

Приложения на байсовската статистика в социалните науки

Анотация на курса

Изследователските проблеми, изискващи приложението на статистически методи, могат да се обособят в три големи групи:

В първата група попадат тези проблеми, които имат теоретично решение. Наистина, като правило, теоретичните решения на един и същ проблем са повече от едно и всяко от тях се основава на различна логика и на различни предположения. Това от своя страна поставя изискването да се познават всички решения и във всяка конкретна ситуация да се прави избор.

Във втората група попадат тези проблеми, които нямат теоретично решение, но в своята практическа дейност изследователите ги свеждат, не съвсем коректно, към аналогични проблеми, които имат теоретично решение.

В третата група попадат тези проблеми, които нито имат теоретично решение, нито могат да се сведат до вече решени проблеми.

Целта на настоящият курс е да покаже как от позициите на една малко позната парадигма в статистиката, наречена байсовска, може да се даде теоретично решение на всеки изследователски проблем и в този смисъл трите групи проблеми да се сведат към първата. Разбира се, тук по-скоро става въпрос за потенциал, понеже е практически невъзможно да бъдат описани изчерпателно всички проблеми, изискващи приложението на статистически методи. По тази причина, половината от курса е посветена на теоретично въведение, а в другата половина са разгледани избрани изследователски проблеми. В теоретичното въведение са изложени принципите, инструментите и техниките на байсовската статистика. Това означава, че в тези теми по необходимост има малко повече математика, но тя е необходима основа за разбирането на различните приложения. А в този курс ще се разглеждат приложения, отнасящи се и до трите групи изследователски проблеми:

Към третата група проблеми се отнасят изводите, основани на информация от непредставителни извадки. Примери за непредставителни извадки напоследък има достатъчно много – анкетите в Интернет, гласуванията по телефона или чрез SMS за различни телевизионни предавания (Песен на Евровизия, Тест на нацията, Big Brother, Star Academy, Музикална ку-ку академия, Вот на доверие, Сблъсък и др.).

Към втората група проблеми се отнасят изводите, основани на информация от представителни извадки, но отнасящи се за няколко взаимно свързани параметъра на генералната съвкупност. Такъв, например, е проблемът за оценяването на процента на гласувалите за някоя партия (кандидат), но само сред действителните гласоподаватели.

Към първата група проблеми се отнася определянето на вероятността нова единица да попадне в конкретна група, формирана по значението на даден признак. Това успешно може се съчетае с техниката на многомерните групировки и по този начин се получава лесна за употреба алтернатива на методите, носещи общото наименование „анализ на риска” или „взимане на решение в ситуация на несигурност”.

Теми

1. Малко история

Теория на вероятностите.

Честотна статистика.

Бейсовска статистика.

[Основна литература: Харалампиев 2007]

[Допълнителна литература: Jaynes 1986; Jaynes 2003: ix-xvi и 1601-1614 (и.);*

Loredo 1990: 82-94 (и.)]

2. Малко математика

Вероятност. Условна вероятност.

Теорема за събиране и умножаване на вероятности. Теорема за пълната вероятност и теорема на Бейс. Априорни и апостериорни вероятности. Метод на максималната ентропия.

Разпределение, плътност на разпределение и функция на разпределение.

[Основна литература: Харалампиев 2007]

[Допълнителна литература: Bretthorst 1990: 1-5 (и.); Dose 2002: 1-5 и 16-17 (и.); Jaynes 1988; Jaynes 2003: 101-117, 201-223, 301-321, 418-428, 601-608 и 1101-1110 (и.); Loredo 1990: 94-103 (и.)]

3. Статистически изводи и заключения относно относителни дялове в генералната съвкупност на базата на априорни вероятности

Безвъзвратен подбор, възвратен подбор, големи генерални съвкупности.

[Основна литература: Харалампиев 2004б: 54-58, 61-63 и 66-67 (к.)]

4. Статистически изводи и заключения относно важни параметри на разпределенията (средни аритметични, стандартни отклонения и т.н.) в генералната съвкупност на базата на априорни вероятности

Безвъзвратен подбор, възвратен подбор, големи генерални съвкупности.

[Основна литература: Харалампиев 2004б: 8-20, 31-42 и 49-53 (к.)]

* Тук и навсякъде по-надолу номерацията на страниците е давана по някои от следните два варианта:
(и.) – по номерацията на страниците в Интернет;
(к.) – по номерацията на страниците в съответната книга.

5. Приложения в областта на социалните науки – непредставителни извадки

Анкети в Интернет.

Телевизионни гласувания – „Песен на Евровизия”, „Вот на доверие”, „Сблъсък”, „Дуел”, „Big Brother”, „Star Academy”, „Музикална ку-ку академия”, „Тест на нацията” и т.н. Проблемът за „победителя”.

Студентска оценка за качеството на преподаване.

[Основна литература: Харалампиев 2004а; Харалампиев 2004б: 68-79 (к.); Харалампиев 2005; Харалампиев 2006]

6. Статистически изводи и заключения относно относителни дялове в генералната съвкупност на базата на апостериорни вероятности

Безвъзвратен подбор, възвратен подбор, големи генерални съвкупности.

Най-вероятно разпределение (най-вероятни относителни дялове) в генералната съвкупност.

[Основна литература: Харалампиев 2004б: 58-60 и 63-65 (к.); Haralampiev 2006: 1-7, 9-12 и 13 (и.)]

7. Статистически изводи и заключения относно важни параметри на разпределенията (средни аритметични, стандартни отклонения и т.н.) в генералната съвкупност на базата на апостериорни вероятности

Безвъзвратен подбор, възвратен подбор, големи генерални съвкупности.

[Основна литература: Харалампиев 2004б: 20-30 и 42-49 (к.)]

8. Приложения в областта на социалните науки – представителни извадки

Предизборни проучвания. Проблемът за процента спрямо действителните гласове.

[Основна литература: Haralampiev 2006]

9. Определяне на вероятността нова единица да попадне в конкретна група, формирана по значенията на един или няколко признака.

Определяне на вероятността нова единица да попадне в конкретна група, формирана по значенията на един или няколко признака при работа с априорни и с апостериорни вероятности.

Банкови правила за отпускане на кредити. Правила за определяне на застрахователни премии. Възможности за приложение в областта на маркетинговите и политическите изследвания.

[Допълнителна литература: Jaynes 2003: 1809-1815 (и.)]

10. Приложение в областта на прогнозирането – доверителни интервали на екстраполационни прогнози.

[Основна литература: Харалампиев 2004б: 79-89 (к.)]

Литература

- Харалампиев, К.** 2004а. Анкетите в интернет: възможност за статистически изводи и интерпретиране на резултатите. Социологически проблеми, брой 3-4. Достъпна в Интернет на адрес: http://www.geocities.com/kaloyan_haralampiev/statija9.pdf
- Харалампиев, К.** 2004б. Нетрадиционен поглед върху традиционни статистически проблеми. Балкани, София
- Харалампиев, К.** 2005. Телевизионните гласувания по телефона – проблемът за „победителя”. Социологически проблеми, брой 3-4. Достъпна в Интернет на адрес: http://www.geocities.com/kaloyan_haralampiev/statija13.pdf
- Харалампиев, К.** 2006. Студентската оценка на преподаването – проблемът за точността на изводите. Юбилеен сборник „30 години катедра „Социология””. Достъпна в Интернет на адрес: http://www.geocities.com/kaloyan_haralampiev/statija12.pdf
- Харалампиев, К.** 2007. За парадигмите в статистиката – байсовска статистика. Достъпна в Интернет на адрес: http://www.geocities.com/kaloyan_haralampiev/doklad1.pdf
- Bretthorst, L.** 1990. An Introduction of Parameter Estimation Using Bayesian Probability. In: Maximum Entropy and Bayesian Methods, P. Fougere (ed.), Kluwer Academic Publishers, Dordrecht the Netherlands. Достъпна в Интернет на адрес: <http://bayes.wustl.edu/glb/intro.pdf>
- Dose, V.** 2002. Bayes in Five Days. Lecture notes from a ten hour tutorial on Bayesian analysis given at the International Max-Planck Research School on bounded plasma. Достъпна в Интернет на адрес: <http://www.ipp.mpg.de/OP/Datenanalyse/Publications/Papers/dose02a.pdf>
- Haralampiev, K.** 2006. Bayesian Inference of Relative Frequency (In the Case of Electoral surveys). Annuaire de l'universite de Sofia “St. Kliment Ohridski”, Fakulte de philosophie, Livre – Sociologie, Tome 99. Достъпна в Интернет на адрес: http://www.geocities.com/kaloyan_haralampiev/article1.pdf
- Jaynes, E.** 1986. Bayesian Methods: General Background. In: Maximum-Entropy and Bayesian Methods in Applied Statistics, J. H. Justice (ed.), Cambridge University Press, Cambridge. Достъпна в Интернет на адрес: <http://bayes.wustl.edu/etj/articles/general.background.pdf>

- Jaynes, E.** 1988. The Relation of Bayesian and Maximum Entropy Methods. In: Maximum-Entropy and Bayesian Methods in Science and Engineering, Vol. 1, G. J. Erickson and C. R. Smith (eds.), Kluwer, Dordrecht. Достъпна в Интернет на адрес:
<http://bayes.wustl.edu/etj/articles/relationship.pdf>
- Jaynes, E.** 2003. Probability Theory: the Logic of Science. Cambridge University Press, Cambridge. В този курс е използвано непълно издание на тази книга, достъпно в Интернет на адрес: <http://omega.albany.edu:8008/JaynesBook.html>
- Loredo, T.** 1990. From Laplace To SN 1987A: Bayesian Inference In Astrophysics. In: Maximum Entropy and Bayesian Methods, P. F. Fougere (ed), Kluwer Academic Publishers, Dordrecht. Достъпна в Интернет на адрес:
<http://bayes.wustl.edu/gregory/articles.pdf>

Изготвил: гл. ас. д-р Калоян Харалампиев