ЕКСПОНЕНЦИЈАЛНИ ФАМИЛИИ ОД РАСПРЕДЕЛБИ НА ВЕРОЈАТНОСТ

Ерблина Зеќири
Институт за математика
Природно-математички
Факултет, Скопје

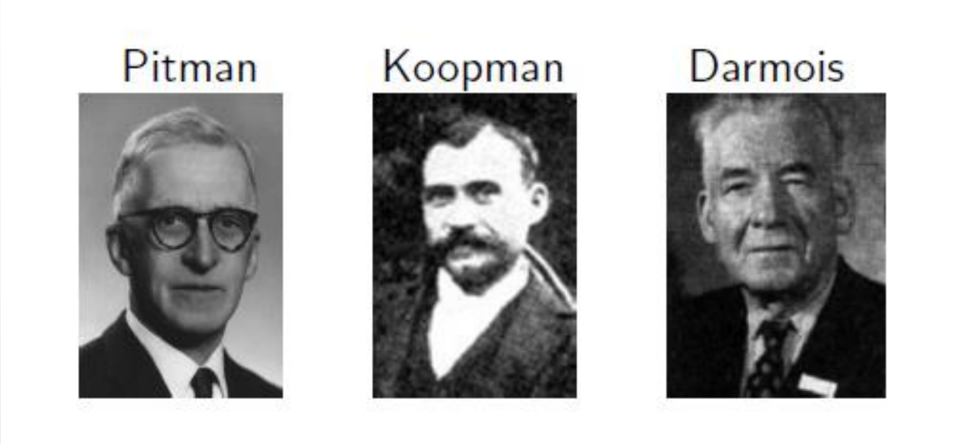
Шести семинар „Математика и примени“, 17 март 2023
Институт за математика, Природно-математички факултет,
Универзитет „Св. Кирил и Методиј“, Скопје

ЕКСПОНЕНЦИЈАЛНИ ФАМИЛИИ НА РАСПРЕДЕЛБА НА ВЕРОЈАТНОСТ

- **Вовед**

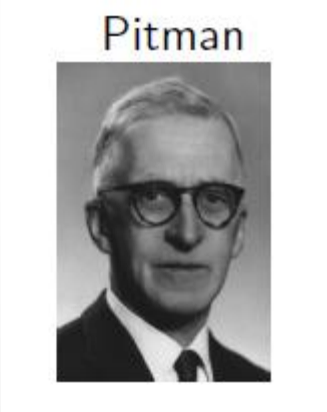
  Поимот експоненцијални фамилии прв пат е воведен од следните познати математичари:

  - E. J. G. Pitman
  - G. Darmois
  - B. O. Koopman



ЕКСПОНЕНЦИЈАЛНИ ФАМИЛИИ НА РАСПРЕДЕЛБА НА ВЕРОЈАТНОСТ

- **Вовед**



E. J. G. Pitman беше австралиски математичар кој даде значителен придонес во статистиката и теоријата на веројатност. Особено, тој е запаметен првенствено како основач на тестот за пермутација на Питман и ефикасноста на Питман.

ЕКСПОНЕНЦИЈАЛНИ ФАМИЛИИ НА РАСПРЕДЕЛБА НА ВЕРОЈАТНОСТ

- **Вовед**

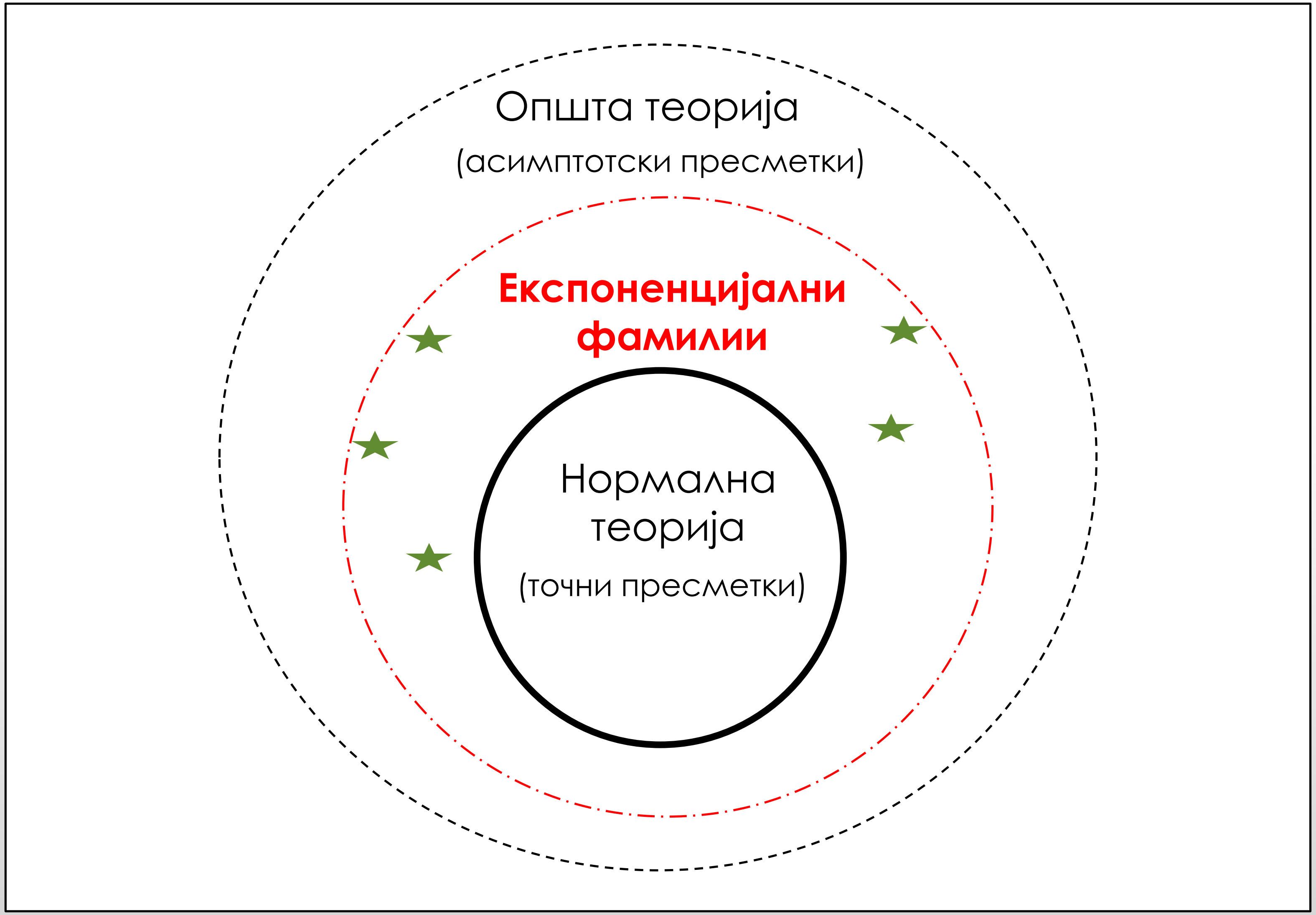
В. О. Koopman бил американски математичар роден во Франција, познат по неговата работа во ергодичната теорија, основите на веројатноста, статистичката теорија и операционото истражување.

ЕКСПОНЕНЦИЈАЛНИ ФАМИЛИИ НА РАСПРЕДЕЛБА НА ВЕРОЈАТНОСТ

- **Вовед**

Georges Darmois бил француски математичар и статистичар. Тој беше пионер во теоријата на доволни статистики.

ЕКСПОНЕНЦИЈАЛНИ ФАМИЛИИ НА РАСПРЕДЕЛБА НА ВЕРОЈАТНОСТ



ЕКСПОНЕНЦИЈАЛНИ ФАМИЛИИ НА РАСПРЕДЕЛБА НА ВЕРОЈАТНОСТ

Зошто?

- Кога избираме непристрасен оценувач за некој параметар, сакаме неговата дисперзија да биде што е можно помала. Под одредени услови постои формула која дава долна граница за дисперзијата (долна граница Рао-Крамер) на кој било непристрасен оценувач. Ниту еден оценувач не може да биде подобар од долната граница. Само во експоненцијалната фамилија може да се постигне таа долна граница.
- Можни се „речиси точни“ теории како што се генерализирани линеарни модели (GLM).
- Наоѓаат примена и во Machine Learning теорија.
- повеќе класични модели во статистиката се всушност експоненцијални фамилии на распределби на веројатност
- голем број на класични методи за оценување на параметри и тестирање на хипотези даваат задоволителни резултати кога се работи во експоненцијална фамилија на распределби на веројатност.

Дефиниција на експоненцијални фамилии од распределби на веројатност

(Не станува збор за екзпоненцијална распределба!)

Експоненцијална фамилија е параметарска фамилија на распределби на веројатност со густина на распределба кои задоволуваат одредени својства што ги прават многу корисни од математичка гледна точка.

Нека Ф е множество од сите распределби на веројатност. Воспоставуваме кореспонденција меѓу Ф и некој простор на параметри $\Theta \subset \mathbb{R}^k$. Ако таа кореспонденција е функција која на секој параметар $\theta \in \Theta \subset \mathbb{R}^k$ му придружува единствена распределба од множеството Ф тогаш Ф се нарекува параметарска фамилија.

Дефиниција на експоненцијални фамилии од распределби на веројатност

Пример: Нека е дадено множеството на сите нормални распределби. Секоја нормална распределба е карактеризирана со нејзината аритметичка средина μ (реален број) и нејзината дисперзија σ^2 (позитивен реален број)

$$X \sim N\left(\mu, \sigma^2\right)$$

множество

$$\Phi$$

кореспонденција

$$\longleftrightarrow$$

простор на параметри

$$\Theta = \mathbb{R} \times \mathbb{R}^+$$

Дефиниција на експоненцијални фамилии од распределби на веројатност

Дефиниција:

За една параметарска фамилија на еднодимензионални непрекинати распределби на веројатност велиме дека е експоненцијална фамилија ако и само ако густината на распределба на кој било член од семејството може да се запише во облик:

$$p_X(x) = h(x) \cdot e^{\left[\eta(\theta) \cdot T(x) - A(\theta)\right]}$$

каде што $h(x), \eta(\theta), T(x), A(\theta)$ се функции т.ш.

$h(x)$ - е функција која зависи само од x

$\eta(\theta)$ - функција која зависи од параметарот θ

$T(x)$ - функција која зависи од x

$A(\eta)$ - Кумулативна функција

Експоненцијални фамилии и доволни статистики

Експоненцијална фамилија

$$p_X(x) = h(x) \cdot e^{\left[\eta(\theta) \cdot T(x) - A(\theta)\right]}$$

Доволни статистики

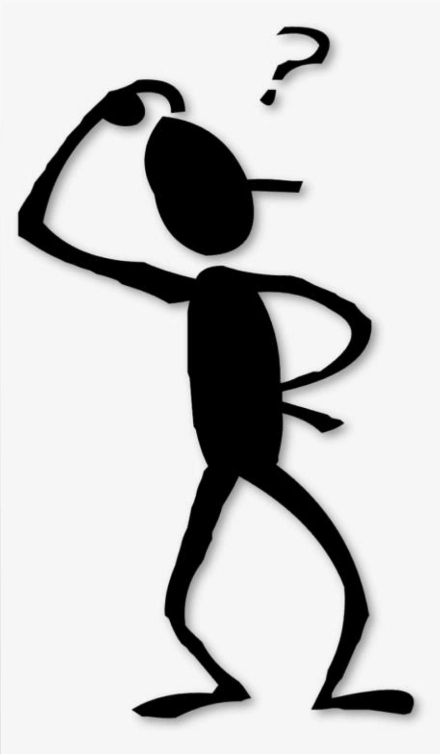
$$p_X(x) = h(x) \cdot g\left(T(x), \theta\right)$$

Знаејќи ја доволната статистика $T(x)$ значи дека знаеме секоја информација за параметарот θ.

Експоненцијални фамилии и доволни статистики

Зошто доволната статистика е доволна?

„Ако сакаме да ја процениме средната висина на граѓаните на САД, нема смисла да се мери растојанието помеѓу градовите во Кина.“



Најпознати распределби кои се дел од експоненцијалната фамилија

- Нормална
- Експоненцијална
- Гама
- Хи квадратн
- Бета
- Пуасонова
- Бернулиева
- Геометриска

Најпознати распределби кои НЕ се дел од експоненцијалната фамилија

- Студентова
- Рамномерна распределба со нефиксирани граници
- Кошиева распределба и.т.н.

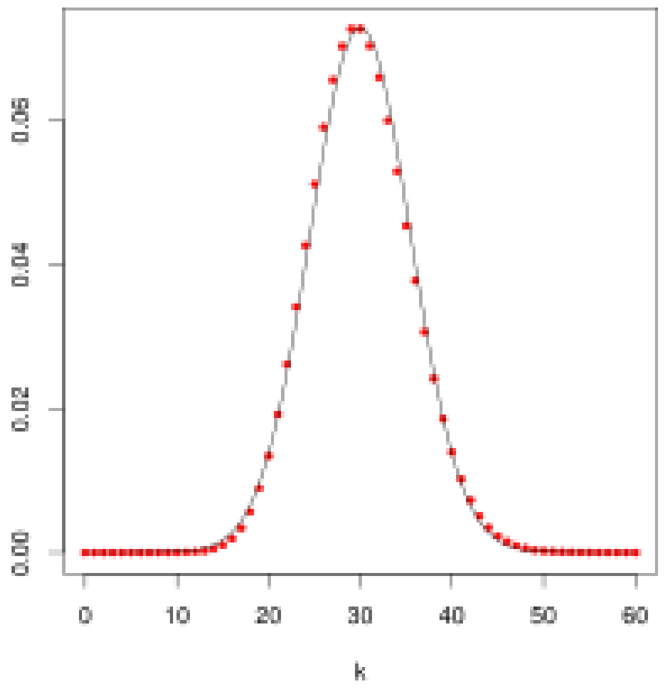
Како да покажеме дека една распределбае дел од експоненцијалната фамилија?

Пуасонова распределба

$$p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$\big\downarrow \text{ трансформација}$$

$$p_X(x) = \frac{1}{x!} e^{[x \log \lambda - \lambda]}$$



$$h(x) = \frac{1}{x!} \qquad \eta(\lambda) = \log \lambda \qquad T(x) = x \qquad A(\eta) = \lambda = e^{\eta}$$

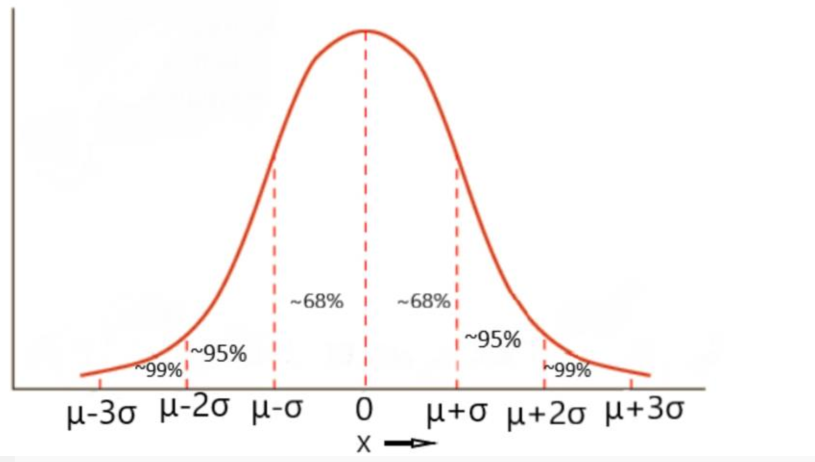
Како да покажеме дека една распределбае дел од експоненцијалната фамилија?

Нормална (Гаусова) распределба

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

трансформација

$$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{\left(\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}\mu^2 - \log\sigma\right)}$$



$$h(x) = \frac{1}{\sqrt{2\pi}} \quad \eta = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} \quad T(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix} \quad A(\eta) = \frac{\mu^2}{2\sigma^2} + \log\sigma = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$$

Експоненцијалната фамилија, математичкото очекување, дисперзијата и другите моменти

Математичко очекување

$$\mu_X = \mathrm{E}(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$$

Математичкото очекување може да се пресмета како **ПРВ ИЗВОД** на функцијата $A(\eta)$.

Општо:

Секој момент од $k-$ти ред може да се пресмета како $k-$ти извод на функцијата $A(\eta)$.

$$\frac{dy}{dx} > \int y\, dx$$

Експоненцијалната фамилија, математичкото очекување, дисперзијата и другите моменти

Нормална (Гаусова) распределба

$$A(\eta) = \frac{\mu^2}{2\sigma^2} + \log\sigma = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2) \qquad \eta_1 = \frac{\mu}{\sigma^2},\ \eta_2 = -\frac{1}{2\sigma^2}$$

диференцираме

$$EX = \frac{\partial A}{\partial \eta_1} = \frac{\eta_1}{2\eta_2} = \mu$$

$$DX = \frac{\partial^2 A}{\partial \eta_1^2} = -\frac{1}{2\eta_2} = \sigma^2$$

ЗАКЛУЧОК

Кои се предностите за моделирање со експоненцијална фамилија?

- Експоненцијалната фамилија е многу поширока од нормалната фамилија која ги содржи само нормалните распределби на веројатност.

- Вклучува широк спектар на врски меѓу варијанса и средна вредност.

- Доволната статистика $T(x)$ е експлицитно дадена во густината на распределба на експоненцијалната фамилија.

Ви благодарам на вниманието!