МАТЕМАТИЧКИ ОМНИБУС, 11 (2024), 63 – 74
http://im-pmf.weebly.com/matematicki-omnibus-kniga-11.html

ЕКСПОНЕНЦИЈАЛНИ ФАМИЛИИ НА РАСПРЕДЕЛБИ НА ВЕРОЈАТНОСТ

Ерблина Зеќири [1]

1. ВОВЕД

Во теоријата на веројатност и статистика, под експоненцијална фамилија се подразбира параметарско множество на распределби на веројатност кои имаат одредено својство. Тие распределби се генерализација на неколку основни веројатносни модели со заеднички специјални особини.

Причината за важноста и широката употреба на експоненцијалните фамилии е тоа што повеќе класични модели во статистиката се всушност експоненцијални фамилии на распределби на веројатност. Како резултат на тоа, голем број класични методи за оценување параметри и тестирање хипотези даваат задоволителни резултати кога се работи во експоненцијална фамилија на распределби на веројатност.

Во овој труд ќе ги објасниме основните дефиниции и својства на експоненцијалните фамилии, а воедно ќе се задржиме и на дел од својствата на најпознатите експоненцијални фамилии како што се бернулиевата распределба, бета распределбата, експоненцијалната распределба, гама распределбата и некои други распределби на веројатност.

2. ЕКСПОНЕНЦИЈАЛНИ ФАМИЛИИ

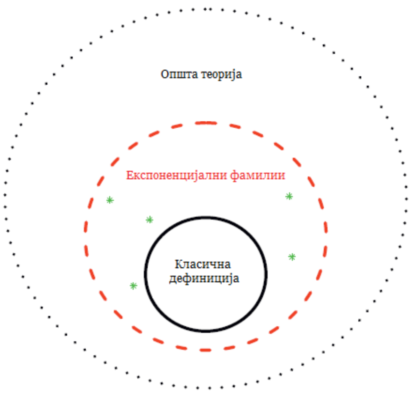
2.1 Кратка историја

Одлични идеи во математиката понекогаш се раѓаат во миг на инспирација, вообичаено објавени пред светот преку некоја револуционерна статија. Такво нешто се случило со Фишеровата статија за максимално подобни оценувачи (Ronald Fisher, 1890-1962), објавена во 1925 година. Но, ништо слично ни одблизу не се случило со експоненцијалните фамилии на веројатност. Теорија за експоненцијалните фамилии се развивала многу securities полека во периодот од 1932 до 1970 година.

Терминот "експоненцијални фамилии" е релативно нов. До средината на 1950-тите години оваа класа на распределби на веројатност била позната како "Купман–Дармоа–Питман" фамилии на распределби на веројатност (во чест на тројцата истакнати статистичари кои работеле посебно, во три различни земји).

Едвин Питман (Edwin John George Pitman), е статистичар познат по своите придонеси во областа на статистиката во средината на 20-тиот век. Најпознат е по својата работа за развој на Питмановиот критериум за ефикасност на оценувачи. Рене Дармоа (René Auguste Joseph Émile Darmois), е француски математичар и статистичар. Иако можеби не е толку познат како некои други статистичари, тој имал значаен придонес во развојот на статистичката теорија, со основање на теоријата на таканаречените експоненцијални фамилии на распределби. Бернард Оскар Купман (Bernard Oscar Koopman), е француско-американски статистичар познат по неговата работа во ергодичната теорија на случајните процеси и основање на теоријата операционите истражувања.

Подолу на слика 1 е прикажана груба шема на историјата на статистиката во 20-тиот век.



Слика 1. Три нивоа на статистичкото моделирање.[3]

Внатрешниот круг ја претставува класичната статистика во која се изучува изведувањето на точни заклучоци преку различни статистички тестови: t-тестови, F-тестови, хи-квадратна статис–тика, ANOVA, повеќедимензионална анализа. Во надворешниот круг е општата теорија базирана на апроксимации како што е централната граничната теорема.

Неколку посебни резултати се наоѓаат надвор од првиот круг, тоа се неколку особини кои ги поврзуваат биномната, поасо-новата, гама и бета распределбите. Овие особини се зелените ѕвездички на сликата. Тие влегуваат во средишниот круг, т.е. во кругот на „експоненцијалните фамилии“, каде што се можни „скоро точни“ заклучоци.

2.2 Дефиниција и основни примери

Ќе ја илустрираме дефиницијата на експоненцијалните фамилии на распределба на веројатност со помош на еден едноставен пример, [2].

Пример 2.2.1 (**Нормална распределба со** $\mu = 0$). [2] Нека X е случајна променлива, $X \sim N\left(0, \sigma^2\right)$. Тогаш густината на распре-делба на случајната променлива X е зададена со

$$p\left(x, \sigma\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, x \in \mathbb{R}.$$

Забележуваме дека оваа функција содржи само еден пара-метар σ. Ако ги воведеме ознаките:

$$\eta\left(\sigma\right) = -\frac{1}{2\sigma^2}, \quad T\left(x\right) = x^2, \quad \psi\left(\sigma\right) = \log\sigma, \quad h\left(x\right) = \frac{1}{\sqrt{2\pi}}, \quad x \in \mathbb{R},$$

тогаш за густината добиваме

$$p\left(x, \sigma\right) = e^{\eta(\sigma)T(x) - \psi(\sigma)} h\left(x\right),$$

за секој $\sigma \in \mathbb{R}^+$, $x \in \mathbb{R}$.

Следно, претпоставуваме дека $(X_1, X_2, \ldots, X_n)$ е примерок од n независни, еднакво распределени случајни променливи $X_i, i = 1, 2, .., n,$ $X_i \sim N\left(0, \sigma^2\right)$, кој одговара на обележјето X. Тогаш за функцијата на подобност имаме:

$$p\left(x_1,x_2,\ldots,x_n,\sigma\right)=\frac{1}{\sigma^n\left(2\pi\right)^{n/2}}e^{-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}},\quad x_1,x_2,\ldots,x_n\in\mathbb{R}.$$

Од обликот на функцијата на подобност, заради теоремата на Фишер за факторизација, добиваме дека статистиката

$$T\left(X_1,X_2,\ldots,X_n\right)=\sum_{i=1}^n X_i^2$$

е доволна статистика за параметарот σ^2, [5].

Слично како погоре, ги користиме ознаките за примерокот

$$\eta\left(\sigma\right)=-\frac{1}{2\sigma^2},\quad T\left(X_1,X_2,\ldots,X_n\right)=\sum_{i=1}^n X_i^2,\quad \psi\left(\sigma\right)=n\log\sigma,\text{ и}$$

$$h\left(x_1,x_2,\ldots,x_n\right)=\frac{1}{\left(2\pi\right)^{n/2}},\ x_1,x_2,\ldots,x_n\in\mathbb{R}.$$

Уште еднаш ја претставуваме функцијата на подобност со соодветните смени и добиваме

$$p\left(x_1,x_2,\ldots,x_n,\sigma\right)=e^{\eta(\sigma)T\left(x_1,x_2,\ldots,x_n\right)-\psi(\sigma)}h\left(x_1,x_2,\ldots,x_n\right).$$

Забележуваме дека во репрезентацијата на функцијата на подобност $p\left(x_1,x_2,\ldots,x_n,\sigma\right)$, статистиката $T\left(X_1,X_2,\ldots,X_n\right)$ има облик $T\left(X_1,X_2,\ldots,X_n\right)=\sum_{i=1}^n X_i^2$. Имајќи предвид дека случајните променливи $X_i, i=1,2,..,n,$ имаат $X_i\sim N\left(0,\sigma^2\right)$ распределба, следува дека случајните променливи $Y_i=\frac{X_i}{\sigma}, i=1,2,..,n,$ имаат $Y_i\sim N\left(0,1\right)$ распределба. Користејќи го фактот дека збирот на квадратите на n независни стандардно нормално распределени случајни променливи, е хи-квадратна распределба со n степени на слобода [5, стр. 14], имаме дека $Y=\sum_{i=1}^n Y_i^2\sim\chi_n^2$. Оттука за густи–

ната на распределба на статистиката $T\left(X_1,X_2,\dots,X_n \right) = \sum_{i=1}^{n} X_i^2$ $= \sigma^2 Y$, добиваме дека:

$$p_T\left(t,\sigma\right)=\frac{e^{-\frac{t}{2\sigma^2}}t^{\frac{n}{2}-1}}{\sigma^n 2^{n/2}\Gamma\left(\frac{n}{2}\right)},\ t>0.$$

Ако сега ги искористиме ознаките:

$$\eta\left(\sigma\right)=-\frac{1}{2\sigma^2},\quad S\left(t\right)=t,\quad \psi\left(\sigma\right)=n\log\sigma,\quad h\left(t\right)=\frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)},\quad t>0,$$

по направените смени, добиваме дека густината на распределба на $T\left(x_1,x_2,\dots,x_n \right)=\sum_{i=1}^{n} x_i^2$ го има истиот облик

$$p_T\left(t,\sigma\right)=e^{\eta(\sigma)S(t)-\psi(\sigma)}h\left(t\right).$$

Јасно е дека нешто интересно се појавува. Почнавме со една основна густина на распределба која ја доведовме до специфичен облик, $p\left(x,\sigma\right)=e^{\eta(\sigma)T(x)-\psi(\sigma)}h\left(x\right)$, за потоа да добиеме дека и функцијата на подобност и доволната статистика $T\left(x_1,x_2,\dots,x_n \right)=\sum_{i=1}^{n} x_i^2$ во таа функција на подобност имаат густина на распределба што ја има истата општа форма. Може да заклучиме дека сите тие припаѓаат во една фамилија на густини на распределба, која претставува еднопараметарска фамилија, поточно експоненцијална фамилија на распределби на веројатност.

Да ја дефинираме таа фамилија во општ случај кога имаме само еден параметар. Слично се дефинира и експоненцијална фамилија на распределби на веројатност за распределби кои имаат повеќе од еден параметар. Во овој труд ќе задржиме само на еднопараметарската фамилија.

Дефиниција 2.2.1 Нека $\left(X_1,X_2,\dots,X_d\right)$ е d-димензионален случаен примерок со распределба $P_\theta,\ \theta\in\Theta\subseteq\mathbb{R}$. Нека $X_1,X_2,\dots,X_d$

се непрекинати случајни променливи. Велиме дека фамилијата на распределби $\{P_\theta,\ \theta\in\Theta\}$ припаѓа на *еднопараметарска експоненцијална фамилија* на распределби на веројатност ако густината на распределба на случајниот вектор $(X_1, X_2, \ldots, X_d)$ може да се претстави во облик

$$p(x,\theta)=e^{\eta(\theta)T(x)-\psi(\theta)}h(x),$$

за некои реалновредносни функции $T(x), \psi(\theta)$ и $h(x)\ge 0$.

Да забележиме дека функциите η, T и h не се единствени (на пример може да ја помножиме функцијата T со некоја константа c и да ја поделиме функцијата η со истата константа).

Дефиниција 2.2.2 Нека $(X_1, X_2, \ldots, X_d)$ е примерок со распре–делба $P_\theta,\ \theta\in\Theta$ која припаѓа на еднопараметарската експоненцијална фамилија. Тогаш статистиката $T(x)$ се нарекува *природно доволна статистика* за фамилијата $\{P_\theta,\ \theta\in\Theta\}$.

Поимот доволна статистика е еден од основните поими во теоријата на статистиката и нејзините примени. Доволноста е воведена во статистичката литература од Роналд А. Фишер (Fisher, 1922). Една доволна статистика треба да содржи во себе целосна информација за непознатите параметри на распредел–бата на примерокот. Во таа смисла, нема да се изгуби ништо при донесување на заклучоци со ограничување на вниманието само на доволната статистика.

Во продолжение ќе спомнеме некои познати распределби кои припаѓаат на експоненцијалната фамилија на распределби на веројатност, како и начинот на кој може да се утврди дека навистина припаѓаат на експоненцијалната фамилија.

Пример 2.2.2 (**Биномна распределба**) [1] Нека случајната променлива X има распределба $X\sim B(n,p)$, за $n\ge 1$ и параметар $p,\ 0<p<1$. Ја претставуваме густината на распределба на случајната променлива X во облик на еднопараметарска експо–ненцијална фамилија на распределби.

$$p(x)=\binom{n}{x}p^x(1-p)^{n-x}=\binom{n}{x}\left(\frac{p}{1-p}\right)^x(1-p)^n=\binom{n}{x}e^{x\log\left(\frac{p}{1-p}\right)+n\log(1-p)},$$

кога $x\in\{0,1,\dots,n\}$.

Означуваме $\eta(p)=\log\left(\frac{p}{1-p}\right), T(x)=x, \psi(p)=-n\log(1-p)$ и $h(x)=\binom{n}{x}$ и тогаш густината на распределба на случајната променлива X добива облик

$$p(x)=\binom{n}{x}p^x(1-p)^{n-x}=h(x)e^{T(x)\eta(p)-\psi(p)},\ p\in(0,1).$$

Значи, фамилијата на распределби на веројатност $\{p(x)\,|\,0<p<1\}$ кога $X\sim B(n,p)$ формира еднопараметарска експоненцијална фамилија на распределби веројатност.

Пример 2.2.3 (**Нормална распределба со позната дисперзија**) Нека X е случајна променлива, $X\sim N(\mu,\sigma^2)$, при што сметаме дека σ дека е позната, а $\mu\in\mathbb{R}$ е параметар. Тогаш за густината на распределбата на случајната променлива X имаме

$$p(x,\mu)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}+\mu x-\frac{\mu^2}{2}},\ x\in\mathbb{R}.$$

Со слични трансформации и смени $\eta(\mu)=\mu, T(x)=x$, $\psi(\mu)=\frac{\mu^2}{2}$ и $h(x)=e^{-\frac{x^2}{2}}$, густината на распределба на $X\sim N(\mu,\sigma^2)$ може да се претстави во облик на еднопараметарска експоненцијална фамилија. Значи, фамилијата на распределби на веројатност $\{p(x,\mu)\,|\,\mu\in\mathbb{R}\}$ формира еднопараметарска експоненцијална фамилија на распределби.

Познати распределби кои се дел од експоненцијалната фамилија на распределби на веројатност се експоненцијалната, гама, хи-квадратна, бета, Поасонова, геометриската распределба како и други распределби.

Природно е да се постави и прашањето кои распределби (ако постојат) не припаѓаат во експоненцијалната фамилија на распределби на веројатност.

Пример 2.2.3 Важен пример за распределба на веројатност која не припаѓа во експоненцијалната фамилија на распределби на веројатност е Кошиевата распределба, со густина на распределба зададена на следниот начин

$$p(x,\alpha)=\frac{1}{\pi\left(1+(x-\mu)^2\right)},\ x\in\mathbb{R}.$$

Да претпоставиме обратно, нека и Кошиевата распределба припаѓа на експоненцијалната фамилија на распределби на веро–јатност. Тоа значи дека постојат функции $\eta(\mu),T(x)$ такви што за сите x,μ важи

$$e^{\eta(\mu)T(x)}=\frac{1}{1+(x-\mu)^2},\ \text{односно}\ \ \eta(\mu)T(x)=-\log\left(1+(x-\mu)^2\right),$$

од каде $\eta(0)T(x)=-\log\left(1+x^2\right)$. Јасно, тогаш $T(x)=-c\log\left(1+x^2\right)$, за некоја константа c.

Со замена во претходната формула добиваме дека за сите x,μ

$$-c\eta(\mu)\log\left(1+x^2\right)=-\log\left(1+x^2\right),$$

од каде следува дека $\eta(\mu)=\frac{1}{c}\frac{\log\left(1+(x-\mu)^2\right)}{\log\left(1+x^2\right)}.$

Оваа значи дека изразот $\frac{\log\left(1+\left(x-\mu\right)^2\right)}{\log\left(1+x^2\right)}$ мора да биде константна функција од x, што е контрадикција. Изборот на $\mu=0$ како специјална вредност на μ не е важен.

2.3 Канонична форма на експоненцијалната фамилија

Нека фамилијата $\left\{P_\theta,\ \theta\in\Theta\right\}$ припаѓа во еднопараметарска експоненцијална фамилија на распределби на веројатност, со густина на распределба од облик $p\left(x,\theta\right)=e^{\eta(\theta)T(x)-\psi(\theta)}h\left(x\right)$. Ако $\eta\left(\theta\right)$ е функција од θ која е инјекција, тогаш може да ја репараметризираме густината на распределба само во однос на η. Односно добиваме $e^{\eta T(x)-\psi(\eta)}h\left(x\right)$.

Дефиниција 2.3.1 Нека $\left(X_1,X_2,\dots,X_d\right)$ е примерок земен од обележје со распределба $P_\eta,\ \eta\in\mathcal{T}\subseteq\mathbb{R}$. Велиме дека фамилијата од распределби $\left\{P_\eta\mid\eta\in\mathcal{T}\right\}$ припаѓа на *каноничната еднопараметарска експоненцијална фамилија* на распределби ако густината на распределба P_η може да се запише во облик

$$p\left(x,\eta\right)e^{\eta T(x)-\psi(\eta)}h\left(x\right).$$

Притоа,

$$\eta\in\mathcal{T}=\left\{\eta\mid e^{\psi(\eta)}=\int\limits_{\mathbb{R}^d}e^{\eta T(x)}h\left(x\right)dx<\infty\right\},$$

во случај на непрекината случајна променлива и

$$\mathcal{T}=\left\{\eta\mid e^{\psi(\eta)}=\sum_{x\in\mathcal{X}}e^{\eta T(x)}h\left(x\right)dx<\infty\right\},$$

во случај на дискретна случајна променлива, за $\mathcal{X}$ пребројливо множество на кое $h(x)>0$.

Дефиниција 2.3.1 За една распределба која припаѓа на каноничната експоненцијална фамилија, параметарот η се вика *природен параметар* и $\mathcal{T}$ се вика *простор на природни параметри*.

2.4 Својство на конвексност

Дадена во канонична форма, експоненцијалната фамилија има некои својства на конвексност.

Теорема 2.4.1 Просторот од природни параметри $\mathcal{T}$ е конвексно множество и $\psi(\eta)$ е конвексна функција на $\mathcal{T}$.

Доказ. Ќе го разгледаме само случајот кога имаме непрекината случајна променлива.

Нека $\eta_1,\eta_2 \in \mathcal{T}$ и нека $0<\alpha<1$. Треба да покажеме дека и $\alpha\eta_1+(1-\alpha)\eta_2 \in \mathcal{T}$ т.е.

$$\int_{\mathbb{R}^d} e^{(\alpha\eta_1+(1-\alpha)\eta_2)T(x)} h(x)\,dx<\infty$$

Имаме,

$$\int_{\mathbb{R}^d} e^{(\alpha\eta_1+(1-\alpha)\eta_2)T(x)} h(x)\,dx=\int_{\mathbb{R}^d} e^{\alpha\eta_1 T(x)} e^{(1-\alpha)\eta_2 T(x)} h(x)\,dx$$

$$=\int_{\mathbb{R}^d} \left(e^{\eta_1 T(x)}\right)^{\alpha}\left(e^{\eta_2 T(x)}\right)^{1-\alpha} h(x)\,dx$$

$$\le\left(\int_{\mathbb{R}^d} e^{\eta_1 T(x)} h(x)\,dx\right)^{\alpha}\left(\int_{\mathbb{R}^d} e^{\eta_2 T(x)} h(x)\,dx\right)^{1-\alpha} \le \infty .$$

Последното важи затоа што од претпоставката $\eta_1,\eta_2 \in \mathcal{T}$ следува дека и $\int_{\mathbb{R}^d} e^{\eta_1 T(x)} h(x)\,dx$ и $\int_{\mathbb{R}^d} e^{\eta_2 T(x)} h(x)\,dx$ се конечни.

Да забележиме дека ние веќе го покажавме неравенството

$$e^{\psi(\alpha\eta_1+(1-\alpha)\eta_2)} \le e^{\alpha\psi(\eta_1)+(1-\alpha)\psi(\eta_2)}.$$

Ова неравенство е еквивалентно со

$$\psi\left(\alpha\eta_1+(1-\alpha)\eta_2\right) \le \alpha\psi(\eta_1)+(1-\alpha)\psi(\eta_2),$$

од каде што добиваме дека $\psi(\eta)$ е конвексна функција на $\mathcal{T}$, [4].

3. ЗАКЛУЧОК

Експоненцијалните фамилии на распределби на веројатност се покажале како моќна и универзална класа на распределби во статистичката теорија и нејзина примена.

Како што спомнавме во трудот, овие распределби имаат неколку извонредни својства, како што е на пример доволната статистика, што ни овозможува да ги извлечеме сите релевантни информации за параметрите од даденото множество со податоци. Ова својство е особено важно за оценување на параметрите и тестирањето на хипотези.

Експоненцијалните фамилии наоѓаат примена во разни области, од физиката и инженерството до биологијата и економијата. Нивната флексибилност и математичка обработливост ги прават фундаментална алатка за статистичарите.

ЛИТЕРАТУРА

[1] O. Barndorff-Nielsen, *Information and Exponential Families,* Matematisk Institut, Aarhus Universitet, 1978.

[2] A. DasGupta, *The Exponential Family and Statistical Applications, In: Probability for Statistics and Machine Learning*, Springer Texts in Statistics. Springer, New York, NY, 2011.

[3] B. Efron, *Exponential Families in Theory and Practice*, Cambridge University Press, 2023.

[4] M. Jordan, *Notes: The exponential Families: Basics,* EECS at UC BerkeleyB, 2009.

[5] И. Стојковска, *Предавања по Основи на статистика,* Универзитет „Св. Кирил и Методиј “, Скопје, Институт за математика, 2013.

[1] Универзитет “Св. Кирил и Методиј”, Скопје
Природно-математички факултет,
Институт за математика
Архимедова 3, 1000 Скопје, Р. Македонија
e-mail: erblina_zeqiri@hotmail.com

Примен: 28.9.2023
Поправен: 19.2.2024
Одобрен: 23.2.2024
Објавен на интернет: 20.11.2024