

Előszó

Valószínűesszámitás és statisztika
számítógép algebrai támogatással

Készült a PTE PMMIK-n a
TÁMOP -4.1.2.A/1-11/1-2011-0098 támogatásával

Klincsik Mihály

A valószínűesszámitás és a statisztika tantárgy a természettudományos és a mérnöki képzések elválaszthatatlan része. Az általános és középiskolai tananyagokban is találkozhatunk játékos véletlen kísérletekkel, az esélyek számításával és a leíró statisztika elemeivel. Azonban a folytonos eloszlások tárgyalásához elengedhetetlen eszköz a függvénytan alapvető ismerete, így a differenciál- és integrálszámitás.

Ez a jegyzet azon hallgatók számára készült, akik most kezdik a valószínűesszámitást tanulni és a statisztikai fogalmakat elsajátítani a felsőoktatás alapszintjén. Szerencsére sok ilyen témával foglalkozó magyar nyelvű egyetemi jegyzettel találkozhatunk. Azonban a számítógép algebrai eszközöket is alkalmazó magyar nyelvű jegyzet még nem született. Mennyivel több és mennyiben más ez a jegyzet, mint egy hagyományos módon készült társa? Döntse el az olvasó illetve felhasználó a kérdést.

Az ma már természetes, hogy a statisztikai számításokhoz számítógépes programokat használunk. Egy jó táblázatkezelő elvégzi a számításokat a beépített függvényei segítségével, megjeleníti az adatokat és segít a statisztikai következtetések megalkotása során. A számítógép algebrai rendszerek képesek az algebrai, a geometriai és az analízisbeli műveleteket szimbolikus módon elvégezni, a levezetésekben segédkezni és megjeleníteni azokat. A tanulási folyamatot élővé teszik, kísérletezni lehet velük és tárolják az ismereteket szimbolikus szinten. Napjainkban a valószínűesszámitási modellek, adott eloszlású valószínűségi változók és a velük végezhető műveletek is beépültek a szimbolikus algebrai rendszerekbe. Mi a Maple számítógép algebrai rendszer ilyen eszközeit mutatjuk be úgy, hogy közben tárgyaljuk a megfelelő matematikai eszközöket is. Tesszük ezt majdnem 20 éves Maple felhasználói tapasztalattal és több, mint 10 éves valószínűesszámitási és statisztikai oktatási tapasztalattal.

A tananyag felépítése nagyrészt megegyezik a magyarországi egyetemeken indított valószínűesszámitás és statisztika kurzusok tananyagaival. Az 1. fejezet a véletlennel összefüggő kísérletek kimeneteleinek leírásával és az eseményalgebra halmaz alapú tárgyalásával foglalkozik. A 2. fejezetben a relatív gyakoriságok alapvető tulajdonságai után a Kolmogorov-féle axiómákkal ismerkedünk meg. Klasszikus valószínűségi mezőre bőszegesen találunk példát a szerencse játékok között. Ezek lehetséges Maple szimulációit mutatjuk be. A folytonos eseménytér tárgyalása indokolja ezen a szinten a σ -algebra és a mérték fogalmával való megismerkedést. A 3.1. fejezetben a kombinatorika elemeit és ezek Maple számításait tárgyaljuk. Kiemeljük, hogy egy bizonyos halmaz elemeinek megszámlálására a Maple kínálja a programozási lehetőséget. Ezzel ellenőrizhetjük, hogy az elméleti úton történt számításaink egyeznek-e a halmaz elemeinek gyakorlati megszámlálásával. A 3.2. fejezet a folytonos eseménytérre vezető véletlen kísérletek valószínűségeit geometriai és Maple eszközökkel modellezi. Mivel Maple-ben felvehetünk egyenletes eloszlású véletlen változót és azzal szimbolikus műveleteket végezhetünk, ezért egyszerűbb esetekben automatikusan megkapjuk a kézzel kiszámolt valószínűségeket. Érdemes tehát ezen a szinten is bevonni a számítógép algebrai eszközt a megismerés folyamatába. A feltételes valószínűség fogalma, az erre épülő szorzás-szabály a témája a 4. fejezetnek. A fogalmak szemléltetésére alkalmas a valószínűségi vagy más néven döntés fa, amely tartalmazza a feltételes valószínűségeket és a szorzás-szabályt is. Az események függetlenségével és olyan hálózatok megbízhatósági számításával foglalkozik az 5. fejezet, melyben az alrendszerek párhuzamos és sorosan kapcsolódnak egymáshoz. A teljes valószínűség-tétel és a Bayes-tétel a témája a 6. fejezetnek. A tételek számításait beépítjük a döntési fa diagramba és annak inverzébe. Összevetjük a fa diagrammal való adat interpretációt és a kontingencia táblázattal való valószínűségek megjelenítését.

A 7. fejezet a valószínűségi változók és eloszlásaik leírásával foglalkozik. A szimbolikus

algebrai eszközök hatékony alkalmazása itt kezdődik. Megjegyezzük, hogy az előző fejezetekben is impliciten benne van a valószínűségi változó alkalmazása, de explicit kifejtése itt kezdődik. A diszkrét és folytonos esetek tárgyalásait szétválasztjuk. Az összekötő fogalom az eloszlásfüggvény és a szétválasztó fogalom az eloszlás és megfelelően a sűrűség függvény. A 8. fejezetben a valószínűségi változók legfontosabb jellemzőit a várható értéket, szórást definiáljuk és számítjuk. A 9. fejezetben a nevezetes diszkrét eloszlások közül az egyenletes, Bernoulli, binomiális, Poisson, hipergeometrikus és geometriai eloszlásokat tárgyaljuk példákkal és Maple megadásokkal. A 10. fejezetben a nevezetes folytonos eloszlások közül az egyenletes, exponenciális, gamma, normál, khi-négyzet és a Student t-eloszlást tárgyaljuk. A két dimenziós eloszlásokat tárgyaljuk a 11. fejezetben. A diszkrét és folytonos eseteket külön választjuk. A feltételes eloszlást, a feltételes eloszlás függvényt és feltételes sűrűség függvényt definiáljuk, számoljuk és szemléltetjük. A valószínűségi változók függetlensége a statisztikában fontos fogalom, ezért megismerkedünk vele. A nevezetes 2 dimenziós eloszlások közül a polinomiális, normál és exponenciális eloszlásokat tárgyaljuk. A vizsgálatokból kiderül, hogy nem független változók megadása igen komplikált módon képezhető. Két változó összege eloszlásának kiszámolásához az együttes eloszlás fogalma is kell. Ezért tárgyaljuk ezt külön a 12. fejezetben. A 13. fejezet a nagyszámok törvényével és a centrális határeloszlással foglalkozik. A Maple szemléltető ereje itt valóban megmutatkozik.

A 14. fejezettől kezdődnek a statisztikai vizsgálatok. Ismertetjük a minta várható értékének, szórásának, mediánjának és móduszának pontbecsléseit és ezek tulajdonságait. A Maple Statistics csomagjában a Mean, StandardDeviation, Median és Mode leírásokat úgy írták meg, hogy ha valószínűségi változó az inputja, akkor arra számolja a megfelelő értékeket és ha egy statisztikai adatsor, akkor annak megfelelő pont becslését számolja. Az átlag és szórás megbízhatósági becslését tárgyalja a 15. fejezet normál eloszlású adatok feltételezésével. Az átlag becslésénél külön kell választani azt az esetet, amikor ismert a szórás, attól amikor ismeretlen. Ugyanis első esetben normál eloszlást, míg a másodikon t-eloszlást kell használnunk. A mintaméret számításánál alkalmazzuk a Monte Carlo-módszert és ezzel válik érzékelhetővé, hogy a Maple hogyan támogatja a véletlen folyamatok törvényszerűségeinek megértését. A 16. fejezetben bemutatjuk a statisztikai hipotézisek tesztelési folyamatát, valamint a tesztelés és becslés kapcsolatát és az elkövetett első és másodfajú hibákat. Felrajzolunk működési jelleggörbét vagy OC (Operating Characteristic) görbét és erőfüggvényt. A khi-négyzet modell tesztet már korábban is használtuk az eloszlás normalitásának tesztelésére. Most egy újabb módszert, az illesztés jósógának tesztjét alkalmazzuk hasonló célra. Végül a 17. fejezetben a lineáris regressziós egyenes illesztésével ennek statisztikai tesztelésével foglalkozunk. Teszteljük a lineáris modell alkalmazhatóságát statisztikai és grafikus úton.

Láthatóan a jegyzet 3 nagy részre tagolható. A valószínűségszámítás klasszikusnak mondható eredményeit az 1-6. fejezetek tartalmazzák. A valószínűségi változók és a hozzá kapcsolódó határérték eredményeket a 7-13. fejezetben találjuk meg. Míg a harmadik rész a 14-17. fejezetekben található és a statisztikai becsléseket, tesztek és a lineáris regressziót foglalja magában.

A jegyzetben leírt Maple kódokat akkor tudja kipróbálni az olvasó, ha van Maple programja (<http://www.maplesoft.com/>) .

Pécs, 2014. március