Villamosmérnök A4

12. gyakorlat (2012. dec. 3-4.)

Nagy számok törvénye, centrális határeloszlás-tétel, konfidencia-intervallumok

Nagy számok törvénye (NSZT): legyenek $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók, melyek közös várható értéke: $\mu \in \mathbb{R}$. Ekkor:

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \longrightarrow \mu, \text{ amint } n \to +\infty.$$

Szóban: az időbeli átlag a térbeli (a súlyfüggvény vagy sűrűségfüggvény szerinti) átlaggal egyezik meg, avagy az empirikus átlag jól közelíti a „valódi" átlagot/várható értéket. Speciálisan, ha az X_ik Bernoulli valószínűségi változók, akkor az empirikus valószínűség az elméleti valószínűséget közelíti!

Centrális határeloszlás-tétel (CHT): legyenek $X_1, X_2, \ldots$ független és azonos eloszlású valószínűségi változók, $\mu := \mathbb{E}X_i \in \mathbb{R}$ és $\sigma := \mathbb{SD}(X_i) \in \mathbb{R}^+$. Ekkor, ha n elég nagy, akkor a valószínűségi változók standardizált összege közelítőleg (standard) normális eloszlású, azaz

$$\lim_{n \to +\infty} \mathbb{P}\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} < x \right) = \Phi(x) \quad (x \in \mathbb{R})$$

Másképp: kicsi szórású független hatások összege normális eloszlású. Speciálisan: az X_i-ket azonos p paraméterű Bernoulli változóknak választva kapjuk a **de Moivre–Laplace tételt**: a $(\mathsf{Bin}(n,p) - np)/\sqrt{np(1-p)}$ standardizált binomiális eloszlás megközelítőleg standard normális eloszlású.

Konfidencia-intervallumok μ-re, ismert σ esetén. Az $N(\mu, \sigma)$ eloszlás n elemű mintájából kapott $\bar{x}_n$ mintaátlaggal szerkesztett p valószínűségű szimmetrikus konfidencia-intervallum:

$$\left(\bar{x}_n - \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{1+p}{2}\right), \bar{x}_n + \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{1+p}{2}\right) \right).$$

Ez a véletlen $\bar{x}_n$ középpontú intervallum p valószínűséggel olyan, hogy tartalmazza a számunkra ismeretlen μ várható értéket. (Ha már megvan $\bar{x}_n$, akkor az intervallum vagy tartalmazza μ-t, vagy nem, nincs benne véletlen, de az így szerkesztett intervallumok az esetek p részében jók.)

A p valószínűségű egyoldali konfidencia-intervallum:

$$\left(\bar{x}_n - \frac{\sigma}{\sqrt{n}} \Phi^{-1}(p), \infty \right).$$

Konfidencia-intervallumok μ-re, ismeretlen szórás esetén. Ilyenkor σ helyett a minta s_n^* szórásával kell számolnunk, és a normális Φ helyett az úgynevezett $n-1$ szabadsági fokú Student t-eloszlás $F_{n-1}(t)$ eloszlásfüggvényével:

$$\left(\bar{x}_n - \frac{s_n^*}{\sqrt{n}} F_{n-1}^{-1}\left(\frac{1+p}{2}\right), \bar{x}_n + \frac{s_n^*}{\sqrt{n}} F_{n-1}^{-1}\left(\frac{1+p}{2}\right) \right).$$

Az F_ν is egy 0-ra szimmetrikus haranggörbeszerű eloszlás, kvantilisei megtalálhatóak a csatolt táblázatban $\nu = 120$-ig. Nagy ν-re a Φ értékeihez konvergálunk, tehát mondjuk $n \geq 250$ mintaméret esetén már teljesen rendben van, ha a Φ táblázatot használjuk.

Hasonlóképpen, az egyoldali konfidencia-intervallum:

$$\left(\bar{x}_n - \frac{s_n^*}{\sqrt{n}} F_{n-1}^{-1}(p), \infty \right).$$

Ha a mintáink nem normális eloszlásból származnak, de sokan vannak, akkor a CHT szerint a mintaátlag már közelítőleg normális, és lehet használni a fenti intervallumokat, ismert vagy ismeretlen σ-ra is.

1. Tudor Tódor egyetemi hallgató. A képzése során összesen 40 különböző kötelező tárgyat kell teljesítenie, hogy diplomát kaphasson. Az egyszerűség kedvéért tegyük fel, hogy ezek egyike sem épül a másikra és, hogy minden őszi szemesztert 8 (új) tárgy felvételével kezdi. Minden tárgyat a többitől függetlenül 0.7 valószínűséggel teljesít elsőre, és 0.3 valószínűséggel elhasal. Egy félévben egy tárgyból legfeljebb háromszor próbálkozhat. Átlagosan hány félév alatt teljesít egy tárgyat? Mi annak a valószínűsége, hogy 10 féléven belül befejezi az egyetemet?

2. Otthon tartunk mindig 100 égőt, melyek élettartama egymástól független exponenciális eloszlású, 5000 óra várható értékkel. Az égőket egymás után használjuk, amint kiégett egy, azonnal kicseréljük a következőre. Az égők hány százaléka ég ki kevesebb, mint 3000 óra alatt?

3. Egy telefonközpontba nagyon sok hívás érkezik be. Tegyük fel, hogy a hívások exponenciális időközönként érkeznek, átlagosan 4 másodpercenként. Mi annak a valószínűsége, hogy 10 óra leforgása alatt több, mint 1000 hívás érkezik be?

4. Nyári csillagos éjszakákon átlag 3 csillaghullást látunk. Mi annak a valószínsűge, hogy egy teljes nyáron (100 napon) több mint 200 csillaghullást látunk? Mi az esélye, hogy több mint 350, de kevesebb, mint 500 csillaghullást látunk?

5. A Színház- és Filmművészeti Egyetemre köztudottan nagyon nehéz bekerülni. A leendő művészek felvételijén a bizottság 0.25 valószínűséggel ítéli meg jónak és eredetinek az alakítást, 0.75 valószínűséggel elutasító. A felvételizők azonban nagyon kitartóak, sokszor is próbálkozhatnak. Mi annak a valószínűsége, hogy 500 végzett színész anno összesen legalább 1000 felvételi meghallgatáson vett részt?

6. Egy átlagos nap egy élelmiszerboltban körülbelül 1000 vásárlás történik. Tudjuk, hogy ha a vásárolt termékek összegének vége 2.5 és 7.5 forint közé esik, akkor 5 forintra, míg egyéb esetekben a legközelebbi tízesre kerekítünk. Tegyük fel, hogy az összeg vége egyenletes eloszlású a $[0, 10]$ intervallumban. Mi annak a valószínűsége, hogy egy nap a bolt a kerekítéseknek köszönhetően

   (a) plusz 1500 forintnál jobb
   (b) 500 és 10000 forint közötti
   (c) nem pozitív
   (d) -1000 és 1000 forint közötti

   egyenleget könyvelhet el?

7. Egy feleletválasztós felvételi teszt 200 kérdést tartalmaz: minden kérdésnél 4 opció van, melyekből pontosan egy helyes.

   (a) Véletlen Viktor véletlenül betévedt a felvételire és minden kérdésnél véletlenül (egyenlő valószínűséggel) megjelölt egy választ. Mi a valószínűsége, hogy felveszik, ha 50 ponttól érvényes a felvételi?
   (b) Mi annak a valószínűsége, hogy pontosan 100 kérdést talál el Viktor?
   (c) A kérdéssor összeállítói szeretnék meghúzni úgy az érvényes felvételi alsó határát, hogy a Véletlen Viktorhoz hasonló tippelőket 90%-os eséllyel ki tudják szűrni. Hány pont legyen az alsó határ?

8. Átlagosan 1000 férfiból 440 magasabb, mint 180 cm. Mi annak a valószínűsége, hogy 100 véletlenül kiválasztott férfi közül legalább 30 alacsonyabb, mint 180 cm?

9. Hányszor kell érmével dobnunk ahhoz, hogy 0.95-nél nagyobb valószínűséggel a fej eredmények száma a dobások számának 49%-a és 51%-a közé essen?

10. Mennyi annak a valószínűsége, hogy 12000 kockadobás során előforduló hatosok száma 1900 és 2150 közé esik?

11. Mi a valószínűsége, hogy a villamosmérnök hallgatók beférnek egy 360 fős előadóban tartott A4 előadásra, ha 500-an vannak, és mindenki, a többiektől függetlenül 0.7 valószínűséggel hallgatja az előadást? Mi a helyzet, ha a terem 370 vagy 380 vagy 390 fős?

12. Egy repülőjáratra, amely minden nap többször is indul, 200 ember fér fel. Mindenki a többiektől függetlenül az esetek 10%-ban lekési a járatát (nem jelenik meg az indulásig). Ezért a repülőtársaság úgy gondolja, hogy növelheti a hasznát, ha 200-nál kicsivel több jegyet ad el, hiszen 200-nál kicsivel több jegy esetén csak kicsi a valószínűsége annak, hogy az eladott extra jegyek miatt bajba keveredik. Hány jegyet adjon hát el, ha azt szeretnék, hogy a járatok harmadánál/tizedénél/századánál forduljon csak elő, hogy valakinek nem jut hely? Hogyan módosul a helyzet, ha az emberek 0.1%-a késik csak el?

13. Egy trópusi országban minden nap esik az eső. Sok éves statisztika alapján kiszámoltuk, hogy egy nap átlagosan 10 mm csapadék esik, a napi szórás pedig 2 mm. Mi a valószínűsége, hogy 100 nap alatt 900 mm-nél is több esik? Mi az esélye, hogy 800 és 1200 mm között lesz a leeső csapadék mennyisége?

14. Mennyi a valószínűsége annak, hogy 50 darab független, azonos eloszlású valószínűségi változó összege a $[0, 30]$ intervallumba esik, ha az eloszlás $[0, 1]$ intervallumon

   (a) egyenletes?
   (b) $f(x) = 2x$ sűrűségfüggvény szerint alakul?

15. Rezső rulettezésre adja fejét a kaszinóban. Taktikája nagyon egyszerű: minden egyes körben 10 dollárt tesz a pirosra. Azt tapasztalja 100 játék után, hogy 300 dollár a vesztesége. Jogos-e a gyanúja, hogy svindlizik a croupier? (A rulett-körön összesen 37 mező van 0-tól 36-ig számozva. Ezek közül egy — a 0 jelű — zöld, a fennmaradó 36-ból pedig 18 piros és 18 fekete.)

16. Egy teherautóra 30 ládát pakolnak fel. Az egyes ládák tömegéről csak annyit tudunk, hogy 10 és 20 kg között egyenletes eloszlásúak, egymástól függetlenül. Mennyi az össztömeg várható értéke és szórása? Mi a valószínűsége, hogy a teljes tömeg nem haladja meg a 470 kg-ot? Maximum hány ládát lehet a teherautóra pakolni, hogy 0.99 valószínűséggel az össztömeg ne haladja meg az 500 kg-os maximális engedélyezett össztömeget?

17. (a) Adott 100 égőnk, melyek élettartama egymástól független exponenciális eloszlású, 5 óra várható értékkel. Tegyük fel, hogy az égőket egymás után használjuk, azonnal kicserélve azt, amelyik kiégett. Becsüljük meg annak valószínűségét, hogy 525 óra után még van működő égőnk!

    (b) Oldjuk meg az előző feladatot úgy, hogy minden égő kicserélése független, a $(0, 0.5)$ intervallumon egyenletes eloszlású ideig tart. Becsüljük meg most annak (is) a valószínűségét, hogy 550 óra elteltével már az összes égő kiégett!

18. Mutassuk meg a CHT alapján, hogy 2 független standard normális eloszlású valószínűségi változó összege is normális 0 várható értékkel és 2 szórásnégyzettel!

19. Idei kékfrankosunk alkoholtartalmát mérjük. Mérési módszerünknek van egy kis véletlen hibája, 0 várható értékkel, 0.5% szórással. Tíz mérés átlaga 12.61% volt. Konstruáljunk egy szimmetrikus 95%-os konfidenciaintervallumot.

20. Öt hallgató egy laborban a következő mérési eredményeket kapta elvileg ugyanarra az áramerősségre egy kissé lestrapált ampermérővel, melynek szórása $2A$: 15.2, 15.7, 14.6, 16.8 illetve $13.9A$. Adjon meg olyan konfidencia-intervallumot, amely az áramerősség várható értékét 0.8, 0.9, illetve 0.99 valószínűséggel tartalmazza! Hány kísérlettel lehetne $1.2A$ hosszúságú 95%-os konfidencia-intervallumot megadni?

21. Egy dobozban piros illetve kék üveggolyók vannak. A pirosak p arányát nem ismerjük. Húzunk visszatevéssel 250 golyót, amiből 102 lett piros. Konstruáljunk egy 95%-os konfidenciaintervallumot p becslésére.

22. Legszebbország miniszterelnöke, Okos Tóbiás úgy érzi, zseniális ötlete támadt a költségvetési hiány kezelésére: bezárja az ország egyetemeit. Hogy biztos legyen a dolgában, megkér egy közvéleménykutató-intézetet, szondázza a lakosság véleményét. Egy ezerfős mintából 511-en ellenezték az ötletet, 102-en támogatták, 387 embernek nem volt véleménye. Mekkora önbizalommal jelentheti ki a miniszterelnök, hogy a lakosság többsége nem ellenzi az ötletet?

23. A 2002-es Salt Lake City Téli Olimpia 1500m-es gyorskorcsolya számában fölmerült, hogy a külső pálya egy kicsit gyorsabb, mint a belső. 24 párban futottak a versenyzők, a 13-adikban volt egy bukás, a többi 23 futam eredménye a táblázatban látható.

    Mivel sok különböző, többé-kevésbé független tényező befolyásolja, hogy egy véletlenül választott versenyző mennyi idő alatt ér célba, föltehetjük, hogy a páronkénti időkülönbségek valamilyen $N(\mu, \sigma)$ normális eloszlást követnek. Az adatok alapján mekkora biztonsággal állítható, hogy tényleg jelentett számottevő előnyt a külső pálya, tehát mondjuk $\mu \geq 0.5$ másodperc?

Race number	Inner lane	Outer lane	Difference
1	107.04	105.98	1.06
2	109.24	108.20	1.04
3	111.02	108.40	2.62
4	108.02	108.58	−0.56
5	107.83	105.51	2.32
6	109.50	112.01	−2.51
7	111.81	112.87	−1.06
8	111.02	106.40	4.62
9	106.04	104.57	1.47
10	110.15	110.70	−0.55
11	109.42	109.45	−0.03
12	108.13	109.57	−1.44
14	105.86	105.97	−0.11
15	108.27	105.63	2.64
16	107.63	105.41	2.22
17	107.72	110.26	−2.54
18	106.38	105.82	0.56
19	107.78	106.29	1.49
20	108.57	107.26	1.31
21	106.99	103.95	3.04
22	107.21	106.00	1.21
23	105.34	105.26	0.08
24	108.76	106.75	2.01
Mean	108.25	107.43	0.82
St.dev.	1.70	2.42	1.78

24. Környezetvédők 16 mintát vettek egy vegyigyár szennyvízéből, és egy rákkeltő anyag arányát mérték: $\bar{x}_{16} = 2.24 \times 10^{-6}$, $s^*_{16} = 1.06 \times 10^{-6}$. Be akarják perelni a céget, hiszen ezen adatok alpján 97.5%-os biztonsággal kijelenthető, hogy a koncentráció túllépi az a értéket, ami az EU szabvány. Viszont a gyár szerint az a mérvadó, hogy az adatok szerint a koncentráció 97.5%-os biztonsággal nem lépi túl a magyar szabvány b értéket. Mennyi a és b?

Quantiles of the t Distribution with ν Degrees of Freedom

	p								
ν	.6	.7	.8	.9	.95	.975	.99	.995	.999
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657	318.309
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750	3.385
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704	3.307
45	0.255	0.528	0.850	1.301	1.679	2.014	2.412	2.690	3.281
50	0.255	0.528	0.849	1.299	1.676	2.009	2.403	2.678	3.261
55	0.255	0.527	0.848	1.297	1.673	2.004	2.396	2.668	3.245
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660	3.232
65	0.254	0.527	0.847	1.295	1.669	1.997	2.385	2.654	3.220
70	0.254	0.527	0.847	1.294	1.667	1.994	2.381	2.648	3.211
75	0.254	0.527	0.846	1.293	1.665	1.992	2.377	2.643	3.202
80	0.254	0.526	0.846	1.292	1.664	1.990	2.374	2.639	3.195
85	0.254	0.526	0.846	1.292	1.663	1.988	2.371	2.635	3.189
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632	3.183
95	0.254	0.526	0.845	1.291	1.661	1.985	2.366	2.629	3.178
100	0.254	0.526	0.845	1.290	1.660	1.984	2.364	2.626	3.174
105	0.254	0.526	0.845	1.290	1.659	1.983	2.362	2.623	3.170
110	0.254	0.526	0.845	1.289	1.659	1.982	2.361	2.621	3.166
115	0.254	0.526	0.845	1.289	1.658	1.981	2.359	2.619	3.163
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.676	3.090