendji and Somé (2021).

Let us first fix the refined definition of discrete associated kernel from Kokonendji and Somé (2018) and state in Theorem 1.2 some important asymptotic properties to be completed in this paper.

**Definition 1.1:** Let $T \subseteq \mathbb{R}$ be the discrete support of the pmf $f$ to be estimated, $x \in T$ a target point and $h > 0$ a bandwidth. A parameterised pmf $K_{x,h} (\cdot)$ on the discrete support $S_x \subseteq \mathbb{R}$ is called ‘discrete associated kernel’ if the following conditions are satisfied:

$$x \in S_x, \quad \lim_{h \to 0} \mathbb{E} (Z_{x,h}) = x \quad \text{and} \quad \lim_{h \to 0} \text{Var} (Z_{x,h}) = \delta \in [0, 1),$$

where $Z_{x,h}$ denotes the discrete random variable with pmf $K_{x,h} (\cdot)$.

The choice of the discrete associated kernel referred to as a ‘second-order’ satisfying $\delta = 0$ in (1), ensures the convergence of its corresponding estimator; and, an elementary example is the naive or Dirac kernel for smoothing a very large sample of discrete data by the empirical estimator. Otherwise, the convergence of its corresponding estimator is not guaranteed; that is for $\delta = \delta(x) \in (0, 1)$ in Definition 1.1, the discrete associated kernel is called of ‘first-order’ like the well-known binomial kernel. It is already shown, via the total mean squared error, that the key interest of discrete smoothing through some count associated-kernel estimators rather than the natural empirical one is for small and moderate sample sizes; see, e.g. Kokonendji and Senga Kiessé (2011, figs. 2 and 3) and Kokonendji and Varron (2016, fig. 1). Thus, any count associated-kernel estimator of pmf must have similar consistency with the Dirac one.

Let $X_1, X_2, \ldots, X_n$ be a sample of independent and identically distributed (i.i.d.) discrete random variables having a nonnegative pmf $f$ on $T \subseteq \mathbb{R}$. In general, the basic estimator of $f$ is not a pmf. Indeed, for some discrete associated-kernels (e.g. binomial, triangular and CoM-Poisson), the total mass of the corresponding estimator is not equal to 1. This limit is explained by the fact that the normalising variable (which is equal to the sum over all the targets belonging to $T$ of the discrete associated-kernel estimator) was assumed to be equal to 1 only to simplify the calculations. More precisely, one can write both estimators as:

$$\hat{f}_n(x) = \frac{\tilde{f}_n(x)}{C_n}, \quad x \in T,$$

with

$$\tilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h_n} (X_i) \quad \text{and} \quad C_n = \sum_{x \in T} \tilde{f}_n(x) > 0,$$

where $(h_n)_{n \geq 1}$ is an arbitrary sequence of positive smoothing parameters that satisfies $\lim_{n \to \infty} h_n = 0$, while $K_{x,h_n} (\cdot)$ is a suitably chosen discrete smoothing kernel function. If $C_n = 1$ as for the kernels of Dirac, Aitchison and Aitken (1976) and Wang and Van Ryzin (1981), one obviously has $\tilde{f}_n = \hat{f}_n$. Hence, more generally than Theorem 2.2 of Ouyang et al. (2006), one has:
Theorem 1.2 (Abdous and Kokonendji 2009): For any \( x \in \mathbb{T} \) and under Assumptions (1) of the second-order (i.e. \( \delta = 0 \)), one has
\[
\tilde{f}_n(x) \xrightarrow{L^2, a.s.} \frac{f(x)}{n \to \infty}
\]
where \( \xrightarrow{L^2, a.s.} \) stands for both ‘mean square and almost surely convergences’. Furthermore, if \( f(x) > 0 \) then
\[
\left( \tilde{f}_n(x) - E\tilde{f}_n(x) \right) \left( \text{Var}\tilde{f}_n(x) \right)^{-1/2} \xrightarrow{D} N(0, 1),
\]
where \( \xrightarrow{D} \) stands for ‘convergence in distribution’ and \( N(0, 1) \) denotes the standard normal distribution.

In this paper we mainly extend Theorem 1.2 of the non-normalised estimator \( \tilde{f}_n \) of (3) to the normalised one \( \hat{f}_n \) of (2), introducing new and non-restrictive assumptions with uniformities on the target point in the limits of (1) and, therefore, changing the types of convergences. As a matter of fact and more importantly, we shall demonstrate the convergence in mean square of the positive normalising random variable \( C_n \) of (3) to 1; which clearly completes the similar result in Kokonendji and Varron (2016, th. 2.1). The following Section 2 states different assumptions and their corresponding results with illustrations on the recent CoM-Poisson kernel estimator. The case of the first-order binomial kernel is briefly discussed with great interest of the practical consistency of its normalised estimator which is pointed out in Section 3 of numerical results. Finally, Section 4 is devoted to the detailed proofs.

2. Results and illustrations

In order to obtain some soft convergences of the pointwise normalised estimator \( \hat{f}_n(x) \) at \( x \), we need quite strong assumptions instead of the most popular (1). In this way, we do not use concentration inequalities as in Kokonendji and Varron (2016) as well as Abdous and Kokonendji (2009) through, for instance, an inequality of Hoeffding (1963).

The first set of assumptions is uniformly in the target \( x \) and it is satisfied, in our knowledge, by all discrete kernels of Definition 1.1 with \( \delta = 0 \):

\[
(A1) : x \in S_X, \lim_{n \to \infty} \sup_{x \in \mathbb{T}} \left| E(Z_{x,h_n}) - x \right| = 0 \quad \text{and} \quad \lim_{n \to \infty} \sup_{x \in \mathbb{T}} \text{Var}(Z_{x,h_n}) = 0.
\]

Hence, the following proposition provides a key point to establish the next result on the pointwise probability convergence of \( \hat{f}_n(x) \) defined in (2).

Proposition 2.1: Under Assumptions (A1), the normalising random variable \( C_n \) converges in mean square to 1.

Theorem 2.2 (Consistency): Under (A1) and for any \( x \in \mathbb{T} \), we have:
\[
\hat{f}_n(x) \xrightarrow{P} f(x),
\]
where ‘\( \xrightarrow{P} \)’ stands for ‘convergence in probability’.
On a finite discrete set \( T \), the pointwise and uniform convergences of a sequence of functions are equivalent; hence, the previous results are guaranteed when the discrete associated kernel satisfies the common set of hypotheses (1) with \( \delta = 0 \) (see Section 4 for further details).

**Corollary 2.3 (Uniform consistency):** Suppose that the set \( T \) is finite. Under Conditions (1) with \( \delta = 0 \), one has

\[
\sup_{x \in T} \left| \hat{f}_n(x) - f(x) \right| \xrightarrow{P} 0.
\]

For making a global comparison between \( \hat{f}_n \) and \( \tilde{f}_n \), we use the absolute \( L^1 \) error. The following proposition highlights the fact that our estimator outperforms, in the sense of the \( L^1 \) error, the unnormalised one. In the continuous framework of estimating a density function, similar conclusions have been reached; see, for example, Devroye and Lugosi (2001, sec. 5.6) and, also, Glad et al. (2003) for \( L^2 \) criterion.

**Proposition 2.4:** Let Assumptions (A1) be satisfied. For any \( \epsilon > 0 \), there exists \( N_\epsilon \in \mathbb{N} \) such that for any \( n \geq N_\epsilon \), we have

\[
\mathbb{E} \left[ \sum_{x \in T} \left| \hat{f}_n(x) - f(x) \right| \right] < \mathbb{E} \left[ \sum_{x \in T} \tilde{f}_n(x) - f(x) \right] + \epsilon.
\]

Regarding to a refined result of the asymptotic normality of \( \hat{f}_n(x) \), it is necessary to quantify the speed of convergence to 0 of \( \sup_{x \in T} \) \( \text{Var}(Z_{x,h_n}) \) and \( \sup_{x \in T} \left| \mathbb{E}[Z_{x,h_n}] - x \right| \) in (A1). We therefore assume that these two sequences satisfy the following second set of conditions:

\[
(A2) \ : \ x \in \mathbb{S}_x, \ \sup_{x \in T} \left| \mathbb{E}(Z_{x,h_n}) - x \right| = \mathcal{O}(h_n) \quad \text{and} \quad \sup_{x \in T} \text{Var}(Z_{x,h_n}) = \mathcal{O}(h_n).
\]

The previous Assumptions (A2) and also (A1) are verified by all the usual kernels of second-order introduced as examples in Section 1.

**Theorem 2.5 (Asymptotic normality):** Let (A2) be satisfied. If the sequence \( (h_n)_{n \geq 1} \) is chosen such that \( \sqrt{n} h_n \to 0 \) as \( n \to \infty \), then, for any \( x \in T \) such that \( f(x) > 0 \), the sequence \( \{\sqrt{n}(\hat{f}_n(x) - f(x))\}_{n \geq 1} \) has a limiting centered normal distribution with variance \( f(x)\{1 - f(x)\} \).

To conclude this section, we highlight some of our previous results on the recent CoM-Poisson kernel estimator of Huang et al. (2022) and compare with the classical binomial one. In fact, we consider the refined version of the CoM-Poisson kernel satisfying (A1) and
Proposition 2.6: Poisson distribution with location (or mean) parameter \( \mu \) for behaviour (5) of this CoM-Poisson kernel which is of the second-order and underdispersed. The following proposition points out the mean and the main key of the variance dispersion with respect to the standard Poisson model. Also demonstrating in the Appendix, sec. 2.2) for more details on the original form, asymptotic properties and the relative dispersion with respect to the standard Poisson model. Also demonstrating in the Appendix, the following proposition points out the mean and the main key of the variance behaviour (5) of this CoM-Poisson kernel which is of the second-order and underdispersed for \( h \in (0, 1) \).

\[ (A2) \text{ as follows: } \mathbb{T} = \mathbb{N} = \mathbb{S}_x \text{ for each } x \in \mathbb{N} \text{ and any } h > 0, \]

\[ K_{x,h}^{\text{CMP}}(z) = \frac{[\lambda(x, 1/h)]^z}{(z!)^{1/h}} \cdot [D(\lambda(x, 1/h), 1/h)]^{-1}, \]

where \( D(\lambda(x, 1/h), 1/h) = \sum_{z=0}^{\infty} [\lambda(x, 1/h)]^z / (z!)^{1/h} \) is the normalising constant and \( \lambda := \lambda(x, 1/h) \) represents a function of \( x \) and \( 1/h \) given by the solution of

\[ \sum_{z=0}^{\infty} \frac{[\lambda(x, 1/h)]^z}{(z!)^{1/h}} (z - x) = 0. \quad (4) \]

This construction implies that \( \mathbb{E}(Z_{x,h}^{\text{CMP}}) = x \) and

\[ \text{Var} \left( Z_{x,h}^{\text{CMP}} \right) = h \{\lambda(x, 1/h)\}^h + \mathcal{O} \left( \{\lambda(x, 1/h)\}^{-h} \right) \text{ as } h \rightarrow 0. \quad (5) \]

Indeed, Huang (2017) proposed the parametrization via the mean of the original CoM-Poisson (Conway-Maxwell-Poisson or CMP) distribution. One can refer to Shmueli et al. (2005), Kokonendji et al. (2008, sec. 4.2), Gaunt et al. (2019) and Toledo et al. (2022, sec. 2.2) for more details on the original form, asymptotic properties and the relative dispersion with respect to the standard Poisson model. Also demonstrating in the Appendix, the following proposition points out the mean and the main key of the variance behaviour (5) of this CoM-Poisson kernel which is of the second-order and underdispersed for \( h \in (0, 1) \).

Proposition 2.6: Let \( Y \) be a count random variable following the mean-parametrized CoM-Poisson distribution with location (or mean) parameter \( \mu \geq 0 \) and dispersion parameter \( \nu > 0 \) such that its pmf \( p(\cdot; \mu, \nu) \) is defined by

\[ p(y; \mu, \nu) := K_{\mu,1/h}^{\text{CMP}}(y), \quad y \in \mathbb{N}. \quad (6) \]

Then \( \mathbb{E}(Y) = \mu \) and, when \( \{\lambda(\mu, \nu)\}^{1/\nu} \rightarrow \infty \) as \( \nu \rightarrow \infty \), the variance of \( Y \) verifies

\[ \text{Var} \left( Y \right) = \frac{1}{\nu} \{\lambda(\mu, \nu)\}^{1/\nu} + \mathcal{O} \left( \{\lambda(\mu, \nu)\}^{-1/\nu} \right). \]

As for the binomial kernel of first-order and underdispersed of Kokonendji and Senga Kiessé (2011), one has: \( \mathbb{T} = \mathbb{N}, \mathbb{S}_x = \{0, 1, \ldots, x + 1\} \) for each \( x \in \mathbb{N} \) and \( h \in (0, 1) \),

\[ K_{x,h}^B(z) = \frac{(x + 1)!}{z!(x + 1 - z)!} \left( \frac{x + h}{x + 1} \right)^z \left( \frac{1 - h}{x + 1} \right)^{x + 1 - z} \]

with \( \mathbb{E}(Z_{x,h}^B) = x + h \rightarrow x \) as \( h \rightarrow 0 \) and

\[ \text{Var} \left( Z_{x,h}^B \right) = \frac{(x + h)(1 - h)}{x + 1}. \quad (7) \]

From Assumptions (1) and through Equation (7), one here has \( \delta = \delta(x) = x/(x + 1) \in [0, 1] \) which does not clearly satisfy the last condition of \( (A1) \) as well as for \( (A2) \). Notice that \( K_{x,h}^B(\cdot) \) is the pmf of the standard binomial distribution \( \mathcal{B}(n, p) \) with \( n = x + 1 \) and \( p := (x + h)/(x + 1) \). Nevertheless, we always use the binomial kernel for smoothing count data of small and moderate sample sizes.
3. Simulations and an application

All numerical studies are here performed using the classical binomial and the recent CoM-Poisson kernel smoothers with the aim to corroborate the previous theoretical results. Computations have been run on PC 2.30 GHz by using the R software R Core Team (2021). Both previous estimators are fitted using the Ake package by Wansouwé et al. (2016) and the mpcmp one of Fung et al. (2020), respectively. We evaluate the performances of these two discrete associated-kernel estimators with cross-validation choices of the optimal bandwidth parameter; see, for example, Kokonendji and Somé (2021) for Bayesian local bandwidths. In fact, the optimal bandwidth \( h_{cv} \) of \( h \) using the cross-validation method is obtained through

\[
h_{cv} = \arg\min_{h > 0} \left[ \sum_{x \in \mathcal{T}} \left\{ \hat{f}_n(x) \right\}^2 - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{n,h-i}(X_i) \right],
\]

where

\[
\hat{f}_{n,h-i}(X_i) = \frac{1}{n-1} \sum_{\ell=1, \ell \neq i}^{n} K_{X_i,h}(X_{\ell})
\]
is being computed as \( \hat{f}_n(X_i) \) without the observation \( X_i \).

We here consider four scenarios which are denoted by A, B, C and D to simulate count datasets with respect to the support of both discrete kernels. These scenarios have been considered to evaluate the performances of both smoothers to deal with zero-inflated, unimodal and multimodal distributions. We shall examine the efficiency of both smoothers via the empirical estimates of \( \hat{C}_n \) and \( \hat{ISE}_n \) of the integrated squared errors (ISE):

\[
\hat{ISE}_n := \frac{1}{N_{sim}} \sum_{t=1}^{N_{sim}} \sum_{x \in \mathcal{T}} \left( \hat{f}_n(x) - f(x) \right)^2 \quad \text{and} \quad \hat{C}_n := \frac{1}{N_{sim}} \sum_{t=1}^{N_{sim}} \sum_{x \in \mathcal{T}} \tilde{f}_n(x),
\]

where \( N_{sim} \) is the number of replications and \( n \) corresponds to the sample size which shall be small, medium and large.

- **Scenario A** is generated by using the Poisson distribution

\[
f_A(x) = \frac{8^x e^{-8}}{x!}, \quad x \in \mathbb{N};
\]

- **Scenario B** comes from the zero-inflated Poisson distribution

\[
f_B(x) = \left( \frac{7}{10} 1_{x=0} \right) + \left( \frac{3}{10} \times \frac{10^x e^{-10}}{x!} \right), \quad x \in \mathbb{N};
\]

- **Scenario C** is from a mixture of two Poisson distributions

\[
f_C(x) = \left( \frac{2}{5} \times \frac{0.5^x e^{-0.5}}{x!} \right) + \left( \frac{3}{5} \times \frac{8^x e^{-8}}{x!} \right), \quad x \in \mathbb{N};
\]
• Scenario D comes from a mixture of three Poisson distributions

\[ f_D(x) = \left( \frac{3}{5} \times \frac{10xe^{-10}}{x!} \right) + \left( \frac{1}{5} \times \frac{22xe^{-22}}{x!} \right) + \left( \frac{1}{5} \times \frac{50xe^{-50}}{x!} \right), \quad x \in \mathbb{N}. \]

Table 1 reports some empirical mean values of \( \hat{C}_n \) and \( \hat{ISE}_n \) with their standard deviations using \( N_{sim} = 500 \) replications from Scenarios A, B, C and D to the corresponding sample sizes \( n = 10, 25, 50, 100, 250, 500 \). For each given subsample and the discrete associated-kernel CoM-Poisson or binomial, we have to compute the related bandwidth \( h_{cv} \) through the cross-validation method before \( \tilde{f}_n, \hat{C}_n, \hat{f}_n \) and finally \( \hat{ISE}_n \). Hence, we observe the following behaviours. First, when the sample size \( n \) increases then all standard deviations of Table 1 steadily decrease towards 0. The normalising constant \( \hat{C}^{CM}_n \) for CoM-Poisson kernel estimator also becomes more and more precise to 1 in absolute value; while the binomial one \( \hat{C}^{B}_n \) moves further away from 1 in absolute value for medium and large sample sizes, in particular for both zero-inflated Scenarios B and C. Next and as expected, the consistent CoM-Poisson smoother is increasingly accurate as sample size increases according to the \( \hat{ISE}^{CM}_n \) criterion. It is seemingly better than the binomial one \( \hat{ISE}^{B}_n \), especially for small and moderate sample sizes \( n \leq 100 \). With enormous surprise and satisfaction, the normalised binomial kernel smoother is also asymptotically consistent in practice, similar to the CoM-Poisson one for all used Scenarios. In fact, this normalising process of \( \hat{f}_n \) by \( C_n \) for obtaining \( \hat{f}_n \) apparently controls the consistency property of \( \hat{f}_n \), even for the discrete first-order associated-kernel not verifying (A1). Finally, we can notify the

| \( n \) | \( \hat{C}^{B}_n \) (0.00066) | \( \hat{C}^{CM}_n \) (0.00059) | \( \hat{ISE}^{B}_n \) (0.00043) | \( \hat{ISE}^{CM}_n \) (0.00022) |
|---|---|---|---|---|
| A 10 | 0.98561 (0.000419) | 0.99860 (0.000234) | 0.00107 (0.000059) | 0.00024 (0.000023) |
| 25 | 0.99309 (0.000173) | 0.99669 (0.000082) | 0.00036 (0.000015) | 0.00008 (0.000004) |
| 50 | 0.99957 (0.000044) | 0.99924 (0.000020) | 0.00004 (0.000002) | 0.00001 (0.000001) |
| 100 | 0.99971 (0.000004) | 1.00000 (0.000002) | 0.00000 (0.000000) | 0.00000 (0.000000) |
| 250 | 0.98688 (0.000007) | 0.99978 (0.000001) | 0.00001 (0.000000) | 0.00000 (0.000000) |
| 500 | 1.00121 (0.000004) | 1.00019 (0.000002) | 0.00000 (0.000000) | 0.00000 (0.000000) |
| B 10 | 0.98129 (0.000188) | 0.95610 (0.000052) | 0.00337 (0.000164) | 0.02605 (0.002615) |
| 25 | 0.99962 (0.000210) | 0.98248 (0.000263) | 0.01323 (0.000759) | 0.01058 (0.000840) |
| 50 | 1.00806 (0.000082) | 1.00096 (0.000007) | 0.00644 (0.000030) | 0.00570 (0.000343) |
| 100 | 1.01201 (0.000549) | 0.99948 (0.000137) | 0.00316 (0.000185) | 0.00338 (0.000209) |
| 250 | 1.01251 (0.000472) | 0.99991 (0.000108) | 0.00054 (0.000083) | 0.00062 (0.000043) |
| 500 | 1.01314 (0.000212) | 0.99974 (0.000051) | 0.00005 (0.000031) | 0.00006 (0.000019) |
| C 10 | 0.91753 (0.000705) | 1.01526 (0.004263) | 0.03427 (0.02122) | 0.02339 (0.00277) |
| 25 | 0.96054 (0.004003) | 1.03188 (0.02678) | 0.01026 (0.00623) | 0.01052 (0.00864) |
| 50 | 0.98841 (0.002547) | 1.03353 (0.01374) | 0.00445 (0.00276) | 0.00509 (0.00428) |
| 100 | 1.00816 (0.001066) | 1.02105 (0.01240) | 0.00275 (0.00139) | 0.00304 (0.00212) |
| 250 | 1.03003 (0.00945) | 1.00878 (0.00851) | 0.00108 (0.00087) | 0.00084 (0.00054) |
| 500 | 1.03054 (0.000807) | 1.00887 (0.00286) | 0.00084 (0.00051) | 0.00048 (0.00022) |
| D 10 | 0.99111 (0.000312) | 0.94899 (0.01556) | 0.02453 (0.00981) | 0.01071 (0.00201) |
| 25 | 1.00151 (0.000901) | 0.98467 (0.01566) | 0.01303 (0.00794) | 0.00980 (0.00962) |
| 50 | 0.99742 (0.000764) | 0.99649 (0.00458) | 0.00289 (0.00112) | 0.00216 (0.00187) |
| 100 | 0.99941 (0.000655) | 0.99748 (0.00195) | 0.00106 (0.00063) | 0.00118 (0.00076) |
| 250 | 1.02035 (0.00034) | 0.99956 (0.00103) | 0.00067 (0.00033) | 0.00044 (0.00037) |
| 500 | 1.01108 (0.00021) | 1.00031 (0.00028) | 0.00027 (0.00014) | 0.00037 (0.00020) |
Figure 1. Empirical distributions of $\sqrt{n}(\hat{f}_{n}(x) - f_{A}(x))$ at $x = 6$ over $N_{\text{sim}} = 500$ independent simulations with $n = 500$ and $h_{n} = (\sqrt{n} \log(n))^{-1}$ using CoM-Poisson (left) and binomial (right) kernels. The smoothed kernel density is displayed in full line, and the centered Gaussian density with the same variance is plotted in dotted line.

importance of normalisation of the discrete associated-kernel estimators of pmfs in practice; see, for example, Wansouwé et al. (2016) and Kokonendji and Somé (2021) for some illustrations in uni- and multivariate cases.

Figure 1 illustrates both empirical distributions of $\sqrt{n}(\hat{f}_{\text{CMP}}^{(6)} - f_{A}(6))$ (left) and $\sqrt{n}(\hat{f}_{B}^{(6)} - f_{A}(6))$ (right) over $N_{\text{sim}} = 500$ replications of Scenario A with the sample size $n = 500$ and the bandwidth $h_{n} = (\sqrt{n} \log(n))^{-1}$ which is chosen here to satisfy the conditions of Theorem 2.5. It is obviously remarkable that the normalised CoM-Poisson kernel estimator is more suitable than the one computed from a binomial kernel, for which the bias increases considerably with the sample size. Once again, these figures confirm the pointwise consistency as well as the pointwise asymptotic normality of the normalised CoM-Poisson kernel estimator, unlike the normalised discrete associated-kernel estimator obtained from the binomial kernel which does not verify our set of hypotheses.

Concerning an application on real data for pointing out the very competitive CoM-Poisson kernel, both discrete kernel estimators are finally used to smooth a count dataset on development days of insect pests on Hura trees with moderate sample size $n = 51$; see Senga Kiessé (2017) and also Huang et al. (2022) for applications using these two discrete associated-kernel estimators among others. Practical performances are here examined via the cross-validation method and the empirical criterion of $\text{ISE}: \hat{\text{ISE}}_{0} := \sum_{x \in \mathbb{T} \subseteq \mathbb{N}} (\hat{f}_{n}(x) - f_{0}(x))^{2}$, where $f_{0}(\cdot)$ is the empirical or naive estimator. The CoM-Poisson kernel appears to be the best with $h_{\text{CMP}}^{CV} = 0.01865$, $\hat{C}_{n}^{\text{CMP}} = 0.99994$ and $\hat{\text{ISE}}_{0}^{\text{CMP}} = 0.00967$ followed by the binomial smoother with $h_{\text{B}}^{CV} = 0.00601$, $\hat{C}_{n}^{\text{B}} = 0.99913$ and finally $\hat{\text{ISE}}_{0}^{\text{B}} = 0.01232$; see Figure 2 for graphical representations. Notice that, for the same dataset, Senga Kiessé (2017) produced $h_{\text{B}}^{CV} = 0.02$ and $\hat{\text{ISE}}_{0}^{\text{B}} = 0.0104$ for the non-normalised binomial estimation $\hat{f}_{n}^{\text{B}}$. While Huang et al. (2022) only presented $h_{\text{CMP}}^{CV} = 0.0251$ without $\hat{\text{ISE}}_{0}^{\text{CMP}}$ for the non-normalised CoM-Poisson estimation $\hat{f}_{n}^{\text{CMP}}$ with precisions to fit a non-zero probability outside of the observed range and also to preserve the sample mean of the dataset.

We also conducted a robustness test on the same insect pest data with a train and test sample sizes of $n = 43$ and 8, respectively. The CoM-Poisson smoother remains the best one with $h_{\text{CMP}}^{CV} = 0.019$, $\hat{C}_{n}^{\text{CMP}} = 0.99994$, $\hat{\text{ISE}}_{0,\text{train}}^{\text{CMP}} = 0.00599$ and $\hat{\text{ISE}}_{0,\text{test}}^{\text{CMP}} = $
Figure 2. Empirical frequency with its corresponding binomial and CoM-Poisson kernel smoothers of count dataset of insect pests on Hura trees with \( n = 51 \); see, for example, Senga Kiessé (2017).

0.00062. The binomial kernel gives \( h_{cv}^B = 0.006, \hat{C}^B_n = 0.99913, \hat{I}\text{SE}^B_{0,\text{train}} = 0.01153 \) and \( \hat{I}\text{SE}^B_{0,\text{test}} = 0.00079 \).

4. Proofs of results

Proof of Proposition 2.1.: First, one easily has the following decomposition:

\[
\mathbb{E} \left[ |C_n - 1|^2 \right] = \text{Var}(C_n) + (\mathbb{E}[C_n] - 1)^2. \tag{8}
\]

We use Equation (3) and the fact that the \( X_i \)'s are i.i.d. to obtain

\[
\text{Var}(C_n) = \sum_{x \in T} \sum_{y \in T} \text{Cov} \left( \hat{f}_n(x), \hat{f}_n(y) \right)
\]

\[
= \frac{1}{n^2} \sum_{x \in T} \sum_{y \in T} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov} \left( K_{x,h_n}(X_i), K_{y,h_n}(X_j) \right)
\]

\[
= \frac{1}{n^2} \sum_{x \in T} \sum_{y \in T} \sum_{i=1}^{n} \text{Cov} \left( K_{x,h_n}(X_1), K_{y,h_n}(X_i) \right)
\]

\[
= \frac{1}{n} \sum_{x \in T} \sum_{y \in T} \text{Cov} \left( K_{x,h_n}(X_1), K_{y,h_n}(X_1) \right)
\]

\[
= \frac{1}{n} \sum_{x \in T} \text{Var} \left[ K_{x,h_n}(X_1) \right] + \frac{1}{n} \sum_{x \in T} \sum_{y \notin \{x\}} \text{Cov} \left( K_{x,h_n}(X_1), K_{y,h_n}(X_1) \right). \tag{9}
\]
The bias term in (8) can be explicitly rewritten as:

\[
\mathbb{E}[C_n] - 1 = \sum_{x \in \mathbb{T}} \mathbb{E}[\tilde{f}_n(x)] - \sum_{x \in \mathbb{T}} f(x)
\]

\[
= \sum_{x \in \mathbb{T}} \sum_{z \in \mathbb{T} \cap \mathbb{S}_x} K_{x, h_n}(z) f(z) - \sum_{x \in \mathbb{T}} \sum_{z \in \mathbb{S}_x} K_{x, h_n}(z) f(x)
\]

\[
= E_{1,n} - E_{2,n},
\]

with

\[
E_{1,n} = \sum_{x \in \mathbb{T}} \sum_{z \in \mathbb{T} \cap \mathbb{S}_x} (f(z) - f(x)) K_{x, h_n}(z)
\]

and

\[
E_{2,n} = \sum_{x \in \mathbb{T}} f(x) \sum_{z \in \mathbb{T} \cap \mathbb{S}_x} K_{x, h_n}(z),
\]

where \( \mathbb{T} \) is the set \( \mathbb{R} \setminus \mathbb{T} \).

Secondly, to make the proof more readable, we divide it in two steps according to the convergences to 0 of both variance and bias terms in Equation (8).

\* Step 1: Convergence to 0 of the variance term in Equation (8).

Under (A1), one can prove that the first term in the right hand side of Equation (9) converges to 0. As a matter of fact, observe first that

\[
\text{Var} \left[ K_{x, h_n}(X_1) \right] - f(x) \left\{ 1 - f(x) \right\} = \left\{ \sum_{z \in \mathbb{T} \cap \mathbb{S}_x} \left( K_{x, h_n}(z) \right)^2 f(z) - f(x) \right\}
\]

\[
+ f(x)^2 - \left( \mathbb{E}[\tilde{f}_n(x)] \right)^2
\]

\[
= : F_{1,n}(x) + F_{2,n}(x). \quad (11)
\]

The sets \( \mathbb{T} \) and \( \mathbb{S}_x \) are discrete. So, one can find a finite constant \( \alpha > 0 \) (which does not depend on \( x \)) such that \( |z - x| \geq \alpha \) for any \( z \) in \( \mathbb{T} \cup \mathbb{S}_x \setminus \{x\} \). Hence, the use of Markov’s inequality and Assumptions (A1) lead to deduce that the first sequence \( (F_{1,n})_{n \geq 1} \) in Equation (11) converges uniformly on \( \mathbb{T} \) to 0 as follows:

\[
\sup_{x \in \mathbb{T}} |F_{1,n}(x)| \leq \sup_{x \in \mathbb{T}} \left\{ f(x) \left( 1 - K_{x, h_n}(x) \right) K_{x, h_n}(x) \right\}
\]

\[
+ \sum_{z \in \mathbb{T} \cap \mathbb{S}_x \setminus \{x\}} |f(z)K_{x, h_n}(z) - f(x)| K_{x, h_n}(z) + f(x) \sum_{z \in \mathbb{T} \cap \mathbb{S}_x} K_{x, h_n}(z)
\]

\[
\leq 4 \sup_{x \in \mathbb{T}} \left\{ 1 - K_{x, h_n}(x) \right\}
\]

\[
\leq 4 \sup_{x \in \mathbb{T}} \mathbb{P}\left( |Z_{x, h_n} - x| \geq \alpha \right)
\]
\[
\leq \frac{4}{\alpha^2} \left\{ \sup_{x \in T} \text{Var} \left( Z_{x,h^n} \right) + \left( \sup_{x \in T} |E \left[ Z_{x,h^n} \right] - x| \right)^2 \right\} \to 0, \quad \text{as } n \to \infty.
\]

Similarly, we show that the second sequence \((F_{2,n})_{n \geq 1}\) in Equation (11) also converges uniformly on \(T\) to 0. Let us be more precise. Note that

\[
|F_{2,n}(x)| = f(x) |E \left[ \tilde{f}_n(x) \right] - f(x)| + E \left[ \tilde{f}_n(x) \right] |E \left[ \tilde{f}_n(x) \right] - f(x)|
\leq 2 \left| E \left[ \tilde{f}_n(x) \right] - f(x) \right| \leq 2 \left\{ \sum_{z \in \mathbb{T} \cap S_x \setminus \{x\}} K_{x,h^n}(z) \left| f(z) - f(x) \right| + f(x) \sum_{z \in \mathbb{T} \cap S_x} K_{x,h^n}(z) \right\}
\leq 6 \mathbb{P} \left( Z_{x,h^n} \neq x \right),
\]

where we have used the fact that if \(z \in \mathbb{T} \cap S_x\), then necessarily \(z \neq x\). Arguing as before, we obtain the expected uniform convergence of \((F_{2,n})_{n \geq 1}\) to 0.

Consequently, from Equation (11) the sequence \{\text{Var}(K_{h^n}(X_1))\}_{n \geq 1}\) converges uniformly on \(T\) to \(\ell_1 : x \in \mathbb{T} \mapsto f(x)\{1 - f(x)\} \in \mathbb{R}\). Finally, there exists \(N \in \mathbb{N}^*\) such that for any \(n \geq N\), we have

\[
\frac{1}{n} \sum_{x \in \mathbb{T}} \text{Var} \left[ K_{x,h^n}(X_1) \right] \leq 2 \frac{1}{n} \sum_{x \in \mathbb{T}} f(x)\{1 - f(x)\} \leq \frac{2}{n},
\]

and the enacted convergence is obtained.

We now deal with the second term on the right hand side of Equation (9). Using the definition of the non-normalised associated-kernel estimator \(\tilde{f}_n(\cdot)\) introduced in Equation (3), one can write that for any \(x, y \in \mathbb{T}\) and all \(n \geq 1\),

\[
\text{Cov} \left( K_{x,h^n}(X_1), K_{y,h^n}(X_1) \right) = \sum_{z \in \mathbb{T} \cap S_x \cap S_y} K_{x,h^n}(z)K_{y,h^n}(z)f(z)
\leq \left( \sum_{z \in \mathbb{T} \cap S_x} K_{x,h^n}(z)f(z) \right) \left( \sum_{z \in \mathbb{T} \cap S_y} K_{y,h^n}(z)f(z) \right)
= \sum_{z \in \mathbb{T} \cap S_x \cap S_y} K_{x,h^n}(z)K_{y,h^n}(z)f(z) - E \left[ \tilde{f}_n(x) \right] E \left[ \tilde{f}_n(y) \right].
\]

It then follows that

\[
\text{Cov} \left( K_{x,h^n}(X_1), K_{y,h^n}(X_1) \right) + f(x)f(y) = \sum_{z \in \mathbb{T} \cap S_x \cap S_y} K_{x,h^n}(z)K_{y,h^n}(z)f(z)
\leq \left( E \left[ \tilde{f}_n(x) \right] - f(x) \right) E \left[ \tilde{f}_n(y) \right]
\leq \left( E \left[ \tilde{f}_n(y) \right] - f(y) \right) f(x).
\]
Thus, one has

$$\sup_{(x,y) \in \mathbb{T}^2 \atop x \neq y} \left| \text{Cov} \left( K_{x,h_n}(X_1), K_{y,h_n}(X_1) \right) + f(x)f(y) \right| \leq G_{1,n} + 2G_{2,n},$$

with

$$G_{1,n} = \sup_{(x,y) \in \mathbb{T}^2 \atop x \neq y} \sum_{z \in \mathbb{T} \cap S_x \cap S_y} K_{x,h_n}(z)K_{y,h_n}(z)f(z) \quad \text{and} \quad G_{2,n} = \sup_{x \in \mathbb{T}} \left| E \left[ \tilde{f}_n(x) \right] - f(x) \right|.$$ 

Following the same arguments used to prove the convergence of $F_{2,n}(x)$ to 0, we show that $(G_{2,n})_{n \geq 1}$ converges to 0. Indeed, observe that

$$\sum_{z \in \mathbb{T} \cap S_x \cap S_y} K_{x,h_n}(z)K_{y,h_n}(z)f(z) = K_{x,h_n}(x)K_{y,h_n}(x)f(x)$$

$$+ \sum_{z \in \mathbb{T} \cap S_x \cap S_y \atop z \neq x} K_{x,h_n}(z)K_{y,h_n}(z)f(z)$$

$$\leq \sum_{z \in S_y \atop z \neq y} K_{y,h_n}(z) + \sum_{z \in S_x \atop z \neq x} K_{x,h_n}(z)$$

$$= \mathbb{P} \left( Z_{y,h_n} \neq y \right) + \mathbb{P} \left( Z_{x,h_n} \neq x \right).$$

Hence, the same lines of proof given before can be reproduced to show that $(G_{1,n})_{n \geq 1}$ converges to 0. The sequence $\{ \text{Cov}(K_{x,h_n}(X_1), K_{y,h_n}(X_1)) \}_{n \geq 1}$ is uniformly convergent on the set $\Delta = \{(x, y) \in \mathbb{T}^2, x \neq y \}$ with limit $\ell_2 : (x, y) \in \Delta \mapsto -f(x)f(y) \in \mathbb{R}$. It can therefore be easily demonstrated that there exists a positive integer $N$ such that for any $n \geq N$, one has

$$\frac{1}{n} \sum_{x \in \mathbb{T}} \sum_{y \in \mathbb{T} \cap \{x\}} \text{Cov} \left( K_{x,h_n}(X_1), K_{y,h_n}(X_1) \right) \leq \frac{1}{n} \sum_{x \in \mathbb{T}} f(x) \sum_{y \in \mathbb{T} \cap \{x\}} f(y) \leq \frac{1}{n}.$$

This completes the proof of the convergence to 0 of the second term on the right hand side of Equation (9) and, finally, the proof of the convergence to 0 of the variance term in Equation (8).

\(\diamond\) Step 2: Convergence to 0 of the bias term in Equation (8).

The sequence $(E_{2,n})_{n \geq 1}$ introduced in Equation (10) clearly converges to 0 since

$$E_{2,n} \leq \sup_{x \in \mathbb{T}} \mathbb{P} \left( Z_{x,h_n} \neq x \right) \to 0, \quad \text{as} \quad n \to \infty.$$

We now use Equation (3) and the same arguments developed in the previous step to obtain that

$$\left| E_{1,n} \right| \leq \sum_{x \in \mathbb{T}} \left| \left( \sum_{z \in \mathbb{T} \cap S_x} f(z)K_{x,h_n}(z) \right) - f(x) + f(x) \left\{ 1 - K_{x,h_n}(x) \right\} \right|$$

$$+ \sum_{x \in \mathbb{T}} f(x) \sum_{z \in \mathbb{T} \cap S_x \atop z \neq x} K_{x,h_n}(z)$$
≤ \sum_{x \in T} |\mathbb{E}[\tilde{f}_n(x) - f(x)]| + 2 \sup_{x \in \mathbb{T}} \mathbb{P}(Z_{x,h_n} \neq x)

≤ \sup_{x \in \mathbb{T}} \frac{\mathbb{E}[\tilde{f}_n(x) - f(x)]}{f(x)} + 2 \sup_{x \in \mathbb{T}} \mathbb{P}(Z_{x,h_n} \neq x) \to 0, \quad \text{as} \ n \to \infty.

Consequently, the bias term in Equation (8) converges to 0. This concludes the proof of the proposition.

Proof of Theorem 2.2.: Note that, for any \( x \in \mathbb{T} \), one can express

\[
\tilde{f}_n(x) - f(x) = \frac{1}{C_n} \left( \tilde{f}_n(x) - f(x) + (1 - C_n)f(x) \right).
\]

Theorem 1.2 of Abdous and Kokonendji (2009) recalls that \( \tilde{f}_n(x) \) converges in mean square to \( f(x) \); such result obviously remains valid in our context. Proposition 2.1 and Slutsky’s theorem complete the proof.

Proof of Corollary 2.3.: It is enough to observe that

\[
\sup_{x \in \mathbb{T}} |\tilde{f}_n(x) - f(x)| = \frac{1}{C_n} \sup_{x \in \mathbb{T}} |\tilde{f}_n(x) - f(x) + f(x) (1 - C_n)|
\]

\[
\leq \frac{1}{C_n} \left\{ |1 - C_n| + \sup_{x \in \mathbb{T}} |\tilde{f}_n(x) - f(x)| \right\}.
\]

Consequently, Proposition 2.1, Theorem 2.2 and the continuous mapping theorem easily allow to deduce the corollary.

Proof of Proposition 2.4.: In view of Equations (2) and (3) and using the Cauchy–Schwarz inequality, one successively obtains

\[
\mathbb{E} \left[ \sum_{x \in \mathbb{T}} |\tilde{f}_n(x) - f(x)| \right]
\]

\[
= \mathbb{E} \left[ \sum_{x \in \mathbb{T}} \left| \frac{1}{C_n} (\tilde{f}_n(x) - f(x)) + \frac{f(x)}{C_n} - f(x) \right| \right]
\]

\[
\leq \mathbb{E} \left[ \frac{1}{C_n} \sum_{x \in \mathbb{T}} |\tilde{f}_n(x) - f(x)| \right] + \mathbb{E} \left[ \left| \frac{C_n - 1}{C_n} \right| \right]
\]

\[
\leq \mathbb{E} \left[ \sum_{x \in \mathbb{T}} |\tilde{f}_n(x) - f(x)| \right] + \mathbb{E} \left[ \left| \frac{C_n - 1}{C_n} \right| \left( 1 + \sum_{x \in \mathbb{T}} |\tilde{f}_n(x) - f(x)| \right) \right]
\]

\[
\leq \mathbb{E} \left[ \sum_{x \in \mathbb{T}} |\tilde{f}_n(x) - f(x)| \right] + 2 \mathbb{E} \left[ \left| \frac{C_n - 1}{C_n} \right| \right] + \mathbb{E} \left[ |C_n - 1| \right]
\]

\[
\leq \mathbb{E} \left[ \sum_{x \in \mathbb{T}} |\tilde{f}_n(x) - f(x)| \right] + \left( \mathbb{E} \left[ (C_n - 1)^2 \right] \right)^{1/2} \left( 1 + 2 \left( \mathbb{E} \left[ C_n^{-2} \right] \right)^{1/2} \right).
\]
Proposition 2.1 and the almost sure absolute boundedness of \( C_n^{-1} \) complete the proof. ■

**Proof of Theorem 2.5.:** From the end of Theorem 1.2, one may first observe that

\[
\sqrt{n} \left( \tilde{f}_n(x) - f(x) \right) = \frac{1}{C_n} \left( \tilde{f}_n(x) - \mathbb{E} \left[ \tilde{f}_n(x) \right] \right) \sqrt{n \text{Var} \left[ \tilde{f}_n(x) \right]} + \frac{\sqrt{n}}{C_n} \left( \mathbb{E} \left[ \tilde{f}_n(x) \right] - f(x) \right)
\]

\[
= \frac{1}{C_n} \left( \tilde{f}_n(x) - \mathbb{E} \left[ \tilde{f}_n(x) \right] \right) \sqrt{n \text{Var} \left[ \tilde{f}_n(x) \right]} + \frac{\sqrt{n}}{C_n} \left( \mathbb{E} \left[ \tilde{f}_n(x) \right] - f(x) \right) + \frac{\sqrt{n}}{C_n} (1 - C_n) f(x).
\] (12)

Let \( (Y_{n,i}) \) be the rowwise i.i.d. triangular array defined by

\[
Y_{n,i} = \frac{K_{x,h_n}(X_i) - \mathbb{E} \left[ K_{x,h_n}(X_i) \right]}{\sqrt{n \text{Var} \left( K_{x,h_n}(X_i) \right)}}, \quad i = 1, \ldots, n.
\]

It is clear that for all \( n \geq 1 \) and any \( i = 1, \ldots, n \),

\[
\mathbb{E} \left[ Y_{n,i} \right] = 0 \quad \text{and} \quad \sum_{i=1}^{n} \mathbb{E} \left[ Y_{n,i}^2 \right] = 1.
\]

Moreover, since \( \text{Var} \left[ K_{x,h_n}(X_1) \right] \longrightarrow f(x)(1 - f(x)) \) as \( n \to \infty \) (see Step 1 of the proof of Proposition 2.1) and using the fact that the \( X_i \)'s are i.i.d., one has

\[
\sum_{i=1}^{n} \mathbb{E} \left[ Y_{n,i}^3 \right] = \frac{\mathbb{E} \left[ \left( K_{x,h_n}(X_1) - \mathbb{E} \left[ K_{x,h_n}(X_1) \right] \right)^3 \right]}{\sqrt{n} \left( \text{Var} \left[ K_{x,h_n}(X_1) \right] \right)^{3/2}} \leq \frac{1}{\sqrt{n} \left( \text{Var} \left[ K_{x,h_n}(X_1) \right] \right)^{3/2}} \longrightarrow 0, \quad \text{as} \quad n \to \infty.
\]

Thus, Lindeberg’s Theorem implies that

\[
\frac{\tilde{f}_n(x) - \mathbb{E} \left[ \tilde{f}_n(x) \right]}{\sqrt{\text{Var} \left[ \tilde{f}_n(x) \right]}} = \sum_{i=1}^{n} Y_{n,i} \xrightarrow{D} \mathcal{N}(0, 1).
\]

Note also that

\[
\sqrt{n \text{Var} \left[ \tilde{f}_n(x) \right]} = \sqrt{\text{Var} \left[ K_{x,h_n}(X_1) \right]} \longrightarrow \sqrt{f(x)} \left( 1 - f(x) \right), \quad \text{as} \quad n \to \infty.
\]

In view of Proposition 2.1, Equation (12) and Slutsky’s theorem, it is therefore sufficient to prove that the two sequences \( \left\{ \sqrt{n} \mathbb{E} \left[ \tilde{f}_n(x) \right] \right\}_{n \geq 1} \) and \( \left\{ \sqrt{n}(1 - C_n) \right\}_{n \geq 1} \) converge in probability to 0 for obtaining the expected convergence in distribution stated in Theorem 2.5.
Indeed, we consider similar arguments as in the proof of the uniform convergence of $F_{2,n}$ on $\mathbb{T}$ (see Step 1 of the proof of Proposition 2.1) and the assumptions of Theorem 2.5 to deduce that
\[
\sqrt{n} \left| \mathbb{E} \left[ \hat{f}_n(x) \right] - f(x) \right| \leq 3\sqrt{n} \sup_{x \in \mathbb{T}} \mathbb{P} (Z_{x,h_n} \neq x) \longrightarrow 0, \quad \text{as } n \rightarrow \infty.
\]

To complete the proof of the theorem, note that by using the results of Step 1 in the proof of Proposition 2.1, there exists $N \in \mathbb{N}^*$ such that for any $n \geq N$, we have
\[
\text{Var} \left[ \sqrt{n}(1 - C_n) \right] = \sum_{x \in \mathbb{T}} \left\{ \text{Var} \left[ K_{x,h_n}(X_1) \right] - f(x) \left\{ 1 - f(x) \right\} \right\} + \sum_{x \in \mathbb{T}} \sum_{y \in \mathbb{T} \setminus \{x\}} \left\{ \text{Cov} \left( K_{x,h_n}(X_1), K_{y,h_n}(X_1) \right) + f(x)f(y) \right\}
\leq \kappa^* h_n \longrightarrow 0, \quad \text{as } n \rightarrow \infty,
\]
where $\kappa^*$ is a positive constant.

Similarly, one can use the results of Step 2 in the proof of Proposition 2.1 to prove
\[
\mathbb{E} \left[ \sqrt{n}(1 - C_n) \right] \leq \kappa^{**} \sqrt{nh_n} \longrightarrow 0, \quad \text{as } n \rightarrow \infty,
\]
where $\kappa^{**}$ is a positive constant. This completes the proof of the theorem.

**Proof of Proposition 2.6.** From Equations (4) and (6), one first has
\[
\mathbb{E} \left[ Y \right] = \mu + \frac{1}{D (\lambda(\mu, v), v)} \sum_{y=0}^{\infty} (y - \mu) \frac{[\lambda(\mu, v)]^y}{(y!)^v} = \mu.
\]

Secondly, one can observe that
\[
\lambda(\mu, v) \frac{\partial \mathbb{E} [Y]}{\partial \lambda(\mu, v)} = \frac{1}{D (\lambda(\mu, v), v)} \sum_{y=0}^{\infty} y^2 \frac{[\lambda(\mu, v)]^y}{(y!)^v}
+ \lambda(\mu, v) \frac{\partial}{\partial \lambda(\mu, v)} \left[ \frac{1}{D (\lambda(\mu, v), v)} \right] \sum_{y=0}^{\infty} y \frac{[\lambda(\mu, v)]^y}{(y!)^v}
= \mathbb{E} \left[ Y^2 \right] + \lambda(\mu, v)D (\lambda(\mu, v), v) \frac{\partial}{\partial \lambda(\mu, v)} \left[ \frac{1}{D (\lambda(\mu, v), v)} \right] \mathbb{E} \left[ Y \right]. \quad (13)
\]
Moreover, a direct calculation of derivative leads us to
\[
\lambda(\mu, v)D (\lambda(\mu, v), v) \frac{\partial}{\partial \lambda(\mu, v)} \left[ \frac{1}{D (\lambda(\mu, v), v)} \right] = -\lambda(\mu, v) \frac{\partial D (\lambda(\mu, v), v)}{D (\lambda(\mu, v), v)}
= -\frac{1}{D (\lambda(\mu, v), v)} \sum_{y=0}^{\infty} y \frac{[\lambda(\mu, v)]^y}{(y!)^v}
= -\mathbb{E} \left[ Y \right]. \quad (14)
\]
Hence, using (13) and (14) we deduce
\[ \text{Var} (Y) = \lambda(\mu, v) \frac{\partial \mathbb{E}[Y]}{\partial \lambda(\mu, v)}. \]

We still consider Equation (4) to obtain
\[ \lambda(\mu, v) = \mu \left( \frac{D(\lambda(\mu, v), v)}{\partial D(\lambda(\mu, v), v) / \partial \lambda(\mu, v)} \right), \]
which implies that
\[
\text{Var} (Y) = \lambda(\mu, v) \frac{\partial}{\partial \lambda(\mu, v)} \left[ \lambda(\mu, v) \frac{\partial \log D(\lambda(\mu, v), v)}{\partial \lambda(\mu, v)} \right] \\
\quad - \lambda(\mu, v) \frac{\partial}{\partial \lambda(\mu, v)} \left[ \lambda(\mu, v) \frac{\partial \log D(\lambda(\mu, v), v)}{\partial \lambda(\mu, v)} \right] + \left\{ \lambda(\mu, v) \right\}^2 \frac{\partial^2 \log D(\lambda(\mu, v), v)}{\left( \partial \lambda(\mu, v) \right)^2}. \quad (15)
\]

According to Gaunt et al. (2019, th. 1) under \( \{ \lambda(\mu, v) \}^{1/v} \to \infty \) as \( v \to \infty \), one can write
\[
\log D(\lambda(\mu, v), v) = v \{ \lambda(\mu, v) \}^{1/v} - \frac{v - 1}{2v} \log \lambda(\mu, v) - \log \left\{ (2\pi)^{(v-1)/2} \sqrt{v} \right\} \\
+ \frac{v^2 - 1}{24v} \{ \lambda(\mu, v) \}^{-1/v} + \frac{v^2 - 1}{48v^2} \{ \lambda(\mu, v) \}^{-2/v} + \mathcal{O} \left( \{ \lambda(\mu, v) \}^{-3/v} \right);
\]
and, therefore, one deduces the following two terms of Equation (15) as:
\[
\lambda(\mu, v) \frac{\partial \log D(\lambda(\mu, v), v)}{\partial \lambda(\mu, v)} = \{ \lambda(\mu, v) \}^{1/v} - \frac{v - 1}{2v} + \mathcal{O} \left( \{ \lambda(\mu, v) \}^{-1/v} \right)
\]
and
\[
\left\{ \lambda(\mu, v) \right\}^2 \frac{\partial^2 \log D(\lambda(\mu, v), v)}{\left( \partial \lambda(\mu, v) \right)^2} = \frac{1 - v}{v} \{ \lambda(\mu, v) \}^{1/v} + \frac{v - 1}{2v} + \mathcal{O} \left( \{ \lambda(\mu, v) \}^{-1/v} \right).
\]
This is enough to complete the proof.

**Acknowledgments**

The authors sincerely thank the Associate Editor and the anonymous referee for their valuable comments.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

For the second coauthor, this work is supported by the EIPHI Graduate School (contract ANR-17-EURE-0002).

**ORCID**

Sobom M. Somé http://orcid.org/0000-0003-4454-5777
References

Abdous, B., and Kokonendji, C.C. (2009), ‘Consistency and Asymptotic Normality for Discrete Associated-Kernel Estimator’, *African Diaspora Journal of Mathematics*, 8, 63–70.

Aitchison, J., and Aitken, C.G.G. (1976), ‘Multivariate Binary Discrimination by the Kernel Method’, *Biometrika*, 63, 413–420.

Chu, C.-Y., Henderson, D.J., and Parmeter, C.F. (2017), ‘On Discrete Epanechnikov Kernel Functions’, *Computational Statistics and Data Analysis*, 116, 79–105.

Devroye, L., and Lugosi, G. (2001), *Combinatorial Methods in Density Estimation*, New York: Springer.

Glad, I.K., Hjort, N.L., and Ushakov, N.G. (2003), ‘Correction of Density Estimators That are Not Densities’, *Scandinavian Journal of Statistics*, 30, 415–427.

Fung, T., Alwan, A., Wishart, J., and Huang, A. (2020), *mpcmp: Mean Parametrized Conway-Maxwell Poisson Regression* (Ver. 0.3.6), R package. Available at http://cran.r-project.org/.

Gaunt, R.E., Iyengar, S., Olde Daalhuis, A.B., and Simsek, B. (2019), ‘An Asymptotic Expansion for the Normalizing Constant of the Conway-Maxwell-Poisson Distribution’, *Annals of the Institute of Statistical Mathematics*, 71, 163–180.

Harfouche, L., Adjabi, S., Zougab, N., and Funke, B. (2018), ‘Multiplicative Bias Correction for Discrete Kernels’, *Statistical Methods and Applications*, 27, 253–276.

Hoeffding, W. (1963), ‘Probability Inequalities for Sums of Bounded Random Variables’, *Journal of the American Statistical Association*, 58, 13–30.

Huang, A. (2017), ‘Mean-Parametrized Conway-Maxwell-Poisson Regression Models for Dispersed Counts’, *Statistical Modelling*, 17, 359–380.

Huang, A., Sippel, L., and Fung, T. (2022), ‘Consistent Second-Order Discrete Kernel Smoothing Using Dispersed Conway-Maxwell-Poisson Kernels’, *Computational Statistics*, 37, 551–563.

Kokonendji, C.C., Mizère, D., and Balakrishnan, N. (2008), ‘Connections of the Poisson Weight Function to Overdispersion and Underdispersion’, *Journal of Statistical Planning and Inference*, 138, 1287–1296.

Kokonendji, C.C., and Senga Kiessé, T. (2011), ‘Discrete Associated Kernels Method and Extensions’, *Statistical Methodology*, 8, 497–516.

Kokonendji, C.C., Senga Kiessé, T., and Zocchi, S.S. (2007), ‘Discrete Triangular Distributions and Non-Parametric Estimation for Probability Mass Function’, *Journal of Nonparametric Statistics*, 19, 241–254.

Kokonendji, C.C., and Somé, S.M. (2018), ‘On Multivariate Associated Kernels to Estimate General Density Functions’, *Journal of the Korean Statistical Society*, 47, 112–126.

Kokonendji, C.C., and Somé, S.M. (2021), ‘Bayesian Bandwidths in Semiparametric Modelling for Nonnegative Orthant Data With Diagnostics’, *Stats*, 4, 162–183.

Kokonendji, C.C., and Varron, D. (2016), ‘Performance of the Discrete Associated Kernel Estimator Through the Total Variation Distance’, *Statistics and Probability Letters*, 110, 225–235.

Kokonendji, C.C., and Zocchi, S.S. (2010), ‘Extensions of Discrete Triangular Distributions and Boundary Bias in Kernel Estimation for Discrete Functions’, *Statistics and Probability Letters*, 80, 1655–1662.

Li, Q., and Racine, J. (2003), ‘Nonparametric Estimation of Distributions With Categorical and Continuous Data’, *Journal of Multivariate Analysis*, 86, 266–292.

Ouyang, D., Li, Q., and Racine, J. (2006), ‘Cross-Validation and the Estimation of Probability Distributions With Categorical Data’, *Journal of Nonparametric Statistics*, 18, 69–100.

R Core Team (2021), *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing. Available at http://cran.r-project.org/.

Racine, J., Li, Q., and Yan, K.X. (2020), ‘Kernel Smoothed Probability Mass Functions for Ordered Datatypes’, *Journal of Nonparametric Statistics*, 32, 563–586.

Senga Kiessé, T. (2017), ‘On Finite Sample Properties of Nonparametric Discrete Asymmetric Kernel Estimators’, *Statistics*, 51, 1046–1060.
Senga Kiessé, T., and Cuny, H.E. (2014), ‘Discrete Triangular Associated Kernel and Bandwidth Choices in Semiparametric Estimation for Count Data’, Journal of Statistical Computation and Simulation, 84, 1813–1829.
Senga Kiessé, T., and Ventura, A. (2016), ‘Discrete Non-Parametric Kernel Estimation for Global Sensitivity Analysis’, Reliability Engineering and System Safety, 146, 47–54.
Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., and Boatwright, P. (2005), ‘A Useful Distribution for Fitting Discrete Data: Revival of the Conway-Maxwell-Poisson Distribution’, Journal of the Royal Statistical Society, Series C, 54, 127–142.
Toledo, D., Umetsu, C.A., Camargo, A.F.M., and Rodrigues de Lara, I.A. (2022), ‘Flexible Models for Non-Equidispersed Count Data: Comparative Performance of Parametric Models to Deal With Underdispersion’, ASTA Advances in Statistical Analysis, 106, 473–497.
Wang, M., and Van Ryzin, J. (1981), ‘A Class of Smooth Estimators for Discrete Distributions’, Biometrika, 68, 301–309.
Wansouwé, W.E., Somé, S.M., and Kokonendji, C.C. (2016), ‘Ake: An R Package for Discrete and Continuous Associated Kernel Estimations’, The R Journal, 8, 258–276.