



دانشگاه رازی

دانشکده علوم

جلسه دفاع از پایان نامه دکتری

عنوان:

تحلیل داده‌های بقا با استفاده از مدل‌های آمیخته فرآیند دیریکله با هسته‌های توزیع وایبل نمایی شده.

ارائه دهنده: علی صالح علی الباجلان

زمان: 1405/05/6 ساعت 14 : 00



ایجناب علی صالح علی الباجلان، موفق به اخذ مدرک کارشناسی آمار در سال 1375 از دانشگاه بغداد عراق و نیز مقطع کارشناسی ارشد در رشته آمار جمعیت در سال 1393 از دانشگاه ساو بهتون انگلستان شدم. و در حال حاضر دانشجوی مقطع دکتری آمار در دانشگاه رازی کرمانشاه می باشم. ایجناب در زمینه تحلیل داده‌های بقا با استفاده از مدل‌های آمیخته فرآیند دیریکله با هسته‌های توزیع وایبل نمایی شده دارای فعالیت نموده‌ام. ایجناب در این سالها از محضر اساتید برجسته‌ای همچون دکتر سلیمان خزائی و دکتر بهرام حاجی جوادی بهره بردم.

اعضای محترم هیأت داوران و راهنمایان پایان نامه

ردیف	نام و نام خانوادگی	سمت
1	دکتر سلیمان خزائی	استاد راهنمای اول
2	دکتر بهرام حاجی جودکی	استاد راهنمای دوم
3	دکتر مهرداد نیاپورست	داور داخلی
4	دکتر حبیب جعفری	داور داخلی
5	دکتر افشین پرورده (از دانشگاه اصفهان)	داور خارجی
6	دکتر عباس رسولی (از دانشگاه زنجان)	داور خارجی
7	دکتر حمزه موسوی	نماینده تحصیلات تکمیلی

چکیده

هدف ما در این پایان‌نامه، تحلیل داده‌های بقا با استفاده از مدل‌های آمیخته‌ی فرایند دیریکله با هسته‌ی توزیع وایبل نمایی شده است. داده‌های بقا در بسیاری از حوزه‌های علمی مانند پزشکی، زیست‌شناسی، مهندسی قابلیت اعتماد و علوم اجتماعی دارای ویژگی‌هایی همچون ناهمگنی، چندمُدی بودن و سانسور هستند که مدل‌سازی آن‌ها را با چالش مواجه می‌سازد. مدل‌های پارامتری ساده مانند نمایی و وایبل نمی‌توانند این پیچیدگی‌ها را به خوبی منعکس کنند و روش‌های ناپارامتری نیز امکان استنباط در سطح پارامترها را فراهم نمی‌سازند.

در این پژوهش، یک مدل آمیخته فرایند دیریکله با هسته توزیع وایبل نمایی شده (DPEWM) برای تحلیل داده‌های بقای راست‌سانسور شده معرفی می‌شود. توزیع وایبل نمایی شده به دلیل انعطاف‌پذیری بالا در تولید اشکال مختلف تابع خطر (افزایشی، کاهشی، ثابت، تک‌نما و وانی‌شکل)، انتخاب مناسبی به عنوان هسته مدل آمیخته است. مدل پیشنهادی با استفاده از چارچوب بیز ناپارامتری، قادر به خوشه‌بندی خودکار داده‌ها بدون نیاز به تعیین تعداد خوشه‌ها از پیش است.

الگوریتم نمونه‌گیری گیبز برای استنباط بیزی پارامترهای مدل توسعه داده شده است که شامل پنج مرحله اصلی است: (۱) به‌روزرسانی بردار پیکربندی خوشه‌ها با استفاده از الگوریتم نیل، (۲) به‌روزرسانی پارامترهای خوشه با استفاده از الگوریتم متروپولیس-هستینگز روی مقیاس لگاریتمی، (۳) به‌روزرسانی ابرپارامترها، (۴) به‌روزرسانی پارامتر دقت فرایند دیریکله و (۵) محاسبه وزن‌های مدل آمیخته با روش چوب‌شکسته.

عملکرد مدل پیشنهادی از طریق مطالعه شبیه‌سازی و تحلیل داده واقعی Faithful مورد ارزیابی قرار گرفته و با مدل‌های رقیب شامل مدل آمیخته فرایند دیریکله با هسته وایبل (DPWM) و مدل آمیخته فرایند دیریکله با هسته بور نوع دوازده سه‌پارامتری (DPB3M) مقایسه شده است. نتایج مطالعه شبیه‌سازی نشان می‌دهد که مدل DPEWM در برآورد توابع چگالی، بقا و نرخ خطر، دارای میانگین مربعات خطای کمتری نسبت به دو مدل دیگر است و فاقد سوگیری قابل توجه می‌باشد. در تحلیل داده واقعی، مدل DPEWM به درستی ساختار دوخوشه‌ای داده‌ها را با احتمال بیش از ۹۳ درصد تشخیص می‌دهد، در حالی که مدل‌های رقیب عملکرد ضعیف‌تری در خوشه‌بندی دارند.

نتایج این پژوهش نشان می‌دهد که مدل آمیخته فرایند دیریکله با هسته وایبل نمایی شده، ابزاری قدرتمند، انعطاف‌پذیر و قابل اعتماد برای تحلیل داده‌های بقا و خوشه‌بندی خودکار داده‌ها محسوب می‌شود و می‌تواند به عنوان جایگزینی مناسب برای مدل‌های رایج در این حوزه مورد استفاده قرار گیرد.

Abstract

Our goal in this thesis is to analyze survival data using Dirichlet process mixture models with exponentiated Weibull kernel. Survival data in many scientific fields such as medicine, biology, reliability engineering, and social sciences have characteristics such as heterogeneity, multimodality, and censoring, which make their modeling challenging. Simple parametric models such as exponential and Weibull cannot adequately reflect these complexities, and nonparametric methods also do not provide inference at the parameter level.

In this research, a Dirichlet process mixture model with exponentiated Weibull kernel (DPEWM) is introduced for analyzing right-censored survival data. The exponentiated Weibull distribution, due to its high flexibility in generating various shapes of the hazard function (increasing, decreasing, constant, unimodal, and bathtub-shaped), is a suitable choice as the kernel of the mixture model. The proposed model, using the Bayesian nonparametric framework, is capable of automatic clustering of data without the need to predetermine the number of clusters.

A Gibbs sampling algorithm has been developed for Bayesian inference of the model parameters, which includes five main steps: (1) updating the cluster configuration vector using Neal's algorithm, (2) updating cluster parameters using the Metropolis-Hastings algorithm on the logarithmic scale, (3) updating hyperparameters, (4) updating the precision parameter of the Dirichlet process, and (5) computing the mixture model weights using the stick-breaking method.

The performance of the proposed model has been evaluated through simulation studies and analysis of the real Faithful dataset and compared with competing models including the Dirichlet process mixture model with Weibull kernel (DPWM) and the Dirichlet process mixture model with three-parameter Burr type XII kernel (DPB3M). The simulation study results show that the DPEWM model has lower mean squared error in estimating density, survival, and hazard functions compared to the other two models and has no significant bias. In the real data analysis, the DPEWM model correctly identifies the two-cluster structure of the data with a probability of over 93%, while the competing models have weaker performance in clustering.

The results of this research show that the Dirichlet process mixture model with exponentiated Weibull kernel is a powerful, flexible, and reliable tool for survival data analysis and automatic clustering of data, and can be used as a suitable alternative to common models in this field.