Communications for Statistical Applications and Methods
2024, Vol. 31, No. 6, 645–659

https://doi.org/10.29220/CSAM.2024.31.6.645
Print ISSN 2287-7843 / Online ISSN 2383-4757

Charlier series approximation for nonhomogenious Poisson processes

Hyung-Tae Ha[1,a]

[a]Department of Applied Statistics, Gachon University, Korea

Abstract

This study investigates the Charlier series approximation for modeling nonhomogeneous Poisson processes. It focuses on mixtures of Poisson distributions and Markov-Modulated Poisson processes to address complex temporal data patterns, such as hospital admission rates. The Charlier series approximation is constructed by expanding probability mass functions using Charlier orthogonal polynomials, which allow for adjustments to reflect higher-order moments like skewness and kurtosis. These polynomials are combined with a Poisson weight function to create flexible approximations tailored to the variability in event rates. Two artificial examples demonstrate the method's effectiveness in capturing dynamic event behaviors. A real-world application to hospital admission data further highlights its practical utility. Performance is assessed using Kullback-Leibler divergence, quantifying the improvement over simple Poisson models. The results show that the Charlier series provides enhanced data fitting and deeper insights into complex probabilistic structures.

Keywords: nonhomogeneous Poisson process, Charlier series approximation, Markov-Modulated Poisson process, mixture of Poisson distributions, hospital admissions data

1. Introduction

1.1. Motivation

Nonhomogeneous Poisson processes are essential statistical tools that model events over time where the occurrence rate varies, making them vital in many scientific fields. Traditional Poisson process models require a constant occurrence rate and may not capture complex event variability, often leading to oversimplified or inaccurate models. This limitation is pronounced in cases of overdispersion or underdispersion, where the variance differs significantly from the mean, affecting the accuracy and reliability of the predictions.

Charlier polynomials, traditionally linked with the Poisson distribution, present a promising solution by allowing adjustments to the probability mass function through expansions (Kokonendji *et al.*, 2010). These adjustments can better align the model with empirical data, particularly by addressing variance misestimations. The flexibility of Charlier polynomial coefficients to reflect the skewness and kurtosis of real-world data enhances the ability to fit the model (Fokianos *et al.*, 2009).

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (RS-2023-00259004) supervised by the IITP (Institute for Information and Communications Technology Planning and Evaluation).

[1]Department of Applied Statistics, Gachon University, Seongnam-ci, Kyunggi-do, South Korea. E-mail: htha@gachon.ac.kr

Published 30 November 2024 / journal homepage: http://csam.or.kr

This study addresses the need that Charlier polynomial series approximation accommodates complex variability in event rates. Although Charlier expansions have been effectively used in homogeneous settings, their application to nonhomogeneous Poisson processes with time-varying rates remains underexplored. This paper contributes by integrating Charlier polynomial series with Poisson processes, thereby enhancing the statistical modeling capabilities for complex time-varying systems. This integration not only refines distributional attributes such as skewness and kurtosis but also improves the models' adaptability and accuracy.

2. Theoretical background

2.1. Nonhomogeneous Poisson processes

Nonhomogeneous Poisson processes are sophisticated stochastic models designed to represent event occurrences where the rate, $\lambda(t)$, varies with time or space. In contrast to homogeneous Poisson processes, which assume constant rates, non-homogeneous Poisson processes use a time-dependent intensity function, $\lambda(t)$, to provide a more realistic representation of scenarios where event probabilities are not static but fluctuate. Two common models for nonhomogeneous Poisson processes are the mixture of Poisson processes and the Markov-Modulated Poisson process (Karlis, 2005). These models provide sophisticated methods for handling data that exhibit varying event rates over time or space, which are not adequately captured by the homogeneous Poisson process.

The mixture of Poisson processes is a statistical model that combines multiple Poisson processes, each with its own rate parameter, into a single process. This approach is particularly useful for modeling data where the event rate varies due to unobserved heterogeneity among subpopulations. Each component of the mixture contributes to the overall process at a rate proportional to its mixing weight, allowing for a flexible representation of the aggregate event rate across different segments of the data. Each component process has its distinct rate function, $\lambda_i(t)$, weighted by coefficients π_i, such that:

$$\lambda(t) = \sum_{i=1}^{\ell} \pi_i \lambda_i(t),$$

ensuring ℓ is the number of mixture Poisson components, $\sum_{i=1}^{\ell} \pi_i = 1$ and each π_i is non-negative. This mixture model allows for flexible adaptation to diverse and complex event patterns, accommodating changes in the intensity function over different segments of time or space.

Markov-modulated Poisson process is a more dynamic model where the rate of the Poisson process is controlled by an underlying Markov chain. This state-dependent mechanism allows the Markov-modulated Poisson process to model systems where the intensity of events varies according to an evolving state of the system, which is governed by the Markov chain transitions. The rate function $\lambda(t)$ is directly influenced by the state of an underlying Markov chain at time t, described as:

$$\lambda(t) = \lambda_{X(t)},$$

where $X(t)$ indicates the current state of the Markov chain, with $\lambda_{X(t)}$ determining the rate specific to that state. This model is exceptionally well-suited for environments, where event rates are inherently dynamic and influenced by various stochastic processes, reflecting the complexities of real-world systems.

2.2. Charlier orthogonal polynomial series

Charlier polynomials, also known as Poisson-Charlier polynomials, are a series of orthogonal polynomials closely associated with the Poisson distribution. They are essential for modeling deviations from standard Poisson behavior, allowing for precise adjustments through series approximations. The general form of the nth Charlier polynomial, $C_n(x;a)$, is defined as:

$$C_n(x;a) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{x!}{(x-k)!} a^{n-k},$$

where a typically represents the mean rate of the Poisson process.

These polynomials are orthogonal with respect to the Poisson weight function $e^{-a}(a^x/x!)$, holding over the set of non-negative integers:

$$\sum_{x=0}^{\infty} C_m(x;a) C_n(x;a) e^{-a} \frac{a^x}{x!} = \delta_{mn} a^n,$$

where δ_{mn} is the Kronecker delta, indicating that C_m and C_n are orthogonal for $m \neq n$. The generating function for Charlier polynomials provides a powerful tool for deriving their properties and simplifying calculations:

$$G(t,x;a) = e^{-at}(1+t)^x = \sum_{n=0}^{\infty} C_n(x;a) \frac{t^n}{n!}.$$

Charlier polynomials also adhere to a three-term recurrence relation, essential for iterative computations:

$$xC_n(x;a) = C_{n+1}(x;a) + (a+n)C_n(x;a) + anC_{n-1}(x;a).$$

The explicit expressions for the first four Charlier polynomials, parameterized by a, which represents the mean of the associated Poisson distribution: The zeroth polynomial $C_0(x;a) = 1$, the first polynomial $C_1(x;a) = x - a$, the second polynomial $C_2(x;a) = (1/2)(x-a)^2 - x$, and the third polynomial $C_3(x;a) = (1/6)(x-a)^3 - (3/2)(x-a)x + a^2$.

These expressions for Charlier polynomials demonstrate increasing complexity with higher degrees, capturing more detailed aspects of sample data. Each polynomial is derived directly from the general formula for Charlier polynomials, and as n increases, the expressions incorporate progressively higher powers of x and a. This detailed structure is crucial in probabilistic models that rely on the Poisson distribution, as it allows for precise adjustments to the model based on empirical data.

3. Methodology

3.1. Charlier series approximation

The Charlier polynomials, when combined with their associated weight function from the Poisson distribution, form a foundation for diverse applications in statistical modeling, particularly through generating functions and probabilistic analyses (Grandell, 1997). The weight function, representative of the Poisson distribution, is defined as follows:

$$w(x;a) = \frac{e^{-a}a^x}{x!},$$

where a is a positive parameter representing the mean of the distribution and x is a non-negative integer (Sundt and Vernic, 2009). A linear combination of Charlier polynomials can be expressed as:

$$p(x; a) = c_0 C_0(x; a) + c_1 C_1(x; a) + c_2 C_2(x; a) + \cdots$$

with c_0, c_1, c_2, ... being constant coefficients. The integration of this series with the Poisson weight function yields the following function:

$$f(x) = e^{-a} \frac{a^x}{x!} \left(c_0 C_0(x; a) + c_1 C_1(x; a) + c_2 C_2(x; a) + \cdots \right)$$

which simplifies to:

$$f(x) = e^{-a} \frac{a^x}{x!} \left(c_0 + c_1 (x - a) + c_2 \left(\frac{1}{2} (x - a)^2 - x \right) + \cdots \right).$$

This framework is particularly useful for expanding the probabilities associated with a Poisson process, which is critical for addressing issues like overdispersion within the distribution. The expanded probability function for a Poisson process with truncation is detailed as follows:

$$f_n(x; a) = w(x; a) \left(c_0 + c_1 C_1(x; a) + c_2 C_2(x; a) + \cdots + c_n C_n(x; a) \right), \tag{3.1}$$

where c_n are the corresponding coefficients. This expanded formulation allows for a refined adjustment of the Poisson distribution, enhancing the model's ability to accurately reflect empirical data characteristics such as skewness and kurtosis, thus providing a more robust tool for statistical analysis.

Understanding the convergence properties and relevant limit theorems for these series is crucial for applying them effectively in probabilistic models. The Charlier series, like other orthogonal polynomial expansions, converges under specific conditions. The convergence of the Charlier series $C_n(x; a)$ for a Poisson variable X with parameter a generally depends on the properties of the weight function and the nature of the coefficients used in the expansion (Barbour *et al.*, 1992). First, for any fixed x, the Charlier series approximation converges pointwise to the function it represents if the series of coefficients (c_n) associated with the expansion decreases sufficiently fast. Specifically, if $|c_n| \leq M\rho^n$ for some $M > 0$ and $0 < \rho < 1$, then the series converges pointwise (Billingsley, 1995). In addition, the Charlier series can uniformly converge on sets of integers if the coefficients c_n decay exponentially fast. This is typically ensured when the function being approximated by the series is smooth enough, such as a bounded and continuous function of the Poisson parameter when interpreted over an interval (Stuart and Ord, 1994).

For large a, the Charlier polynomials exhibit behavior similar to Hermite polynomials, which implies that the Charlier polynomial of a suitably normalized Poisson variable converges to the corresponding Hermite polynomial. This relationship underlies the application of the central limit theorem (Glynn and Iglehart, 1990). As $a \to \infty$, and for functions f approximated by the Charlier series $\sum_{n=0}^{\infty} c_n C_n(x; a)$, the normalized sum $(X - a)/\sqrt{a}$ where X is Poisson(a), converges in distribution to a standard normal distribution. This means the Charlier series can be used to approximate functions of Poisson-distributed variables that are normalized to be asymptotically Gaussian, enhancing their utility in statistical inference (Lehmann and Casella, 1998; Johnson *et al.*, 2005). The rate at which the Charlier series converges to the target function or distribution can be quantitatively assessed using the square norm of the remainder of the series. For a given truncation at N, the error $\|f - \sum_{n=0}^{N} c_n C_n(\cdot; a)\|$ depends on both N and a (Shorack and Wellner, 2009). For adequately smooth functions, the error in approximation by the Charlier series can decrease exponentially with N and is inversely proportional to some power of a. This rate is particularly effective when modeling functions with significant deviations from typical Poisson expectations, such as heavy tails or high peak kurtosis (Hall and Heyde, 2014).

3.2. Estimation

The estimation of coefficients $c_0, c_1, \ldots, c_n$ in the Charlier series approximation is integral to accurately modeling the empirical distribution. These coefficients are typically derived through moment matching techniques, where the theoretical moments of the Charlier series, mean, variance, skewness, and kurtosis, are aligned with the empirical moments observed in the data. This alignment is essential for ensuring that the expansion captures the fundamental characteristics of the target distribution.

The coefficients c_i are determined through orthogonal projection, given the theoretical higher order moments $\mu(\cdot)$ of a target distribution $f(\cdot)$:

$$\sum_{\ell=0}^{\infty} C_h(x; a) w(x; a) \sum_{i=0}^{n} c_i C_i(x; a) = \sum_{\ell=0}^{\infty} C_h(x; a) f(x), \quad h = 0, 1, \ldots, n. \tag{3.2}$$

This yields a linear system:

$$\sum_{i=0}^{n} c_i \sum_{\ell=0}^{\infty} w(x; a) C_i(x; a) C_h(x; a) = \sum_{k=0}^{h} \delta_{h,k} \mu(k), \quad h = 0, 1, \ldots, n,$$

where $\delta_{h,k}$ represents the coefficient of x^k in $C_h(x)$. Solving this system provides the coefficients c_h as (Ha and Provost, 2007):

$$c_h = \frac{1}{\rho_h} \sum_{k=0}^{h} \delta_{h,k} \mu(k), \quad h = 0, 1, \ldots, n.$$

The resulting approximation function $f_n(x)$ is formulated as:

$$f_n(x) = w(x; a) \sum_{i=0}^{n} \left(\frac{1}{\rho_i} \sum_{k=0}^{i} \delta_{i,k} \mu(k) \right) C_i(x; a).$$

3.3. Generalization for computation

The Poisson distribution's role as a foundational weight function in generating Charlier polynomials is crucial for approximating target distributions. This method leverages Poisson properties to establish a baseline for aligning theoretical models with empirical data. A linear combination of Charlier polynomials, applied with the Poisson distribution, refines this initial model, acting as a pseudo-polynomial approximant. This process minimizes differences between the target and Poisson-based approximations by adjusting the parameters of the Charlier series to better fit the empirical data, capturing the unique characteristics of the distribution.

A ratio approximant emerges, where the modified Charlier series, aligned with the Poisson weight function, approximates the ratio of two distributions. This sophisticated approach addresses cases where simple Poisson models fail to capture complex data patterns. An appropriate initial approximant is critical. The Poisson distribution, while a useful starting point, often struggles with overdispersion or multimodality in complex data. Generalizing the initial approximant to adaptive members of the Poisson family, like mixtures of Poisson distributions, provides a more flexible framework. This reduces the burden on Charlier polynomial adjustments, enhancing model fidelity and robustness. Incorporating adaptive Poisson family distributions as initial approximants fosters more nuanced statistical tools, advancing probabilistic modeling capabilities to tackle increasingly complex data in scientific and applied fields. This refined approach includes broader Poisson distributions as weight functions, facilitating nuanced modeling of complex data distributions.

The coefficients c_i of the Charlier polynomials are determined through the method of orthogonal projection. For a generalized weight function $\psi(x;\alpha)$ with parameters α, which may represent a mixture of Poisson distributions, the coefficients are computed like equation (3.2) as follows:

$$\sum_{\ell=0}^{\infty} C_h(x;\alpha)\psi(x;\alpha)\sum_{i=0}^{n} c_i C_i(x;\alpha) = \sum_{\ell=0}^{\infty} C_h(x;\alpha) f(x), \quad h = 0,1,\ldots,n. \tag{3.3}$$

The associated Charlier polynomials $C_i(x;\alpha)$ can be obtained via the orthogonality equation, that is,

$$\sum_{x=0}^{\infty} C_m(x;\alpha)C_n(x;\alpha)\psi(x;\alpha) = \theta_{mn},$$

where θ_{mn} is the orthogonality factor. This integral setup accounts for the domain of the Poisson distribution and its extensions, which is from 0 to infinity, given the discrete nature of these distributions. This projection leads to a linear system:

$$\sum_{i=0}^{n} c_i \sum_{\ell=0}^{\infty} \psi(x;\alpha)C_i(x;\alpha)C_h(x;\alpha) = \sum_{k=0}^{h} \delta_{h,k}\mu(k), \quad h = 0,1,\ldots,n,$$

where $\delta_{h,k}$ represents the coefficient of x^k in $C_h(x;\alpha)$. The integral equation reflects the weighted inner products over the expanded support of the Poisson mixtures. Solving this linear system yields the coefficients c_h as:

$$c_h = \frac{1}{\rho_h}\sum_{k=0}^{h} \delta_{h,k}\mu(k), \quad h = 0,1,\ldots,n.$$

The approximation function $f_n(x)$ resulting from this series is then formulated as:

$$f_n(x) = \psi(x;\alpha)\sum_{i=0}^{n}\left(\frac{1}{\rho_i}\sum_{k=0}^{i}\delta_{i,k}\mu(k)\right)C_i(x;\alpha).$$

Given the polynomial nature of $C_i(x)$, which can be expanded as $C_i(x;\alpha) = \sum_{\ell=0}^{i}\delta_{i,\ell}x^{\ell}$, we can reframe the series as (Ha and Provost, 2007):

$$f_n(x) = \psi(x;\alpha)\sum_{\ell=0}^{n}\left(\sum_{i=\ell}^{n}\frac{\delta_{i,\ell}}{\rho_i}\sum_{k=0}^{i}\delta_{i,k}\mu(k)\right)x^{\ell},$$

and thus:

$$f_n(x) = \psi(x;\alpha) \sum_{\ell=0}^{n} \xi_\ell x^\ell,$$

where

$$\xi_\ell = \sum_{i=\ell}^{n} \frac{\delta_{i,\ell}}{\rho_i} \sum_{k=0}^{i} \delta_{i,k}\mu(k). \tag{3.4}$$

Hence, this equation is computationally efficient because it allows for the calculation of coefficients without needing to derive the associated Charlier type orthogonal polynomial series.

3.4. Performance measure and optimal tunning parameter

To measure the difference between a target distribution and an estimated distribution, you might consider using several statistical metrics such as the Kullback-Leibler divergence (KL divergence), mean squared error, or even a simple histogram-based comparison. Here, we will focus on a straightforward implementation using the KL divergence, which measures the information lost when one distribution is used to approximate another. KL divergence is a measure of how one probability distribution diverges from a second, expected probability distribution. For discrete probability distributions P (the theoretical distribution) and Q (the estimated distribution), the KL divergence is given by:

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log\left(\frac{P(i)}{Q(i)}\right),$$

where $P(i)$ and $Q(i)$ are the probabilities of the i^{th} event in the respective distributions. We assume that the probabilities in P and Q are nonzero for all i where $P(i)$ is non-zero to avoid undefined logarithms and the probabilities are normalized, meaning that the sum of all probabilities in each distribution equals one.

To further refine our approach in determining the optimal degree of polynomial adjustment using Charlier orthogonal polynomials, we utilize the KL divergence as a selection criterion. This divergence quantifies the difference between the true distribution and the distribution estimated by successive polynomial approximations (Provost and Ha, 2015).

As we incrementally increase the degree of the polynomial, we compute the KL divergence for each step. Our goal is to identify the degree to which the KL divergence reaches a local minimum for the first time. This local minimum suggests the most effective polynomial degree for approximating the distribution with minimal informational loss. At this point, we achieve an optimal balance between model complexity and accuracy, ensuring the polynomial adjustment is well-tuned to the data without underfitting or overfitting.

4. Numerical examples

We conduct two comprehensive artificial examples to evaluate the performance of the model under pre-defined parameters, focusing on two distinct but widely applicable nonhomogeneous Poisson processes. These processes are instrumental in modeling complex time-varying intensities that are often observed in real-world scenarios. In addition, we conduct a real world data application.

4.1. Mixture of Poisson distributions

We consider a mixture of two Poisson distributions as a statistical model used to describe data that are generated from two different Poisson processes. The parameter π controls the relative contribution of each Poisson distribution to the overall mixture, allowing a wider variety of data characteristics than a single Poisson distribution. This approach is particularly beneficial in real-world scenarios where data might be influenced by multiple underlying processes with different rates of event occurrence. The probability mass function is a weighted sum of the probability mass functions of two individual Poisson distributions. Let us define $\lambda_1 = 5$ (mean rate of occurrences for the first Poisson process), $\lambda_2 = 15$ (mean rate of occurrences for the second Poisson process), and $\pi = 0.3$ (the proportion of the mixture that comes from the first Poisson distribution). Consequently, $1 - \pi$ (0.7 in this case) will be the proportion coming from the second distribution.

The probability mass function of a mixture of two Poisson distributions for observing k events is given by:

$$P(X = k) = \pi \times P_1(X = k) + (1 - \pi) \times P_2(X = k),$$

where $P_1(X = k) = e^{-\lambda_1}(\lambda_1^k/k!)$ and $P_2(X = k) = e^{-\lambda_2}(\lambda_2^k/k!)$.

To obtain the exact moments of a mixture of two Poisson distributions, you typically need to calculate the expected values of the powers of the random variable representing the mixture. Here is the mathematical approach to calculate the n^{th} moment (expected value of X^n) of this mixture. The moment n of a random variable X following a mixture of two Poisson distributions can be calculated using the formula:

$$E[X^n] = \pi \cdot E\left[X_1^n\right] + (1 - \pi) \cdot E\left[X_2^n\right],$$

where $E[X_i^n]$ is the n^{th} moment of the Poisson distribution with the parameter λ_i. For a Poisson distribution with parameter λ, the n^{th} moment can use the following relation based on factorial moments:

$$E\left[X^n\right] = \sum_{k=0}^{n} S(n,k)\lambda^k,$$

where $S(n,k)$ are the Stirling numbers of the second kind, which count the number of ways to partition a set of n objects into k non-empty subsets.

The Kullback-Leibler divergence values presented indicate the degree of approximation accuracy between the exact mixture of two Poisson distributions and approximating distributions of a simple Poisson distribution and the Charlier series approximants.

- Base Poisson distribution: With a KL divergence of 0.537723, the simple Poisson model exhibits a considerable deviation from the exact mixture, indicating a substantial approximation error. This high value suggests that the simple Poisson distribution is significantly limited in capturing the dual-rate nature inherent in the mixture, resulting in a poor fit.

- 6^{th} Charlier series approximant: The approximation accuracy improves with the 6^{th} Charlier series, which yields a KL divergence of 0.23146. This improvement suggests that the 6^{th} order approximant more effectively captures the probability distribution nuances of the mixture compared to the simpler Poisson model.

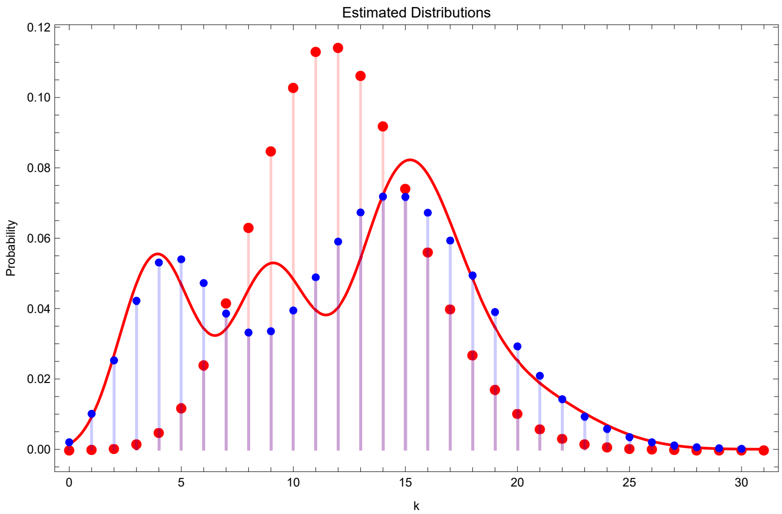

Figure 1: *PMFs: Exact(blue dots), simple Poisson approximant(red dots), and* 8^{th}*-degree Charlier approximant(red line).*

- 8^{th} Charlier series approximant: Demonstrating the best fit, the 8^{th} Charlier series approximant further reduces the KL divergence to 0.0190387. This marked decrease indicates that the 8^{th} order approximant closely replicates the actual characteristics of the mixture, offering a highly precise approximation.

Figure 1 provides a visual representation of these differences. It displays the probability mass functions of the exact mixture alongside those of the simple Poisson distribution and the 8^{th} degree Charlier series approximant. This visual comparison emphasizes the enhanced accuracy achieved through higher-order Charlier approximations in modeling complex probabilistic behaviors.

4.2. Markov-modulated Poisson process

The Markov-modulated Poisson process is leveraged to model systems where the intensity of the Poisson process is governed by an underlying Markov process. In this model, the state transitions dictate the rate changes, making it suitable for environments where the intensity dynamics is influenced by identifiable states of a system. The example aims to determine the model's effectiveness in tracking these transitions and the corresponding impact on event intensities.

The transition matrix **P** in a Markov-Modulated Poisson process is fundamental as it defines the probabilities of transitioning from one state to another in discrete time steps. Each element p_{ij} of the matrix represents the probability of transitioning from state i to state j. For a system with n states, **P** is an $n \times n$ matrix given by:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}.$$

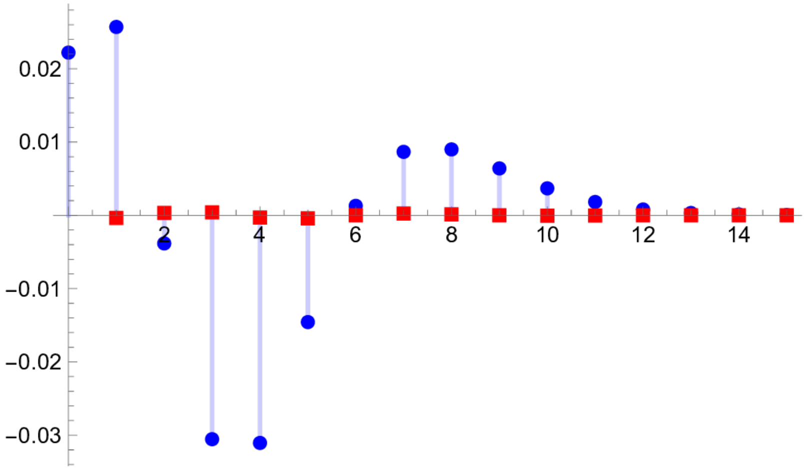

Figure 2: *PMF Differences: Exact and base Poisson approximant (blue dots) vs. exact and 4^{th}-degree Charlier approximant (red squares).*

For the given example with four states, the transition matrix is:

$$\mathbf{P} = \begin{bmatrix} 0.7 & 0.1 & 0.1 & 0.1 \\ 0.2 & 0.5 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.6 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{bmatrix}.$$

Each state i is associated with a Poisson rate λ_i, which determines the intensity of the Poisson process when the system is in state i. The vector of rates for all states is denoted as $\boldsymbol{\lambda}$. For a system with n states, $\boldsymbol{\lambda}$ is:

$$\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_n].$$

For the example with four states, the rates are given by:

$$\boldsymbol{\lambda} = [3, 4, 5, 2].$$

At each time step, the function generates the number of events based on the current state's Poisson rate using the Poisson distribution function. The state transitions are dictated by randomly selecting a new state based on the current state's transition probabilities. The output of this Markov-Modulated Poisson process provides a distribution of the total number of events for multiple trials over a specified period, which reflects the influence of both the stochastic state transitions and the Poisson-distributed event occurrences, showcasing the dynamics of the Markov-Modulated Poisson process.

For a Markov-Modulated Poisson Process with transition matrix $\mathbf{P}$ and rate vector $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_n]$, the cumulant generating function $K_X(\theta, t)$ for the Markov-Modulated Poisson process is given by

$$K_X(\theta, t) = \log\left(\boldsymbol{\alpha}^\top e^{t(\mathbf{Q}+\boldsymbol{\Lambda}(e^\theta-1))}\mathbf{1}\right),$$

where $\mathbf{Q} = \mathbf{P} - \mathbf{I}$ is the generator matrix of the Markov chain, with $\mathbf{I}$ being the identity matrix, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix with the Poisson rates on the diagonal, $\boldsymbol{\alpha}$ is assumed to be the steady-state probabilities to represent the long-term proportion of time the Markov chain spends in each state, that is, a probability vector to satisfy $\mathbf{P} \cdot \boldsymbol{\alpha} = \boldsymbol{\alpha}$ and $\mathbf{1}$ is a column vector of ones, the n^{th} cumulant $\kappa_n(t)$ is obtained by differentiating the cumulant generating function $K_X(\theta, t)$ with respect to θ and evaluating it at $\theta = 0$:

$$\kappa_n(t) = \left. \frac{\partial^n K_X(\theta, t)}{\partial \theta^n} \right|_{\theta=0}.$$

For practical purposes, the first and second cumulants (that is, the mean and variance) are, respectively, $\mu_X(t) = \boldsymbol{\alpha}^\top e^{t(\mathbf{Q}+\boldsymbol{\Lambda})}\mathbf{1}$ and $\sigma_X^2(t) = \boldsymbol{\alpha}^\top (e^{t(\mathbf{Q}+\boldsymbol{\Lambda})}\boldsymbol{\Lambda} e^{t(\mathbf{Q}+\boldsymbol{\Lambda})^\top})\mathbf{1} - (\mu_X(t))^2$. And the third and fourth moments are, respectively, related to the cumulants by $\kappa_3(t) + 3\kappa_2(t)\mu_X(t) + \mu_X^3(t)$ and $\kappa_4(t) + 4\kappa_3(t)\mu_X(t) + 6\kappa_2(t)\mu_X^2(t) + \mu_X^4(t)$.

In the context of assessing the accuracy of approximations for a given distribution, Figure 2 provides a visual comparison of probability mass function differences between the exact and the base Poisson distribution, and between the exact and the 4^{th} degree Charlier series density approximant. The figure likely illustrates how each approximant aligns with the exact distribution visually, giving an intuitive sense of how well each model captures the underlying distributional characteristics.

The exact moments of the Markov Modulated Poisson Process, specifically the zeroth (1), first (3.43966), second (16.5), third (96.181), and fourth (648.491), encapsulate critical statistical properties such as total probability, mean, variance, skewness, and propensity for extreme values. A fundamental component of our analysis involves the base Poisson distribution with a rate parameter equal to the first moment (3.43966). This base distribution is crucial, as it forms the foundation upon which the polynomial adjustments of the Charlier series approximant are developed. By integrating these moments into the Charlier series, we enhance the polynomial representation, leading to a refined expression:

$$1.68994 - 0.575486a + 0.135092a^2 - 0.0141733a^3 + 0.000755002a^4,$$

where a represents the variable in the polynomial. This adjusted series provides a more accurate approximation of the underlying stochastic process.

The KL divergence results for various polynomial approximations to a reference distribution are as follows:

- **Base Poisson (**0.0260418**):** Indicates a minor but noticeable divergence from the reference, reflecting the simplicity of the Poisson model.
- **d = 2 (**0.000442214**):** Shows a significantly improved fit over the base model, with the second-degree polynomial capturing more characteristics of the reference.
- **d = 3 (**0.000490774**):** Slightly worse than d=2, potentially due to overfitting at this level of polynomial complexity.
- **d = 4 (**4.44737×10^{-6}**):** Demonstrates a minimal KL divergence, indicating an excellent approximation to the reference distribution.
- **d = 5 (**5.98039×10^{-6}**):** Although still very close, shows a slight increase in divergence compared to $d = 4$, suggesting diminishing returns with higher polynomial degrees.

These findings highlight that, while polynomial adjustments can significantly enhance the approximation quality, the choice of polynomial degree must balance complexity with performance, with the aim of optimizing the accuracy without introducing unnecessary complexity or instability.

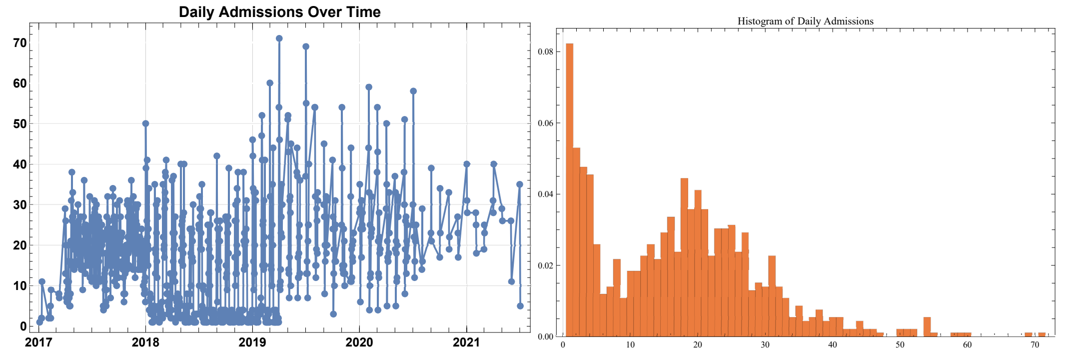

Figure 3: *Daily number of hospital admission over time (left panel) and histogram of daily admission (right panel).*

4.3. Hospital admissions data

Our study employs a publicly available dataset hosted on GitHub in the repository titled [hospital admissions data analysis](https://github.com/kateue/Hospital-Admissions-Data-Analysis). This data set comprises hospital admission records that are ideal for analyzing patterns and intensity variations over time. The data include detailed timestamps of patient admissions, providing a granular view of admission frequencies, which are essential for modeling Poisson processes.

Utilizing this dataset allows us to apply the mixture of Poisson distributions to a real-world scenario where admission rates are likely to vary due to factors such as time of day, day of the week, and public health trends. The left panel of Figure 3 displays the daily number of hospital admissions, illustrating that the occurrence rate varies over time, indicative of a non-homogeneous Poisson process. The right panel demonstrates the complex distributional characteristics of these admissions, further emphasizing the variability in daily rates.

The primary objective of incorporating this dataset is to validate the proposed statistical models by applying them to data that exhibit non-homogeneous properties. We aim to assess the robustness of the models in capturing the dynamics of hospital admissions that are influenced by complex underlying processes.

We first establish an initial approximation using the Poisson distribution framework. The parameters of these models, including the mixture ratios p_i and the rate parameters λ_i of the Poisson distributions, are then optimized using the Newton-Raphson algorithm.

Figure 4 illustrates the mixture of three Poisson distributions and its respective expansions, providing a visual representation of how these statistical models are applied and the effects of their polynomial enhancements. This visualization aids in understanding the incremental benefits and limitations of expanding the model complexity through additional Charlier polynomial terms.

In the mixture model of three Poisson distributions, the estimated parameters represent the mixture weights (probabilities) and the rate parameters λ for each of the Poisson components. The weights of the mixture (p_i) are estimated as $p_1 = 0.167222$, $p_2 = 0.300697$ and $p_3 = 0.532081$. It indicates that approximately 16.72% and 30.07% of the data are expected to be generated from the first and second Poisson components, respectively. Thus, the third component contributes approximately 53.21% of the data, making it the most significant contributor among the three.

And the rate parameters (λ_i) are $\lambda_1 = 35.8614$; this rate parameter for the first Poisson distribution suggests a relatively high frequency of events. A higher λ indicates a higher mean and variance, as

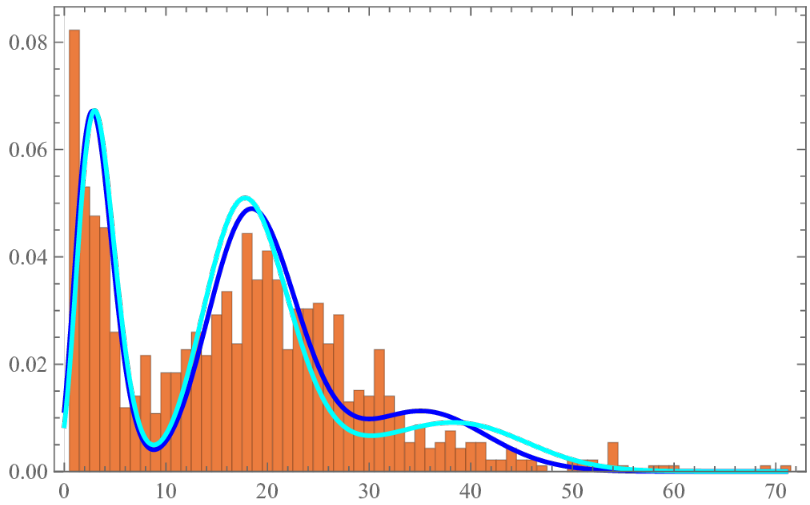

Figure 4: *Hospital admission data: Histogram, mixture of three Poisson distributions (blue line), and 3^{th}-degree series approximant (Cyan line).*

both are equal to λ in a Poisson distribution. This component likely represents a segment of the data with frequent occurrences. And the rate parameter $\lambda_2 = 3.27494$ for the second Poisson distribution is significantly lower than that of the first, indicating fewer events. This component accounts for a smaller portion of the data, characterized by fewer frequent occurrences. And the rate $\lambda_3 = 18.9284$ for the third Poisson distribution falls between the other two, suggesting a moderate level of event occurrences.

The estimated parameters of this mixture model provide insight into the distributional characteristics of the data set. The different λ values highlight the presence of varied behavior within the dataset, with some subsets of the data showing high frequencies of events and others showing much lower frequencies. Meanwhile, the mixture weights p_i help quantify the proportion of the dataset attributed to each behavior type. Such a model is particularly useful in scenarios where the data is heterogeneous, arising from multiple underlying processes that cannot be adequately described by a single Poisson distribution.

We generate the associated Charlier-type orthogonal polynomials, which facilitate the development of series approximations through linear combinations. It is important to note that the effectiveness of the linear combination adjustments depends significantly on the quality of the initial approximation. If the initial model already closely approximates the target data, further adjustments made by adding linear combinations may offer diminishing improvements. This is due to the inherent limitations in enhancing a model that is already a good fit, which consequently reduces the impact of additional polynomial series terms. The coefficients ξ_i for $i = 0, 1, 2,$ and 3 of the approximant of the probability mass function of the third degree series, as specified in equation (3.4), were estimated as follows: $\xi_0 = 0.74964$, $\xi_1 = 0.107293$, $\xi_2 = -0.00714586$, and $\xi_3 = 0.000116469$.

The results using the difference measure reveal that among various models, the third-degree polynomial adjustment (d = 3) with the lowest difference measure value (0.0000434733) provides the best fit to the data, indicating superior accuracy in capturing complex patterns without overfitting. Conversely, the second-degree (d = 2) model with a difference measure value of 0.00226874 and the fourth-degree (d = 4) model with a value of 0.00136258 show poorer fits, suggesting they either underfit or introduce unnecessary complexity.

5. Concluding remarks

This study introduced the Charlier series that approximates probability mass functions as a powerful tool to improve statistical models. Using Charlier polynomials, we can better represent probability mass functions, increasing accuracy, and addressing overdispersion issues in Poisson processes.

We also demonstrated how advanced statistical models, such as the mixture of Poisson distributions and the Markov-modulated Poisson process, are effective for analyzing complex data, such as hospital admission rates. These models have been tested through artificial examples and real-world applications, showing their ability to capture dynamic patterns.

However, fully integrating Charlier polynomials with mixed Poisson models is still a developing area of research. This integration could lead to significant improvements in model accuracy. Advancing polynomial expansions within Poisson distributions could transform data analysis tools, fulfilling the need for sophisticated techniques that can handle complex patterns.

The analyzed models have provided clear insights into data variability and the precision of complex approximations compared to simpler models. This underscores the importance of choosing the right model for different analytical scenarios and pushes for the adoption of more advanced, tailored statistical methods that incorporate polynomial expansions.

References

Barbour AD, Holst L, and Janson S (1992). *Poisson Approximation*, Oxford University Press, New York.

Billingsley P (1995). *Probability and Measure*, John Wiley and Sons, New York.

Fokianos K, Rahbek A, and Tjøstheim D (2009). Poisson autoregression, *Journal of the American Statistical Association*, **104**, 1430–1439.

Glynn PW and Iglehart DL (1990). Simulation output analysis using standardized time series, *Mathematics of Operations Research*, **15**, 1–16.

Grandell J (1997). *Mixed Poisson Processes*, Chapman and Hall/CRC, London.

Ha H-T and Provost SB (2007). A viable alternative to resorting to statistical tables, *Communications in Statistics—Simulation and Computation*, **36**, 1135–1151.

Hall P and Heyde CC (2014). *Martingale Limit Theory and Its Application*, Academic Press, New York.

Johnson NL, Kemp AW, and Kotz S (2005). *Univariate Discrete Distributions*, John Wiley and Sons, Hoboken, New Jersey.

Karlis D (2005). EM algorithm for mixed Poisson and other discrete distributions, *Astin Bulletin*, **35**, 3–24.

Kokonendji C, Senga Kiessé A, and Zocchi B (2010). Discrete distributions in statistical modelling for insurance data, *Insurance: Mathematics and Economics*, **47**, 217–225.

Lehmann EL and Casella G (1998). *Theory of Point Estimation*, Springer, New York.

Provost SB and Ha H-T (2015). Distribution approximation and modeling via orthogonal polynomial sequences, *Statistics*, **50**, 454–470.

Shorack GR and Wellner JA (2009). *Empirical Processes with Applications to Statistics*, SIAM, Philadelphia, Pa.

Stuart A and Ord JK (1994). *Kendall's Advanced Theory of Statistics, Volume 1: Distribution Theory*, Wiley.

Sundt R and Vernic A (2009). *Recursions for Convolutions and Compound Distributions with Insurance Applications*, Springer, Dordrecht.

Received July 05, 2024; Revised August 14, 2024; Accepted August 18, 2024