Bolla Mariann - Csicsman József

Algoritmikus modellek és tanulóalgoritmusok a statisztikában

2011

Támogatás:
Készült a TÁMOP-4.1.2.A/1-11/1-2011-0064 számú, a „Természettudományos (matematika és fizika) képzés a műszaki és informatikai felsőoktatásban” című projekt keretében.

Készült:
a BME TTK Matematikai Intézet gondozásában

Szakmai felelős vezető:
Ferenczi Miklós

ISBN 978-963-279-893-6

Előszó

Jegyzetünk azoknak a hallgatóknak készült, akik matematikai statisztika és többváltozós statisztika tanulmányaik után szeretnék megismerni a modern statisztikai modelleket és módszereket is. A klasszikus statisztika fogalomrendszere és legtöbb tétele a XX. század első felében lett kidolgozva, elsősorban valószínűségszámítási alapokon. Ebben jelentős szerepet játszott az angolszász, orosz és indiai iskola. Érdekes, hogy olyan kulcsfontosságú eredmények, mint a Cramér–Rao egyenlőtlenség, Rao–Blackwellizálás, és a Wald-féle szekvenciális döntési eljárás a II. Világháború idején születtek meg, utóbbi töltények gazdaságos minőségellenőrzésére.

A XX. század közepére kifejlesztették a többváltozós statisztikai eljárásokat is, amelyek széleskörű alkalmazásának azonban csak a nagy teljesítményű számítógépek elterjedése nyitott utat a XX. század második felében (BMDP, SPSS programcsomagok), hiszen ezek a módszerek nagyméretű adatmátrixok és kovarianciamátrixok szinguláris- és spektrális felbontásán alapulnak. Nagyjából ezeket az ismereteket foglalja össze a BME matematikus képzés BSc és MSc statisztika anyagának gerincét képező Bolla–Krámli, Statisztikai következtetések elmélete (Typotex, 2005 és 2012) könyv.

Az 1970-80-as években azonban már ez a tényanyag sem bizonyult elégségesnek. Valós életbeli (biolológiai, pszichológiai, szociológiai) adatrendszerekkel foglalkozva azt találtuk, hogy a klasszikus módszerek nem alkalamazhatók mindig közvetlenül, illetve a probémák sokszor túlmutattak a tanult (elsősorban többváltozós normális eloszlású mintákra kifejlesztett) módszerek alkalmazhatósági körén (diszkrét, nem-paraméteres szituációk, időben is változó megfigyelések). L. Breiman, Statistical modeling: the two cultures (Statist. Sci. 16) 2001-es cikkében szintén rámutat arra, hogy gyakorlati problémákkal szembesülve a klasszikus apparátus néha csődöt mond. Az ún. második kultúra egy algoritmikus szemléletet visz a klasszikusba, ami azonban nem a numerikus módszerek automatikus alkalmazását jelenti, hanem olyan elméleti algoritmusok kifejlesztését, melyek az információelmélet, a Hilbert-terek, sőt akár a gráfelmélet eszköztárát használják magas színvonalon. Ebbe az eszköztárba szeretnénk betekintést nyújtani.

Ilyen módon a tankönyv egy, a modern statisztikai módszerek iránt érdeklődő hallgatók számára a BME-n kétévente tartott kurzus anyaga, de használható témalabor vagy diplomamunka készítéséhez is, illetve az elméleti részek kihagyásával a leírt algoritmusok nagyméretű adatrendszerek adatbányászatával foglalkozó szakemberek számára is hasznosak lehetnek. Az algoritmikus modellek köre egyre terjed, itt csak a legfontosabbakat foglaltuk össze, de utalunk egyéb, hasonló célú eljárásokra, illetve bőséges szakirodalmat közlünk a részletek iránt érdeklődőknek. Az elméleti részek tanulmányozása pedig az arra fogékony olvasók kezébe ötleteket és eszközöket adhat hasonló szituációk kezelésére.

Bolla Marianna,
Csicsman József

Budapest, 2013. július 5.

Tartalomjegyzék