١١٤٩٦٨

دانشگاه اصفهان

دانشکده علوم

گروه آمار

پایان نامه‌ی کارشناسی ارشد رشته‌ی آمار گرایش آمار ریاضی

مطالعه‌ای بر توزیع پواسن تعمیم‌یافته

استاد راهنما :

دکتر محمد حسین علامت‌ساز

پژوهشگر:

سید محمد جواد طاوسی

اسفندماه ۱۳۸۷

۱۳۸۸/۴/ ۶

مرکز اطلاعات و مدارک علمی ایران

۱۱۴۹۶۸

کلیه حقوق مادی مترتب بر نتایج مطالعات ابتکارات
و نوآوری‌های ناشی از تحقیق موضوع این پایان‌نامه
متعلق به دانشگاه اصفهان است.

شیوه نگارش پایان نامه
رعایت شده است.
تحصیلات تکمیلی دانشگاه اصفهان

دانشگاه اصفهان

دانشکده علوم

گروه آمار

پایان‌نامه‌ی کارشناسی ارشد رشته‌ی آمارگرایش آمار- ریاضی

سید محمد جواد طاووسی

تحت عنوان

مطالعه ای بر توزیع پواسن تعمیم یافته

در تاریخ ۸۷/۱۲/۱۰ توسط هیأت داوران زیر بررسی با نمره ۱۸/۰۳ و با درجه عالی به تصویب نهایی رسید.

۱- استاد راهنمای پایان‌نامه دکتر محمد حسین علامت ساز با مرتبه‌ی علمی استاد

امضاء

۲- استادداور داخل گروه پایان‌نامه دکتر ایرج کاظمی با مرتبه‌ی علمی استاد یار

امضاء

۳- استاد داورخارج از گروه دکتر جعفر احمدی با مرتبه‌ی علمی دانشیار

امضاء

امضای مدیر گروه

تقـــدیـــر و تشـــکر

منّت خدای را عزّوجلّ که طاعتش موجب قربت است و به شکر اندرش مزید نعمت. هر نفسی که فرو می‌رود، ممّد حیات است و چون بر می آید، مفرّح ذات. پس در نفسی دو نعمت موجود است و بر هر نعمتی شکر واجب.

از دست و زبان که برآید کز عهده شکرش به در آید

شکر و سپاس بی‌کران تنها خدای یکتا را سزاست که انسان را آفرید و به او عقل را عطا نمود و بدینوسیله او را اشرف مخلوقات خویش قرار داد. و انسان این خلیفه خداوند بر روی زمین توانست با بهره‌گیری از این امانت گرانبها طبیعت را مسخر خویش ساخته و مسائلی بسیار پیچیده را در روی این کره خاکی حل نماید. اینک که به لطف حضرت حق، نگارش این پایان‌نامه به اتمام رسیده است ضمن سپاسگزاری به درگاه باریتعالی و سپاس ویژه از عنایات حجت برحقش بقیه ا... الاعظم حضرت مهدی(عج) و سلطان عشق ولی نعمت ایرانیان حضرت امام رضا(ع)، لازم می‌دانم که از استاد بزرگوارم جناب آقای دکتر علامت‌ساز که دقت نظر و تأمل درخور تحسین ایشان در انجام این کار همواره برای من راهنما و راهگشایی ارزنده بوده است، نهایت سپاسگزاری و قدردانی را به عمل آورده و برای ایشان آرزوی موفقیت و سلامت نمایم.

همچنین از اساتید محترم جناب آقای دکتر کاظمی (از دانشگاه اصفهان) و جناب آقای دکتر احمدی (از دانشگاه فردوسی مشهد مقدس) که زحمت داوری این پایان نامه را متحمل شدند کمال تشکر خود را ابراز داشته و توفیق روزافزونشان را خواستارم.

از خانواده عزیز و دلسوزم و همچنین کلیه دوستان عزیزی که به نحوی در ارائه این پایان‌نامه به من کمک کردند صمیمانه متشکرم. همچنین در پایان از تمامی اساتیدی که در طول دوران تحصیل دلسوزانه چراغ راه من بوده و از شمع وجودشان بهره‌مند شده‌ام صمیمانه سپاسگزارم و موفقیت، بهروزی و سعادت آنها را از ایزد متعال خواستارم.

س.م.ج. طاوسی

اسفند ۱۳۸۷

تقـدیـم بـه

پیشگاه مقدس یگانه قطب دایره امکان حضرت بقیه آ... الاعظم(عج) روحی و ارواح العالمین لتراب مقدمه الفداء

و

همه اندیشمندان، خرد ورزان و علم اندوزان گیتی

و

پدر و مادر مهربان، دلسوز و فداکارم

و

همسر عزیزم.

چکیده

مدل پواسون معمولی برای مدل بندی انواع مختلف داده‌های شمارشی که می‌دانیم یا حدس می‌زنیم نسبت واریانس به میانگین جامعه (شاخص پراکندگی) متفاوت از یک است، اغلب کارایی ندارد. دراین حالات یک مدل جایگزین، توزیع پواسون تعمیم‌یافته است. این توزیع علاوه برسادگی، به دلیل داشتن دوپارامتر، از انعطاف پذیری بالایی برخوردار است. بسته به اینکه پارامتر دوم آن مثبت یا منفی باشد، دارای خاصیت بیش پراکنش یا کم پراکنش است. وقتی مقدار پارامتر دوم صفر باشد، این توزیع به شکل پواسون معمولی در می‌آید. این توزیع کاربردهای زیادی در بسیاری از علوم و حوزه‌های مطالعه شامل مهندسی، تحلیل بقا، زیست شناسی، ژنتیک و فرآیندهای شاخه‌ای دارد.

در این تحقیق توزیع پواسون تعمیم‌یافته را به اختصار معرفی کرده، به بیان خاستگاه‌های آن پرداخته و برخی خواص آماری آن را بررسی می‌نماییم. همچنین اثبات دیگری برای توزیع احتمال بودن آن براساس لم تفاضلی اویلر ارائه نموده و ثابت می‌کنیم که این توزیع، آمیخته‌ای از توزیع پواسون و همچنین توزیعی مرکب است. شیوه‌های برآورد و آزمون‌های فرض پارامترها و نیکویی برازش توزیع پواسن تعمیم‌یافته را همراه با مثال‌های عددی ارائه نموده و در برازش داده‌های شمارشی به دلیل شباهت این توزیع با توزیع دوجمله‌ای منفی به مقایسه برخی خواص این دو توزیع و مقایسه مدل‌های برازشی آن دو برای داده‌های واقعی می‌پردازیم. در نهایت نیز به مطالعه برخی دیگر از مدل‌ها و تعمیم‌های کلی‌تری از این توزیع می‌پردازیم.

هدف از انجام این تحقیق آشنایی بیشتر با توزیع پواسون تعمیم‌یافته، خواستگاه، ویژگی‌های آماری و توزیعی آن می‌باشد. شناخت آزمون‌های نیکویی برازش، برآوردهای این توزیع و این که تحت چه شرایطی استفاده از این توزیع نسبت به توزیع دو جمله‌ای منفی (به عنوان یک رقیب) برای برازش مدل با داده‌های شمارشی بهتر می‌باشد نیز از دیگر اهداف این پایان نامه است.

واژگان کلیدی : آمیخته پواسن، برآورد ماکسیمم درستنمایی، بیش پراکنش، تابع توزیع تجربی، توزیع پواسن تعمیم‌یافته، توزیع پواسن تعمیم‌یافته مرکب، توزیع پواسن مرکب، کم پراکنش.

فهرست مطالب

عنوان صفحه

فصل اول : مقدمه

۱-۱ موضوع و پیشینه تحقیق ۱

۱-۲ اهداف تحقیق ۴

۱-۳ ساختار پایان نامه ۴

فصل دوم : توزیع پواسن تعمیم‌یافته و خاستگاه آن

۲-۱ مقدمه ۵

۲-۲ توزیع پواسن تعمیم‌یافته (GPD) ۶

۲-۳ نمایش ترسیمی توزیع پواسن تعمیم‌یافته ۱۱

۲-۴ خاستگاه توزیع پواسن تعمیم‌یافته ۱۵

۲-۴-۱ حد توزیع دوجمله‌ای منفی تعمیم‌یافته(GNBD) ۱۵

۲-۴-۲ حد توزیع‌های شبه دوجمله‌ای(QBD) ۱۶

۲-۴-۳ حد توزیع مارکف-پولیای تعمیم‌یافته ۱۷

۲-۴-۴ فرایند پواسن تعمیم‌یافته ۱۷

۲-۴-۵ GPD به عنوان توزیع تعداد متقاضیانی که در یک دوره اشتغال سرویس می‌گیرند (کاربرد در فرآیند صف) ۲۲

۲-۴-۶ GPD به عنوان یک توزیع آمیخته پواسن ۲۴

۲-۴-۷ GPD به عنوان یک توزیع پواسن مرکب ۲۷

۲-۴-۸ GPD بر اساس معادلات دیفرانسیل تفاضلی ۲۹

فصل سوم : برخی خواص توزیع پواسن تعمیم‌یافته

۳-۱ مقدمه ۳۴

۳-۲ بسط لاگرانژ ۳۴

عنوان | صفحه

۳-۳ میانگین و واریانس توزیع پواسن تعمیم‌یافته ۳۶
۳-۴ تابع مولد احتمال توزیع پواسن تعمیم‌یافته ۳۸
۳-۵ تک مدی بودن مدل‌های توزیع پواسون تعمیم‌یافته ۳۸
۳-۶ تابع مولد گشتاور توزیع پواسن تعمیم‌یافته ۴۱
۳-۷ ضریب چولگی و کشیدگی توزیع پواسون تعمیم‌یافته ۴۳
۳-۸ رابطه بین احتمال‌های دقیق و تجمعی ۴۴
۳-۹ امید ریاضی انحراف از میانگین توزیع پواسن تعمیم‌یافته ۴۵
۳-۱۰ گشتاورهای صحیح منفی(وارون) ۴۵
۳-۱۱ ماتریس اطلاع فیشر ۴۶
۳-۱۲ رفتار دنباله‌ای GPD ۴۹
۳-۱۳ خاصیت پیچش GPD ۴۹
۳-۱۴ تفاضل دو متغیر GP ۵۰
۳-۱۵ قدر مطلق تفاضل دو متغیر GP ۵۱
۳-۱۶ ارتباط با توزیع‌های نرمال و گوسی معکوس ۵۲

فصل چهارم : استنباط کلاسیک در توزیع پواسن تعمیم‌یافته و کاربردهایی از این توزیع

۴-۱ مقدمه ۵۵
۴-۲ برآورد نقطه‌ای پارامترهای توزیع پواسون تعمیم‌یافته ۵۶
۴-۲-۱ برآورد گشتاوری ۵۶
۴-۲-۲ برآورد بر اساس میانگین نمونه و فراوانی اول (طبقه صفر) ۵۷
۴-۲-۳ برآورد ماکسیمم درستنمایی (ML) ۵۷
۴-۲-۴ روش برآورد اختلافات وزنی (WD) ۶۵
۴-۲-۵ روش برآورد نرخ های تغییر وزنی تجربی (EWRC) ۷۴
۴-۳ برآورد فاصله‌ای پارامترهای توزیع پواسون تعمیم‌یافته ۷۸

عنوان | صفحه

۴-۳-۱ برآورد فاصله‌ای برای λ زمانی که θ یا ω معلوم است ۷۸

۴-۳-۲ برآورد فاصله‌ای برای θ یا ω زمانی که λ معلوم است ۸۱

۴-۴ آزمون فرض در توزیع پواسون تعمیم‌یافته ۸۲

۴-۴-۱ آزمون برای λ ۸۲

۴-۴-۲ آزمون برای θ ۸۴

۴-۴-۳ آزمون توزیع پواسون تعمیم‌یافته در برابر توزیع پواسون ۸۶

۴-۵ آزمون نیکویی برازش توزیع پواسون تعمیم‌یافته ۸۷

۴-۵-۱ آزمون خی‌دو ۸۷

۴-۵-۲ آزمون‌های نیکویی برازش تابع توزیع تجربی (EDF) ۹۵

۴-۶ مقایسه توزیع پواسن تعمیم‌یافته با توزیع دوجمله‌ای منفی ۱۰۵

۴-۶-۱ بررسی نظری ۱۰۵

۴-۶-۲ مثال عددی ۱۱۲

فصل پنجم : مدل‌های دیگری از توزیع پواسن تعمیم‌یافته

۵-۱ مقدمه ۱۱۷

۵-۲ توزیع پواسن تعمیم‌یافته بریده شده ۱۱۸

۵-۳ توزیع پواسن تعمیم‌یافته مرکب ۱۱۹

۵-۳-۱ تعریف توزیع ۱۱۹

۵-۳-۲ گشتاورهای مرکزی توزیع ۱۲۰

۵-۳-۳ یک خاصیت CGPD ۱۲۰

۵-۴ خانواده کلی توسعه یافته توزیع‌های پواسن تعمیم‌یافته (EGCGPD) ۱۲۲

۵-۴-۱ تعریف ۱۲۲

۵-۴-۲ رابطه بازگشتی برای $C_j(\lambda_1, \theta)$ ۱۲۲

۵-۴-۳ گشتاورهای فاکتوریل EGCGPD ۱۲۳

عنوان

صفحه

۵-۴-۴ میانگین EGCGPD-II ۱۲۳

۵-۴-۵ تابع درستنمایی ۱۲۳

۵-۴-۶ حالات خاص ۱۲۴

۵-۴-۷ ویژگی‌های EGCGPD ۱۲۴

بحث و پیشنهاد در باب ادامه تحقیق ۱۲۶

واژه نامه فارسی به انگلیسی ۱۲۷

منابع و مآخذ ۱۳۰

فهرست شکل‌ها

عنوان **صفحه**

شکل ۲-۱ (الف) نمودار توابع احتمال توزیع پواسن تعمیم‌یافته به ازای $\theta = -0/1$ و مقادیر مختلف λ به طور همزمان ۱۱

شکل ۲-۱ (ب) نمودار توابع احتمال توزیع پواسن تعمیم‌یافته به ازای $\theta = -0/1$ و مقادیر مختلف λ به طور مجزا ۱۲

شکل ۲-۲ (الف) نمودار توابع احتمال توزیع پواسن تعمیم‌یافته به ازای $\lambda = 8$ و مقادیر مختلف θ به طور مجزا ۱۳

شکل ۲-۲ (ب) نمودار توابع احتمال توزیع پواسن تعمیم‌یافته به ازای $\lambda = 8$ و مقادیر مختلف θ به طور همزمان ۱۴

شکل ۳-۱ نمودار تابع جرم احتمال GPD به ازای $\lambda = 15$ و $\theta = -0/3$ ۵۳

شکل ۳-۲ نمودار تابع جرم احتمال GPD به ازای $\lambda = 15$ و $\theta = -0/1$ ۵۳

شکل ۳-۳ نمودار تابع جرم احتمال GPD به ازای $\lambda = 15$ و $\theta = 0/1$ ۵۴

شکل ۴-۱ نمودار تابع $H(\theta)$ برای پیشامد E_1 در حالتی که $\bar{x} \leq 2$ ۶۲

شکل ۴-۲ نمودار تابع $H(\theta)$ برای پیشامد E_1 در حالتی که $j \geq 2,\ j < \bar{x} < j+1 \leq k$ ۶۴

شکل ۴-۳ (الف) نمودار اریبی برآوردگرها برای $\lambda = 0/8$ ۷۶

شکل ۴-۳ (ب) نمودار اریبی برآوردگرها برای $\theta = 0/2$ ۷۶

شکل ۴-۴ (الف) نمودار اریبی برآوردگرها برای $\lambda = 3$ ۷۶

شکل ۴-۴ (ب) نمودار اریبی برآوردگرها برای $\theta = 0/4$ ۷۶

شکل ۴-۵ (الف) نمودار اریبی برآوردگرها برای $\lambda = 5$ ۷۷

شکل ۴-۵ (ب) نمودار اریبی برآوردگرها برای $\theta = 0/2$ ۷۷

شکل ۴-۶ (الف) نمودار اریبی برآوردگرها برای $\lambda = 7$ ۷۷

عنوان | صفحه

شکل ۴-۶ (ب) نمودار اریبی برآوردگرها برای $\theta = ۰/۲$ ۷۷

شکل ۴-۷ نمودار توزیع فراوانی‌های داده‌های شپشک چوب ۸۹

شکل ۴-۸ نمودار فراوانی‌های برآورد شده به روش ماکسیمم درستنمایی برای توزیع پواسون ۹۱

شکل ۴-۹ نمودار فراوانی‌های برآوردشده به روش ماکسیمم درستنمایی برای توزیع پواسون تعمیم‌یافته ۹۱

شکل ۴-۱۰ نمودارتوزیع داده‌های تعداد خسارات بیمه کنندگان اتومبیل سال ۱۳۸۶ استان مازندران ۹۳

شکل ۴-۱۱ نمودارهایی از توابع جرم احتمال GP و NB : خط ثابت، توزیع GP را نشان می‌دهد و خط نقطه چین نشان دهنده NB است. در ردیفهای ۱تا۴، به ترتیب $\mu = ۲۵$، $\mu = ۱۵$، $\mu = ۵$ و $\mu = ۰/۵$ می باشد. ستون های ۱تا۳ به ترتیب $D = ۱,۵$، $D = ۱۰$ و $D = ۳۰$ را دارند (محور عمودی نشان دهنده تابع جرم احتمال است) ۱۱۱

فهرست جدول‌ها

عنوان **صفحه**

جدول ۴-۱ توزیع فراوانی‌های تعداد دستبردهای پنج ماه متوالی در شهر A ۶۱

جدول ۴-۲ توزیع فراوانی‌های تعداد دستبردهای پنج ماه متوالی در شهر B ۶۳

جدول ۴-۳ مقایسه روش‌های ML، WD، MC و EWRC برای برآورد پارامترهای GPD ۶۹-۷۲

جدول ۴-۴ مثالی از طبقه‌های بیشتری از داده‌های شبیه سازی شده GPD و برآورد پارامترها ۷۳

جدول ۴-۵ مثال دیگری از طبقه‌های بیشتری از داده‌های شبیه سازی شده GPD و برآورد پارامترها ۷۳

جدول ۴-۶ توزیع ۴۰۲ شپشک چوب در ۱۲۲ تخته ۸۸

جدول ۴-۷ آزمون نیکویی برازش مدل پواسون معمولی به داده‌های شپشک چوب ۹۰

جدول ۴-۸ آزمون نیکویی برازش مدل پواسن تعمیم‌یافته به داده‌های شپشک چوب ۹۰

جدول ۴-۹ برآورد پارامترهای توزیع پواسون تعمیم‌یافته با استفاده از روش درستنمایی برای داده‌های شپشک چوب ۹۰

جدول ۴-۱۰ فراوانی‌های مشاهده شده و برآورد شده تحت مدل‌های پواسون معمولی و پواسون تعمیم‌یافته ۹۲

جدول ۴-۱۱ توزیع داده‌های تعداد خسارات بیمه کنندگان اتومبیل سال ۱۳۸۶ استان مازندران .. ۹۳

جدول ۴-۱۲ آزمون نیکویی برازش مدل پواسون به داده‌های تعداد خسارات بیمه کنندگان اتومبیل سال ۱۳۸۶ استان مازندران ۹۴

جدول ۴-۱۳ آزمون نیکویی برازش مدل GP به داده‌های تعداد خسارات بیمه کنندگان اتومبیل سال ۱۳۸۶ استان مازندران ۹۴

جدول ۴-۱۴ سطوح تجربی برای آزمون‌های EDF وقتی $\alpha = ۰/۱$ ۱۰۱

جدول ۴-۱۵ مقایسه توان (به درصد) برای آزمون‌های EDF وقتی $\alpha = ۰/۰۵$ و $n = ۲۵$ ۱۰۲

جدول ۴-۱۶ مقایسه توان (به درصد) برای آزمون‌های EDF وقتی $\alpha = ۰/۰۵$ و $n = ۵۰$ ۱۰۲

جدول ۴-۱۷ آزمون‌های EDF برای داده‌های تصادف در سطح $\alpha = ۰/۰۵$ ۱۰۳

عنوان | صفحه

جدول ۴-۱۸ فراوانی‌های تعداد تومورهای ۱۵۸ بیمار NF2 ..... ۱۱۳

جدول ۴-۱۹ آماره‌های داده های شمارشی تومور ۱۱۳

جدول ۴-۲۰ آماره‌های مقایسه‌ای MLE برای ۴ مدل (NB، GP، ZINB و ZIGP) ۱۱۵

جدول ۵-۱ حالات خاص $1/C_j(\lambda_1,\theta)$ در EGCGPD ۱۲۴

مخفف‌ها

WD	Weighted discrepancies	اختلافات وزنی
MLE	Maximum likelihood estimation	برآورد ماکسیمم درستنمایی
GP	Generalized Poisson	پواسن تعمیم‌یافته
EDF	Empirical distribution function	تابع توزیع تجربی
cdf	cumulative distribution function	تابع توزیع تجمعی
pmf	probability mass function	تابع جرم احتمال
pdf	probability density function	تابع چگالی احتمال
ch.f	characteristic function	تابع مشخصه
pgf	probability generating function	تابع مولد احتمال
cgf	cumulant generating function	تابع مولد انباشتک
mgf	moment generating function	تابع مولد گشتاور
LT	Laplace transformation	تبدیل لاپلاس
LPD	Lagrangian probability distribution	توزیع احتمال لاگرانژی
GPD	Generalized Poisson distribution	توزیع پواسون تعمیم‌یافته
ZIGPD	Zero-inflated generalized Poisson distribution	توزیع پواسن تعمیم‌یافته صفرآماسیده
CGPD	Compound generalized Poisson distribution	توزیع پواسون تعمیم‌یافته مرکب
NBD	Negative binomial distribution	توزیع دوجمله‌ای منفی
ZINBD	Zero-inflated negative binomial distribution	توزیع دوجمله‌ای منفی صفر آماسیده

توزیع دوجمله‌ای منفی تعمیم‌یافته

GNBD	Generalized negative binomial distribution	
QBD	Quasi-binomial distribution	توزیع شبه دوجمله‌ای

خانواده کلی توسعه یافته توزیع‌های پواسون تعمیم‌یافته

EGCGPD	Extended general class of generalized Poisson distribution	
CV	Coefficient of variation	ضریب تغییرات
CI	Confidence interval	فاصله اطمینان
ML	Maximum likelihood	ماکسیمم درستنمایی
r.v.	random variable	متغیر تصادفی
i.i.d.	independent and identically distributed	مستقل و همتوزیع
MC	Minimum Chi-square	مینیمم خی‌دو
EWRC	Empirical weighted rates of change	نرخ های تغییر وزنی تجربی
GLIM	Generalized linear model software	نرم‌افزار مدل‌های خطی تعمیم‌یافته
LLR	Logarithm likelihood ratio	نسبت لگاریتم درستنمایی

فصل اول

مقدمه

۱-۱ موضوع و پیشینه تحقیق

نظریه توزیع‌ها یکی از مباحث مهم آمار است. آشنایی هرچه بیشتر با توزیع‌ها و خواص آنها ما را در کاربرد بهتر و دقیق تر آنها یاری خواهد داد. یک موضوع اساسی در زمینه تحلیل داده‌های شمارشی، انتخاب و تعیین توزیع مناسب جهت برآورد و پیش بینی رفتار توزیع است. برای تعیین مدل مناسب آماری برای داده‌ها و تجزیه و تحلیل‌های مربوطه لازم است که در ابتدا توزیع‌هایی که می‌توانند به‌عنوان برازش روی داده‌ها به‌کار روند را مورد بررسی قرارداده و مناسب‌ترین مدل را انتخاب کرد. با شناخت توزیع، مهم‌ترین استنباط‌های آماری در رابطه با پارامتر مدل مورد استفاده و برآورد آنها است.

مدل احتمال پواسون الگویی بسیار مناسب برای بسیاری از پدیده‌های تصادفی فراهم می‌کند. ازآنجا که در رده‌ی توزیع‌های سری توانی[1] با تکیه گاه مشتمل بر اعداد صحیحی نامنفی، توزیع پواسون است که دارای میانگین و واریانس برابری است، هر پدیده تصادفی را که در آن نوعی شمارش مورد توجه باشد، در صورت برابری میانگین و واریانس، می توان با استفاده از توزیع پواسون مدل بندی کرد. این شمارش ممکن است تعداد تصادفات رانندگی

[1] Power series distributions

منجر به مرگ در طی یک سال، تعداد ذرات رادیو اکتیو در واحد زمان، تعداد مشتریان وارد شده به یک اداره، تعداد مرگ و میر در بازه زمانی معین و یا به طور کلی تعداد پیشامدهای رخ داده در واحد زمان یا مکان باشد. در تمام این موارد چنین فرض می‌شود که احتمال وقوع یک پیشامد در طول واحد بازه زمان یا مکان ثابت می‌ماند. همچنین وقوع یا عدم وقوع یک پیشامد در بازه ای از زمان یا مکان، مستقل از رخداد یا عدم رخداد پیشامدهای دیگر در آن بازه است و احتمال وقوع دو پیشامد یا بیشتر در بازه زمانی کوچک، قابل اغماض است. تعداد پیشامدها در بازه‌های ناهمپوش نیز مستقل‌اند. بر طبق این فرض‌ها که اصول موضوعه یک فرایند پواسون همگن را تشکیل می‌دهند، تعداد رخدادهای یک پیشامد در دوره‌ای از زمان یا مکانی خاص دارای توزیع پواسون با پارامتر ثابت λ است که λ میانگین تعداد وقوع پیشامد در آن دوره است و با واریانس توزیع برابر می باشد. اما مدل پواسون معمولی برای مدل بندی انواع مختلف داده‌های شمارشی که می‌دانیم یا حدس می‌زنیم نسبت واریانس به میانگین جامعه یا شاخص پراکندگی آن متفاوت از یک است، اغلب ناکارا است. برای مثال رفتار اجتماعی بسیاری از حشرات باعث می‌شود که آنها در دسته‌های چندتایی زندگی کنند. لذا مدل پواسون قادر به نمایش این داده‌ها نمی‌باشد و اغلب برازش نامناسبی به این داده‌ها ارائه می‌دهد. در این حالات مدل‌های مختلفی از جمله دوجمله‌ای منفی کاربرد دارند اما یک مدل جایگزین بسیار خوب برای توزیع پواسون استاندارد، *توزیع پواسون تعمیم یافته*[1] (GPD) است. توزیع پواسون تعمیم یافته علاوه بر سادگی، به دلیل داشتن دو پارامتر λ و θ، از انعطاف پذیری بالایی برخوردار است. بسته به اینکه پارامتر دوم آن مثبت یا منفی باشد، دارای خاصیت بیش پراکنش یا کم پراکنش است. این توزیع وقتی مقدار پارامتر دوم صفر باشد به شکل پواسون معمولی در می‌آید. میانگین و واریانس، هر دو با افزایش λ افزایش یافته، با کاهش آن کاهش می‌یابند. همچنین وقتی θ مثبت است، میانگین و واریانس با افزایش θ نیز افزایش می‌یابند اما واریانس سریعتر افزایش می‌یابد. در نتیجه این توزیع می‌تواند برازش بسیار خوبی را به داده‌های شمارشی ارائه دهد. جاناردان و شفر (۱۹۷۷) گزارش دادند که در برازش مدل پواسون تعمیم‌یافته به ۹۲ مجموعه از داده‌های زیستی از جمله داده‌های مربوط به ناهنجاری‌های کروموزومی در گلبول‌های سفید انسان، در ۸۹ مورد مدل پواسون تعمیم‌یافته بسیار عالی برازش داده شده است. این توزیع در بسیاری از علوم از جمله کشاورزی، اقتصاد، مهندسی، صنعتی(ساخت)، پزشکی، زیست شناسی، بوم شناسی، ژنتیک، تحلیل بقا، بازاریابی، نظریه صف و فرآیندهای شاخه‌ای کاربرد دارد. خواص گوناگون و کاربردهای دیگر این مدل را می‌توان در کارهایی از کانسول و شنتون (۱۹۷۳)، چارلمبایدز (۱۹۷۴)،

[1] Generalized Poisson Distribution

جاناردان و همکاران(۱۹۷۹)، کومار و کانسول (۱۹۸۰) و کانسول (۱۹۸۷) یافت. از زمینه‌های کاربرد بسیار مهم این توزیع می‌توان باروری خانوار، مدل‌بندی داده‌های تصادف، ترخیص‌های بیمارستانی، تحلیل رگرسیون و نظریه ریسک بیمه را برشمرد.

توزیع پواسن به شیوه‌های متعددی تعمیم‌یافته است. کانسول و جین در سال ۱۹۷۳ یک توزیع پواسن تعمیم‌یافته جدید را ارائه کردند که بعدها به وسیله کانسول و شکری (۱۹۸۵) اصلاح شد. کمپ (۱۹۸۶) نشان داد که روش برآورد ماکسیمم درستنمایی(ML[1])، از بین روش‌های دیگر برآورد پارامترها، می‌تواند به عنوان روش نمره گذاری (امتیاز بندی) با استفاده از مجموع‌های وزنی اختلافات میان فراوانی‌های مشاهده شده و مورد انتظار مورد ملاحظه قرار گیرد. نمونه‌هایی که وزن‌های آنها، (الف) وابسته به پارامترها نیستند، (ب) وابسته به پارامترها هستند و یا (ج) هم به پارامترها و هم به فراوانی‌های مشاهده شده وابسته هستند؛ در مقاله کمپ آورده شده است. کانسول (۱۹۸۸) برخی از مدل‌هایی(براساس معادلات دیفرانسیل تفاضلی) که منجر به تعمیمی از توزیع پواسن تحت عناوین توزیع پواسن تعمیم‌یافته، توزیع پواسن لاگرانژی، توزیع دنباله‌ها با توان‌های تعدیل شده و فرایند پواسن تعمیم‌یافته می‌شوند را ارائه کرد. سپس کانسول و فامویه (۱۹۸۸) برآورد ماکسیمم درستنمایی برای این توزیع هنگامی که میانگین نمونه بزرگتر از واریانس آن باشد را ارائه کردند. همچنین کانسول در سال ۱۹۸۹ درکتاب خود مشتمل بر ده فصل، توزیع پواسون تعمیم‌یافته را مفصل‌تر معرفی کرده، منشأ و برخی از خواص آماری این توزیع و آزمون‌هایی برای تشخیص مدل و همچنین شیوه‌هایی از استنباط آماری پارامترها را ارائه کرد. فامویه و لی (۱۹۹۲) برآورد پارامترهای این توزیع را به روش اختلافات وزنی فراوانی‌های مشاهده شده و مورد انتظار نیز ارائه دادند. همچنین در مقاله خود یک روش جدید وزن دارکردن برای برآورد پارامترهای این توزیع را مورد بحث قرار داده و مشاهده کردند اریبی تحت آن روش بهتر از دیگر روش‌های برآوردیابی است اما واریانس آن بیشترین مقدار است. آمباگاسپیدیا و بالاکریشنان (۱۹۹۴) به توضیح و بررسی توزیع پواسن تعمیم‌یافته مرکب، گشتاورهای مرکزی و برخی خواص آن پرداختند. لرنر و همکاران (۱۹۹۷) اثبات دیگری برای توزیع احتمال بودن پواسن تعمیم‌یافته ارائه کرده و نشان دادند که توزیع پواسن تعمیم‌یافته یک توزیع پواسن مرکب است که توزیع ترکیبی حالت خاصی از خود این توزیع است. فامویه (۱۹۹۹) به ارائه و مقایسه آزمون‌های نیکویی برازش توزیع پواسن تعمیم یافته پرداخت. توئنتر در سال ۲۰۰۰ اثبات دیگری را

[1] Maximum Likelihood