Greek Statistical Institute:

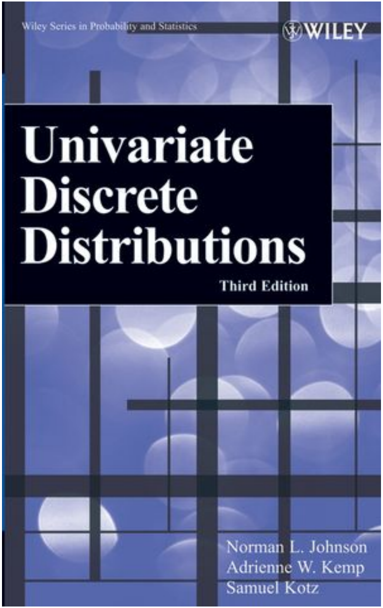
Workshop on "Discrete Distributions" in Memory of Adrienne Freda KEMP

Title: Discrete Distributions for Kernel Smoothing

Prof. Célestin C. KOKONENDJI

Email Address: celestin.kokonendji@univ-fcomte.fr

Harokopio University of Athens: **2024.01.27**



Thanks also to Adrienne Freda KEMP (1930 - 2022)

Outline:

1 Motivation & Introduction: Cont. w.r. Discr.

2 Definitions & (Asymptotical) Properties

3 Constructions: Two Approaches

4 Conclusion & Further Problems

- Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**:832-837.
- Parzen E (1962) On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**:1065-1076.

1. Motivation & Introduction

Let $X_1, \ldots, X_n$ be an iid n-sample from **unknown** pdf/pmf f on $\mathbb{T} \subseteq \mathbb{R}$

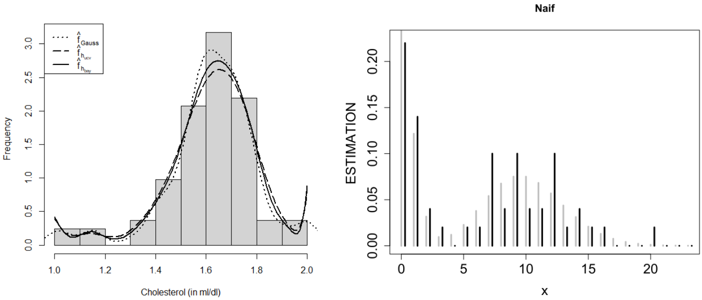
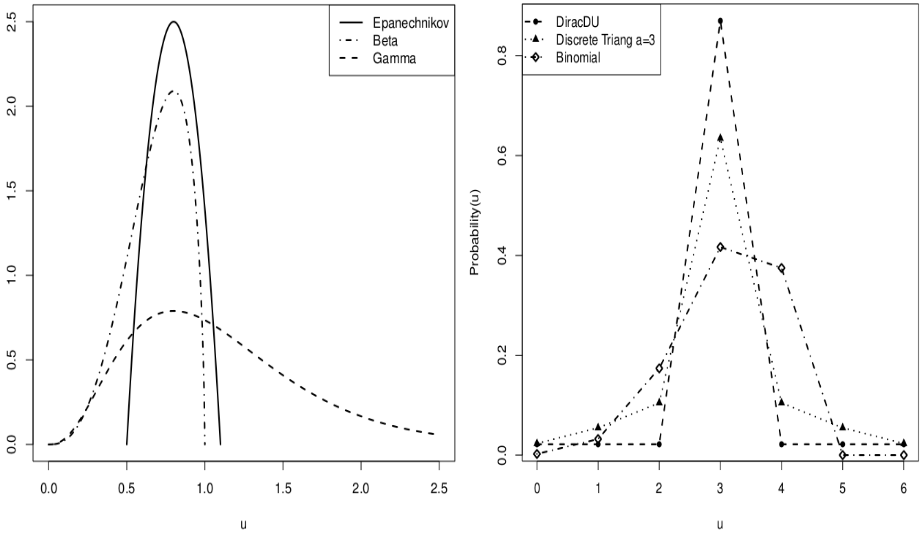


Figure: Continuous $\mathbb{T} = [a, b]$ and Discrete $\mathbb{T} = \{0, 1, \ldots\}$ kernel smoothings.

$$\widetilde{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \forall x \in \mathbb{R} : \textit{Rosenblatt}\,(1956);\ \textit{Parzen}\,(1962)$$

$$= \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i),\ (\textit{asymm.}) : \textit{Kokonendji \& SengaKiesse}(2007, 2011)$$

Figure: Shapes of ***continuous*** and **discrete** associated kernels:
(***a***) Epanechnikov, beta and gamma with same $x = 0.8$ and $h = 0.3$;
(**b**) DiracDU, discrete triangular and binomial with $x=3$ and $h = 0.13$.



Associated **Discrete** Kernel Estimations of f ($n \to \infty$ & $n < n_0$):

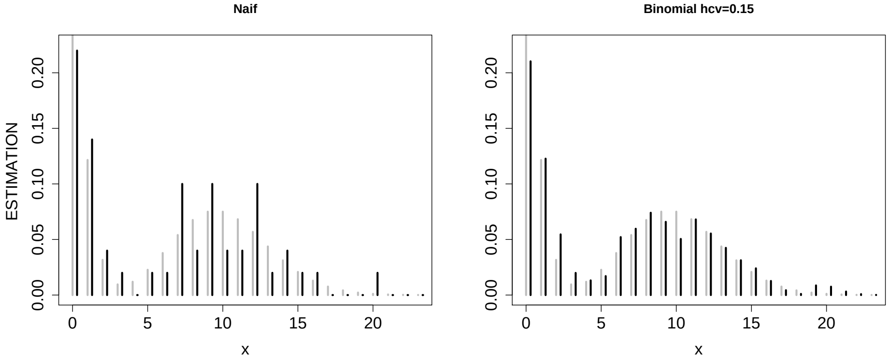


Figure: Discrete smoothings using empirical (or naive) and binomial kernel estimators of simulated data from $f = 0.4\mathcal{P}(0.5) + 0.6\mathcal{P}(10)$ with $n = 50$.

$$\widetilde{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i),\ \forall x \in \mathbb{T}:\ \textit{discrete set};$$

(i) $h > 0$ is the smoothing parameter: $h = h(n) \to 0$ as $n \to \infty$.

(ii) $K_{x,h}(\cdot)$ is the associated-kernel (a pmf) linked to x and h.

The **Mean Integrated Square Error** (MISE) criterion:

$$MISE(\widetilde{f}_n) = IBias^2\left(\widetilde{f}_n\right) + IVar\left(\widetilde{f}_n\right) \textcolor{red}{\rightarrow 0} \text{ or } \nrightarrow 0 \textbf{ ?! - Do the job.}$$

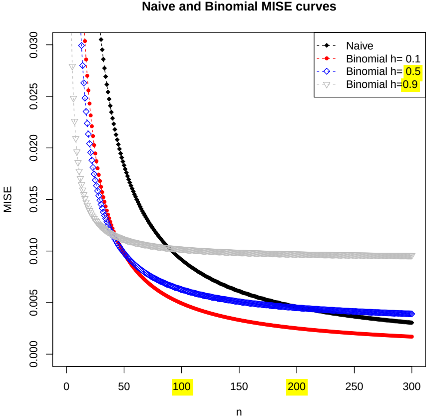


Figure: MISE curves (n_0?) for Dirac (or Naive) and binomial kernels with f.

(**!!!**) Discrete associated-kernels: Definition - Properties

(**??**) - Constructions - Further problems.

2. Definitions & Properties

Definition – *Discrete Associated-Kernel (DAK)*

Let $x \in \mathbb{T}$ be in discrete support of f to be estimated, and $h > 0$. A pmf $K_{x,h}$ (parameters x and h) on support $\mathbb{S}_{x,h}$ is a **DAK** if

$$(\mathbf{i})\ x \in \mathbb{S}_{x,h}\ (\mathbf{ii})\ \lim_{h\to 0} \mathbb{E}Z_{x,h} = x\ (\mathbf{iii})\ \lim_{h\to 0} Var\, Z_{x,h} = \underline{\sigma^2 \in [0,1)},$$

for all $x \in \mathbb{T}$, with the discrete r.v. $Z_{x,h} \sim K_{x,h}$.

Remarks:

(**i**) $\mathbb{T} \cap \mathbb{S}_{x,h} \neq \emptyset$ (**ii**) $\mathbb{E}Z_{x,h} = x + \mathbf{a}(x,h)$ (**iii**) $Var\, Z_{x,h} = \mathbf{b}(x,h)$, with $\mathbf{a}(x,h) \to 0$ and $\mathbf{b}(x,h) \to \sigma^2$ as $h \to 0$ and $\forall x \in \mathbb{T}$.

$$(asymm.) \qquad K_{x,h}(\cdot) \hookleftarrow \frac{1}{h} K\left(\frac{x-\cdot}{h}\right) \qquad (symm.).$$

$\sigma^2 \neq 0$: *first-order*, not converge (as Binomial), close neighbors;
$\underline{\sigma^2 = 0}$: *second-order*, converge (as Dirac). K_θ = type of kernel.

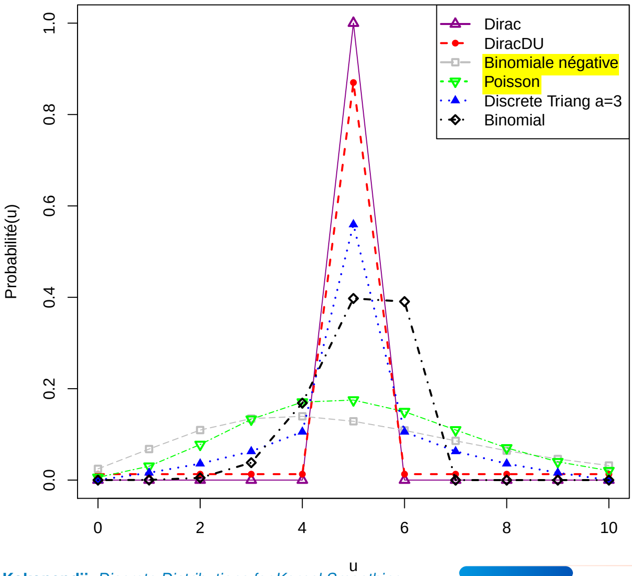
Table: Summary of **only seven** discrete associated-kernels!

Type of kernel K_θ	$\mathbb{S}_{x,h}$	$\mathbb{E}Z_{x,h} - x$	$Var\, Z_{x,h}$	Authors
0. Dirac ($h = 0$)	$\{x\}$	0	0	—
1. DiracDU	$[\![0, c-1]\!]$	$\dfrac{c(c-2x-1)}{2(c-1)}h$	$\cdots \to 0$	Aitchison, Aitken(1976)
2. DiracDU-ext$\mathbb{Z}$	$\mathbb{Z} := \pm\mathbb{N}$	0	$\dfrac{(1+h)h}{(1-h)^2}$	Wang, Van Rizin(1981)
3. Discrete Triangular*	$[\![x, x \pm m]\!]$	0	$\cdots \to 0$	Kokonendji et al.(2007)
4. CoM-Poisson	$\mathbb{N}$	0	$\cdots \to 0$	Huang et al.(2022)
5. Double Poisson	$\mathbb{N}$	$\approx h$	$\approx (x+h)h$	Kokonendji et al.(2023)
6. Gamma-count	$\mathbb{N}$	$\approx h$	$\cdots \to 0$	Kokonendji et al.(2023)
7. Binomial	$[\![0, x+1]\!]$	h	$\to \dfrac{x}{x+1}$	Kokonendji, SK.(2011)
–. Poisson	$\mathbb{N}$	h	$x + h \to x$	Kokonendji, SK.(2011)
–. Negative binomial	$\mathbb{N}$	h	$\to \dfrac{x(2x+1)}{x+1}$	Kokonendji, SK.(2011)

with $\lim_{h\to 0} Var\, Z_{x,h} = 0$, in $(0,1)$, and in $(0,\infty)$; SK = Senga Kiessé.

*Discrete Epanechnikov (Chu et al.,2017) **+** Discrete symmetric optimal (SK, Durrieu, 2024)

Figure: Comparative graphics of some DAKs on the edge $\underline{x = 5}$.



2.1. Properties of DAK Estimator $\widetilde{f}_{n,h}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i)$:

- **Criterion MISE**: Let $f^{(k)}$ be the kth finite difference of f. Then

$$\mathbb{E}\widetilde{f}_{n,h}(x) = \mathbb{E}f(Z_{x,h}) = \sum_{k\geq 0} \frac{1}{k!}\mathbb{E}\left(Z_{x,h} - \mathbb{E}Z_{x,h}\right)^k f^{(k)}(\mathbb{E}Z_{x,h}) \text{ and}$$

$$MISE\left(\widetilde{f}_{n,h}\right) = \sum_{x\in\mathbb{T}} Bias^2\widetilde{f}_{n,h}(x) + \sum_{x\in\mathbb{T}} Var\,\widetilde{f}_{n,h}(x),$$

with $\quad Bias\,\widetilde{f}_{n,h}(x) = f(\underline{\mathbb{E}Z_{x,h}}) - f(x) + \frac{1}{2}(\underline{Var\,Z_{x,h}})f^{(2)}(x) + o(h)$

and $\quad Var\,\widetilde{f}_{n,h}(x) = \frac{1}{n}f(x)\left[\left\{\underline{\mathbb{P}(Z_{x,h} = x)}\right\}^2 - f(x)\right] + o(1/h).$

- **Normalizing constant**: If $\quad C_{n,h} := \sum_{x\in\mathbb{T}} \widetilde{f}_{n,h}(x) > 0 \quad$ then

$$\mathbb{E}C_{n,h} = 1 + \sum_{x\in\mathbb{T}} Bias\,\widetilde{f}_{n,h}(x) \text{ and } Var\,C_{n,h} = \sum_{x\in\mathbb{T}} Var\,\widetilde{f}_{n,h}(x).$$

$\underline{C_{n,h} = 1}$ for Dirac, DiracDU, DiracDU-ext$\mathbb{Z}$, & $\neq 1$ for the others.

Note: The two first finite differences of $g : \mathbb{N} \to \mathbb{R}$ can be defined (quasi-symmetrically) by

$$g^{(1)}(x) = \begin{cases} \frac{1}{2}\{g(x+1) - g(x-1)\}, & \text{if } x \in \mathbb{N}\setminus\{0\} \\ g(1) - g(0), & \text{if } x = 0 \end{cases}$$

and

$$g^{(2)}(x) = \begin{cases} \frac{1}{4}\{g(x+2) - 2g(x) + g(x-2)\}, & \text{if } x \in \mathbb{N}\setminus\{0,1\} \\ \frac{1}{4}\{g(3) - 3g(1) + g(0)\}, & \text{if } x = 1 \\ \frac{1}{2}\{g(2) - 2g(1) + g(0)\}, & \text{if } x = 0. \end{cases}$$

2.2. Asymptotic properties of non-normalized DAKE $\widetilde{f}_{n,h}(x)$

Since $h = h_n \to 0$ as $n \to \infty$, one can write $\widetilde{f}_{n,h} =: \widetilde{f}_n$.

Theorem – *MSE, Strong consistency & Asymptotic normality* (Abdous & Kokonendji, 2009)

For any $x \in \mathbb{T}$ and for **DAK** of second-order (i.e., $\sigma^2 = 0$), one has

$$\widetilde{f}_n(x) \xrightarrow[n\to\infty]{L^2,\, a.s.} f(x),$$

where " $\xrightarrow{L^2,\, a.s.}$ " stands for both "mean square and almost surely convergences". Furthermore, if $f(x) > 0$ then

$$\left\{\widetilde{f}_n(x) - \mathbb{E}\widetilde{f}_n(x)\right\}\left\{Var\,\widetilde{f}_n(x)\right\}^{-1/2} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,1),$$

where " $\xrightarrow{\mathcal{D}}$ " stands for "convergence in distribution" and $\mathcal{N}(0,1)$ denotes the standard normal distribution.

2.3. Asymptotic properties of normalized DAKE $\widehat{f}_n(x) := \widetilde{f}_n(x)/C_n$

The first set of assumptions is **uniformly** in the target $x \in \mathbb{T}$:

(**A1**): $x \in \mathbb{S}_{x,h_n}$, $\lim_{n\to\infty} \sup_{x\in\mathbb{T}} \left|\mathbb{E}Z_{x,h_n} - x\right| = 0$ and $\lim_{n\to\infty} \sup_{x\in\mathbb{T}} Var\, Z_{x,h_n} = 0.$

Proposition (Esstafa, Kokonendji & Somé, 2023)

Under (**A1**), the normalizing r.v. $C_n \xrightarrow[n\to\infty]{L^2} 1.$

Theorem – *Consistency* (Esstafa et al., 2023)

Under (**A1**) and $\forall x \in \mathbb{T}$, we have: $\widehat{f}_n(x) \xrightarrow[n\to\infty]{\mathbb{P}} f(x),$

where " $\xrightarrow{\mathbb{P}}$ " stands for "convergence in probability".

Corollary – *Uniform consistency* (Esstafa et al., 2023)

For **DAK** ($\sigma^2 = 0$) and $\mathbb{T}$ finite: $\sup_{x\in\mathbb{T}} \left|\widehat{f}_n(x) - f(x)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$

The second set of conditions is to quantify the speed of convergence to 0 in (**A1**) for **a refined result** of the asymptotic normality:

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

Theorem – *Asymptotic normality* (Esstafa et al., 2023)

Let **(A2)** be satisfied. If $(h_n)_{n\geq 1}$ is chosen such that $\sqrt{n}h_n \to 0$ as $n \to \infty$, then, for any $x \in \mathbb{T}$ such that $f(x) > 0$,

$$\sqrt{n}\left\{\widehat{f}_n(x) - f(x)\right\} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, f(x)\{1 - f(x)\}).$$

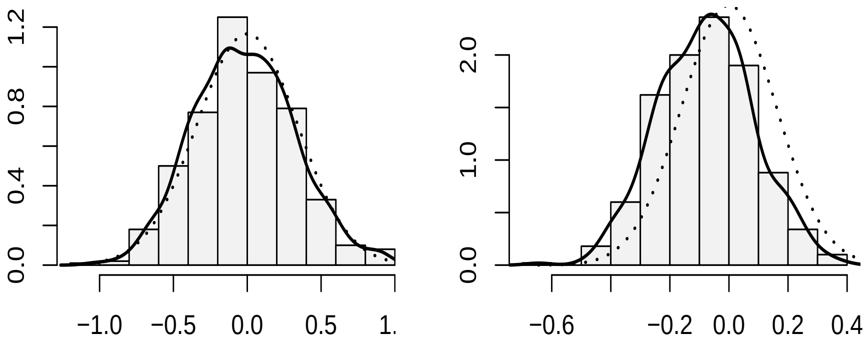
NB: $\widehat{f}_n = \widetilde{f}_n / C_n$ outperforms $\widetilde{f}_n$ (**but numerically for** $\sigma^2 \neq 0$**!**).

Proposition – *Global comparison in L^1* (Esstafa et al., 2023)

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

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

Figure: Empirical distributions of $\sqrt{n}\{\widehat{f_n}(x) - f(x)\}$, $f := \mathcal{P}(8)$, at $x = 6$ over $N_{sim} = 500$ 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; the centered Gaussian density with same variance is plotted in dotted line



2.4. Bandwidth selections – crucial choice (for discrete)!

Cross-validation (global): $\widetilde{h}_{cv} = \arg\min_{h>0} \left(\sum_{x\in\mathbb{T}} \widetilde{f}^2_{n,h}(x) - \frac{2}{n}\sum_{i=1}^{n} \widetilde{f}_{n,h;-i}(X_i) \right)$

Bayes global: $\widetilde{h}_{Bg} \propto \int_{h>0} h\,\pi(h) \prod_{i=1}^{n} \widetilde{f}_{n,h;-i}(X_i)\,dh$

Bayes local: $\widetilde{h}_{B\ell}(x) \propto \int_{h>0} h\,\pi(h)\,\widetilde{f}_{n,h}(x)\,dh, \quad x \in \mathbb{T}$

Bayes adaptive: $\widetilde{h}_{Ba}(i) \propto \int_{h_i>0} h_i\,\pi(h_i)\,\widetilde{f}_{n,h;-i}(X_i)\,dh_i, \quad i = 1,\ldots,n$

with $\widetilde{f}_{n,h;-i}(X_i) := \frac{1}{n-1}\sum_{\ell=1,\ell\neq i}^{n} K_{X_i,h}(X_\ell)$, $\widetilde{f}_{n,h} = \widetilde{f}_n$ and $\pi =$ prior distr.

3. Constructions

Two approaches to build a DAK & from Existing or New pmf:

1. Direct DAKE: an *intuitive technique* of DAKE depending on the author.s, thus I deduced the corresponding DAK; e.g.,

- New pmf: DiracDU (Aitchison & Aitken, 1976), and DiracDU-ext$\mathbb{Z}$ (Wang & Van Rizin, 1981).
- Existing pmf: CoM-Poisson (Huang et al., 2022).

2. Method MDAK: a *systematic* or *rigorous approach* using `the bell shape` of DAK like that of `Napoleon's hat!`

- New pmf: Discrete Triangular (Kokonendji et al.,2007), and Asymm. Disc. Triang. (Kokonendji, Zocchi, 2010).
- Existing: Binomial (Kokonendji, Senga Kiessé, 2011), and Double Poisson, Gamma-count (Kokonendji et al.,2023).

MDAK: the bell shape like that of Napoleon's hat or $\mathcal{N}(x, h^2)$

Mode-Dispersion (MD) for AK: MDAK

From Libengué & Kokonendji (2017), the MD method is applicable to underdispersed count distributions (i.e., $Var < \mathbb{E}$).

Definition – *Mean-Dispersion-Ready (MDR) pmf*

A *MDR pmf* K_θ is a underdispersed parametrized pmf with support $\mathbb{S}_\theta \subseteq \mathbb{R}$, $\theta \in \Theta \subseteq \mathbb{R}^2$, such that K_θ has second moments with **mode** $m \in \mathbb{R}$ and admitting a **dispersion** parameter D.

Remark – (Kokonendji et al., 2023)

Let K_θ be a MDR pmf on $\mathbb{S}_\theta \subseteq \mathbb{R}$. Then:

(i) the mode m of K_θ always belongs to $\mathbb{S}_\theta$;

(ii) if μ is the mean of K_θ then $K_\theta(m) \geq K_\theta(\lfloor \mu \rfloor)$, where $\lfloor \cdot \rfloor$ denotes the integer part.

From a **unimodal** MDR pmf K_θ, the MDAK method requires a solution of: $\boxed{(\theta(m, D)) = (x, h)}$

4. Conclusion & Further Problems

Type of kernel K_θ	$\mathbb{T}$	$C_{n,h}$	Construction	Authors
0. Dirac ($h = 0$)	$\mathbb{T}$	1	—	—
1. DiracDU	$[\![0, c-1]\!]$	1	*DAKE* - new	Aitchison, Aitken(1976)
2. DiracDU-ext$\mathbb{Z}$	$\mathbb{Z} := \pm\mathbb{N}$	1	*DAKE* - new	Wang, Van Rizin(1981)
3. Discrete Triangular*	$\mathbb{T}$	$\neq 1$	MDAK - new	Kokonendji et al.(2007)
4. CoM-Poisson	$\mathbb{N}$	$\neq 1$	*DAKE* - exist	Huang et al.(2022)
5. Double Poisson	$\mathbb{N}$	$\neq 1$	MDAK - exist	Kokonendji et al.(2023)
6. Gamma-count	$\mathbb{N}$	$\neq 1$	MDAK - exist	Kokonendji et al.(2023)
7. Binomial	$[\![0, m]\!] \subseteq \mathbb{N}$	$\neq 1$	MDAK - exist	Kokonendji, SK.(2011)
–. Poisson	$\mathbb{N}$	$\neq 1$	MDAK - exist	Kokonendji, SK.(2011)
–. Negative binomial	$\mathbb{N}$	$\neq 1$	MDAK - exist	Kokonendji, SK.(2011)

with $\lim_{h\to 0} Var(Z_{x,h}) = 0$, in $(0,1)$, and in $(0,\infty)$.

4. Conclusion & Further Problems

1. **Probabilistic** Interpretations/Explanations of $C_n \neq 1$ but around 1**?**
2. **Underdispersed** pmf is not enough to build a DAK; e.g., BerPoi, generalized Poisson, underdispersed Poisson, BerG, hyper-Poisson.
3. **Extensions** of the DAK range: e.g.,
   - *convergent* version of the flexible binomial (Poisson) kernel
   - *asymmetric* version of the categorical DiracDU kernel
   - *discrete trigonometric kernel*.
4. **Multivariate** version of DAK (e.g., Kokonendji & Somé, 2018):
   - $\prod_{\ell=1}^{d} K_{x_\ell, h_\ell}^{[\ell]}(\cdot)$: *multiple* or *mixte*, with independent components
   - $\mathbf{K}_{\mathbf{x},\mathbf{H}}(\cdot)$: with correlated components, where $\mathbf{H} = \left(h_{jk}\right)_{j,k}$ is a $(d \times d)$-matrix of bandwidths and $\mathbf{x} = (x_1, \ldots, x_d)^\top$
5. **Recursive** version of DAKE: an attempt by Aboubacar & Kokonendji (2024) for data streams.
6. **DAKE for other** f: wpmf, (sp) regression, hazard, discrimination, etc.

References

1. Aitchison J & Aitken CGG (1976) Multivariate binary discrimination by the kernel method. *Biometrika* **63**:413-420.
2. Chu C-Y, Henderson DJ & Parmeter (2017) On discrete Epanechnikov kernel functions. *Comput. Statist. Data Anal.* **116**:79-105.
3. Huang A, Sippel L & Fung T (2022) Consistent second-order discrete kernel smoothing using dispersed Conway-Maxwell-Poisson kernels. *Comput. Statist.* **37**:551-563.
4. Senga Kiessé T & Durrieu G (2024) On a discrete symmetric optimal associated kernel for estimating count data distributions. *Statist. Probab. Lett.* **208**:110078.
5. Wang M & Van Ryzin JA (1981) class of smooth estimators for discrete distributions. *Biometrika* **68**:301-309.
6. **Kokonendji &**, e.g., Abdous, Aboubacar, Esstafa, Libengué, Senga Kiessé, Somé, Zocchi: **https://ckokonen.pages.math.cnrs.fr/**

... Thank You [Merci / Singîla] for Interest!

Perhaps in 2026.07 at Besançon/Bangui for my sixties!