Kísérlettervezés (kiegészítő anyag) – Megoldások

Adatelemzés – elméleti ismeretek

Valószínűségszámítási alapfogalmak

- Valószínűségi változó (*random variable*): X
- Várható érték, átlag (*expected value*, *average*, *mean*):

$$\mu = \mathbb{E}X = \sum_{i=1}^{n} p_i x_i$$

- Szórásnégyzet (*variance*):

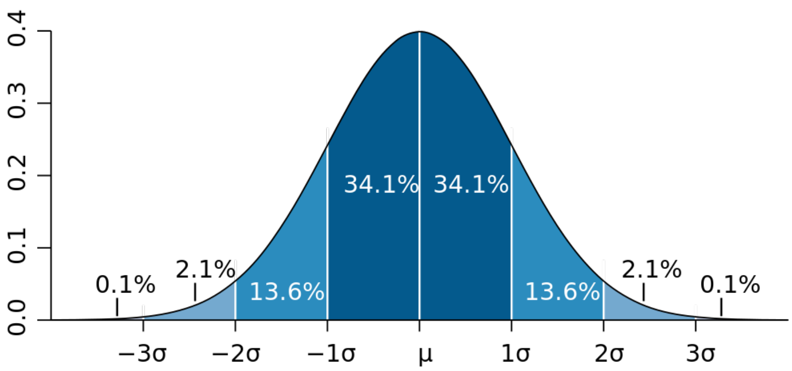
$$\sigma^2 = \mathbb{E}\,(X-\mu)^2 = \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

- Szórás (*standard deviation*):

$$\sigma = \sqrt{\mathbb{E}\,(X-\mu)^2} = \sqrt{\sum_{i=1}^{n} p_i (x_i - \mu)^2}$$

Konfidencia – normális eloszlás

A normális eloszlású változó
- az esetek 68%-ában legfeljebb 1σ messze kerül μ-től,
- az esetek 95%-ában legfeljebb 2σ messze kerül μ-től,
- az esetek 99,7%-ában legfeljebb 3σ messze kerül μ-től.



1. ábra. Konfidenciaintervallumok

Statisztikai alapfogalmak

- Megfigyelések: t darab, $x_1, \ldots, x_t$
- Tapasztalati átlag (*sample mean*):

$$m = \bar{x} = \frac{x_1 + \cdots + x_t}{t}$$

- Korrigált tapasztalati szórás (*unbiased sample standard deviation*):

$$s = \sqrt{\frac{(x_1 - m)^2 + \cdots + (x_t - m)^2}{t-1}} = \sqrt{\frac{\sum_{i=1}^{t} (x_i - m)^2}{t-1}}$$

Figyeljük meg, hogy a korrigált tapasztalati értékeknél t helyett $(t-1)$-gyel osztunk. Ennek oka, hogy t-vel osztva a kapott érték általában alábecsli a teljes populáció szórását. Belátható, hogy $(t-1)$-gyel osztva a valódi szórást jobban közelítő értéket kapunk. Ezt nevezzük Bessel-féle korrekciónak (`https://en.wikipedia.org/wiki/Bessel's_correction`).

Kísérlettervezés

A centrális határeloszlás-tételből (*central limit theorem*) következik, hogy tetszőleges eloszlású jellemző (véges μ várható értékkel és σ szórással) tapasztalati átlaga (m valószínűségi változó) $t \to \infty$ esetén normális eloszlású, μ várható értékkel és $\sigma_m = \frac{\sigma}{\sqrt{t}}$ szórással (ahol σ a megfigyelés alatt álló valószínűségi változó szórása).
Ökölszabály: ismert szórásnál $t > 30$ után kezd elfogadható lenni ez a közelítés.
A fentiek miatt a megfigyeléseink kiértékeléséhez használhatjuk a normális eloszláshoz tartozó konfidenciaintervallumokat. Eszerint m 68%-os valószínűséggel maximum σ_m távolságra esik μ-től. Tehát (ha m-et és σ_m-et ismerjük, és μ-t szeretnénk ez alapján becsülni) μ értékéről kijelenthetjük, hogy

- 68% valószínűséggel $m - \sigma_m$ és $m + \sigma_m$ közé esik, valamint ehhez hasonlóan
- 95% valószínűséggel $m - 2\sigma_m$ és $m + 2\sigma_m$ közé esik,
- és 99,7% valószínűséggel $m - 3\sigma_m$ és $m + 3\sigma_m$ közé esik.

Sokszor azonban σ nem ismert (a mérésekre azért van szükség, hogy becsléseket kaphassunk μ és σ értékére), ekkor elfogadható a $\sigma_m \approx \frac{s}{\sqrt{t}}$ (ahol s a korrigált tapasztalati szórás), amennyiben $t \geq 100$.

1. Kísérlet kiértékelése

Infrastruktúránk méretezését megnehezíti, hogy egy adott feladattípus végrehajtási ideje a körülmények függvényében ingadozik, például lapozás, memória szemétgyűjtés, memória cache találatok stb. változékonysága folytán. Ezért összeállítottunk egy valós munkaterhelést jól jellemző benchmarkot, és ennek többszöri lefuttatása során a futási időket átlagolva szeretnénk meghatározni a rendszer átlagos teljesítményét.

a) Az első tíz futtatás eredményei: 37 s, 34 s, 35 s, 39 s, 57 s, 41 s, 36 s, 35 s, 61 s, 35 s. Mennyi ez alapján a rövid kísérlet alapján a tapasztalati átlag és tapasztalati szórás?
Megoldás
Vajon mitől lehetnek ilyenek az értékek?
A tapasztalati átlag:

$$m = \frac{37 + 34 + 35 + 39 + 57 + 41 + 36 + 35 + 61 + 35}{10} = 41 \text{ [s]}$$

Tapasztalati átlagtól eltérések és eltérések négyzetei:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	
x_i	37	34	35	39	57	41	36	35	61	35	[s]
$x_i - m$	-4	-7	-6	-2	$+16$	0	-5	-6	$+20$	-6	[s]
$(x_i - m)^2$	16	49	36	4	256	0	25	36	400	36	[s^2]

Tapasztalati átlagtól eltérések: mennyi ezeknek az összege/átlaga? Miért 0?
Tapasztalati átlagtól eltérések négyzete: jól érezhető, hogy a nagy eltérést jobban büntetjük, összes négyzetes eltérés 858 s^2. Innen eltérések négyzetes közepe (átlagos négyzetes eltérés gyöke):

$$s^* = \sqrt{\frac{(x_1 - E)^2 + ... + (x_t - E)^2}{t}} = \sqrt{\frac{858 \text{ s}^2}{10}} \approx 9{,}26 \text{ s}$$

Ez lenne a sokaság szórása, ha ez a 10 adatpont lenne a teljes sokaság. Mivel ez a 10 adatpont csak a sokaságból vett minta, valójában $(t-1)$-et (jelen esetben 9-et) kell a nevezőbe írni, hogy az úgynevezett korrigált szórást kapjuk:

$$s = \sqrt{\frac{(x_1 - E)^2 + ... + (x_t - E)^2}{t - 1}} = \sqrt{\frac{858 \text{ s}^2}{9}} \approx 9{,}76 \text{ s}$$

Ez utóbbi már valóban az eredeti valószínűségi változó szórását közelíti (ha nagyon sokszor ismételnénk meg ezt a 10 hosszú kísérletsorozatot, a tapasztalati szórás várható értéke a szórás lenne).

b) Nagyobb léptékben futtatva a kísérletet, a benchmark 10 000 futtatása átlagban 44,3 másodpercig tartott, 11,6 másodperc tapasztalati szórással. Mennyire lehetünk biztosak a kapott eredmény pontosságában?
Megoldás
Mivel jóval több mint 100 megfigyelésből állt a kísérlet, a tapasztalati átlag jó közelítéssel normális eloszlással szór a tényleges várható érték körül. Ennek a Gauss-haranggörbének a szórása (az ismeretlen valószínűség változó szórását a kísérletből adódó tapasztalati szórással helyettesítve):

$$\sigma_m = \frac{\sigma}{\sqrt{t}} \approx \frac{s}{\sqrt{t}} \approx \frac{11{,}6\text{ s}}{\sqrt{10\ 000}} = 0{,}116\text{ s}$$

Tehát azt mondhatjuk, hogy a benchmark várható végrehajtási ideje $44{,}3\text{ s} \pm 0{,}1\text{ s}$. A normális eloszlás konfidenciaintervallumairól tanultak alapján 99,7% konfidencia mellett jelenthetjük ki, hogy a végrehajtási idő várható értéke a $[44{,}0\text{ s}, 44{,}6\text{ s}]$ intervallumba esik.

2. Kísérlettervezés

Egy modellezett folyamat átbocsátóképességére szimuláció alapján szeretnénk egy közelítő értéket és hozzá tartozó konfidencia-intervallumot meghatározni.

a) Hány szimuláció mérési eredményeiből számoljunk átlagot?
Megoldás
Ha még nincsenek közelítéseink a szórásra, az ökölszabály szerint legalább 100 megfigyelést kell végezni.

b) Az így elvégzett mérési eredmények tapasztalati közepe 500 kérés/s; a tapasztalati szórás 10%. Szeretnénk, hogy 95% konfidencia mellett egy legfeljebb 40 kérés/s széles intervallumba essen az átbocsátóképesség. Hány mérést végezzünk még?
Megoldás
A tapasztalati szórás 10%, azaz $s = 10\% \cdot 500$ kérés/sec $= 50$ kérés/sec.
Ha a 95%-os (vagyis 2 szórásnyi sugarú) konfidenciaintervallum szélessége (sugár kétszerese) maximum 40 kérés/sec, akkor a normális eloszlás szórása (σ_m) maximum 10 kérés/sec lehet.
Ez a tapasztalati átlag mint valószínűségi változó szórása a tényleges várható érték körül, értéke:

$$\sigma_m = \frac{\sigma}{\sqrt{t}} \approx \frac{s}{\sqrt{t}} \approx \frac{50\text{ kérés/sec}}{\sqrt{t}} \leq 10\text{ kérés/sec}$$

Innen $5 \leq \sqrt{t}$, azaz legalább 25 megfigyelésből kell számolni az átlagot. Persze az egész normális eloszlásos közelítés csak 100 megfigyelés fölött alkalmazható – de ennyi megfigyelést már el is végeztünk, tehát nem kell még többet mérni. A jelenlegi kísérlet eredménye alapján:

$$\sigma_m = \frac{s}{\sqrt{t}} \approx \frac{50\text{ kérés/sec}}{\sqrt{100}} = 5\text{ kérés/sec}$$

Tehát a $2\sigma_m$ sugarú intervallum szélessége 20 kérés/sec, így kétszeres pontosságot garantálhatunk 95% konfidencia mellett. (Avagy 40 kérés/sec széles (négy szórásnyi sugarú) intervallumot is garantálhatnánk 99,9936% konfidenciával.)