Задача на германски тенкови

Филип Николовски

Машински факултет, Скопје
Универзитет „Св. Кирил и Методиј“ во Скопје

Шести семинар „Математика и примени“, 17 март 2023
Институт за математика, Природно-математички факултет
Универзитет „Св. Кирил и Методиј“ во Скопје

Зошто германски тенкови?

Малку историски контекст:

- Периодот е 1940/1941 година и Втората светска војна е во ек
- Вие сте аналитичар кој работи за военото разузнување на Сојузниците
- Сојузничките војски постојано уништуваат или заробуваат германска воена опрема, вклучително и тенкови
- Вам ви поставуваат задача: врз основа на сериските броеви на тенковите да се изврши проценка на вкупниот број на тенкови со кои располага германската армија

Прашање:

Како ќе се справите со оваа задача?

Преведено на статистички јазик

Под претпоставка дека нумерирањето на тенковите е последователно и почнува со 1, поставената задача се сведува на следново:

Да се конструира оценувач $\widehat{N}$ за максимумот N на дискретна рамномерна распределба дефинирана на множеството $\Omega = \{1, 2, \ldots, N\}$ врз основа на примерок $S = \{x_1, x_2, \ldots, x_n\} \subset \Omega$ со големина n, ако најголемата вредност во примерокот е m, т.е. ако $\max S = m$

Ќе разгледаме два оценувачи на максимумот: едниот едноставен и наивен, а другиот со посложена математичка дефиниција

Наивен оценувач

- Наједноставен оценувач е очигледниот:

$$\widehat{N} = \max S = m,$$

т.е. да сметаме дека максимумот на распределбата се совпаѓа со максималната набљудувана вредност во примерокот

- Иако ова е сосема оправдано, сепак се покажува дека ваквата оценка во пракса не е најдобра (во просек дава не баш добри резултати)
- Ова најмногу се должи на фактот што оценувачот ги занемарува останатите елементи на примерокот и, секако, неговата големина

Подобрен оценувач

- Во продолжение ќе покажеме како може да се конструира оценувач $\widehat{N}$ за вредноста на N со подетална анализа на примерокот
- Нека M е случајна променлива која го означува максимумот примерокот, т.е. најголемата наљудувана вредност
- Поаѓаме од следново прашање: колку изнесува веројатноста при случаен избор на примерок $S = \{x_1, x_2, \ldots, x_n\} \subset \Omega$ да важи $\max S = m$
- Со други зборови: $P(M = m) =?$

Подобрен оценувач

- Ако примерокот има големина n и максимум m, тогаш $n-1$ елемент во него се избрани од множеството $\{1, 2, \ldots, m-1\}$ (ова важи бидејќи сите елементи на примерокот се различни меѓу себе, па сите се помали од максимумот m и уште важи $m \geqslant n$)
- Оттука, за $m \geqslant n$:

$$P(M = m) = \frac{\text{број на примероци со максимум} = m}{\text{вкупен број на примероци}} = \frac{\binom{m-1}{n-1}}{\binom{N}{n}},$$

  и, формално, $P(M = m) = 0$ за $m < n$
- Математичкото очекување на M може да се искористи за конструкција на подобар оценувач на N

Подобрен оценувач

Имаме:

$$E(M) = \sum_{m=n}^{N} m\, P(M = m) = \sum_{m=n}^{N} m \binom{m-1}{n-1} \binom{N}{n}^{-1} = \frac{n(N+1)}{n+1},$$

од каде следи:

$$N = E(M)\left(1 + \frac{1}{n}\right) - 1$$

Со замена $E(M) = m$ (емпириската вредност за M) се добива израз за оценувачот:

$$\widehat{N} = m\left(1 + \frac{1}{n}\right) - 1$$

Дали овој оценувач е добар?

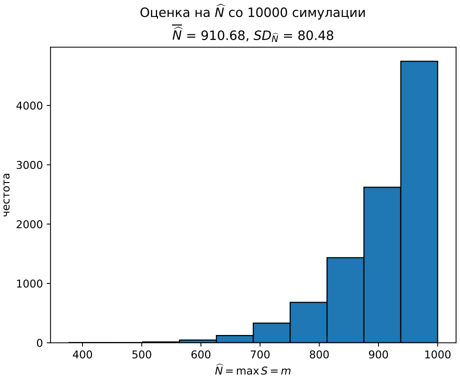
- Во пракса се покажало дека овој оценувач е многу ефикасен
- Во литературата се сретнуваат следниве податоци за тенковите:

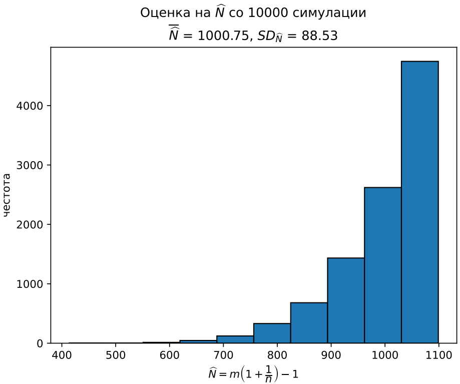
Месец	**Статистичка оценка**	**Вистински број**	**Број според разузнување**
јуни 1940	169	122	1000
јуни 1941	224	271	1550
август 1942	327	342	1550

Споредба на двата оценувачи

- За да го испитаме однесувањето на оценувачите, ќе спроведеме симулација врз основа на популација со познати параметри
- За симулацијата избираме $\Omega = \{1, 2, \ldots, 1000\}$, т.е. $N = 1000$
- Големината на примероците е $n = 10$
- Симулацијата се состои од избирање примерок, утврдување на вредноста m и оценка на N со двата оценувачи Резултатите ги прикажуваме на хистограми и ги пресметуваме просечните оценки и нивните стандардни девијации добиени при симулирањето

Резултатите се следни:





Се забележува дека во просек во просек вториот оценувач е подобар (за него важи $E(\widehat{N}) = N$, значи е непристрасен)

Еден пример на примена

- Добиениот резултат може да го искористиме за да оцениме колку банкноти од 10 денари има во оптек
- Претпоставката е дека нумерацијата е последователна и дека почнува со AA000001
- За оваа оценка на располагање имаме примерок од 71 банкнота собрани во Скопје во периодот декември ’22 – март ’23
- При анализата сериските броеви ги претвораме во природни броеви

Еден пример на примена

- Најмалиот набљудуван сериски број е АБ094730, т.е. $1\,094\,730$
- Најголемиот набљудуван сериски број е БК626364, т.е.

$$m = 43\,626\,364$$

- Големината на примерокот е $n = 71$
- За оцenката на максимумот на распределбата, т.е. бројот на банкноти во оптек, имаме:

$$\widehat{N} = m\left(1+\frac{1}{n}\right) - 1 = 43\,626\,364\left(1+\frac{1}{71}\right) - 1 \approx 44\,240\,819$$

Пример на примена

- Следи $\widehat{N} \approx 44\,240\,819$ што одговара на банкнота со сериски број

БЛ240819

- Значи во оптек има приближно $44\,240\,819$ банкноти од 10 денари
- Заклучуваме дека приближно $442\,448\,190$ од денарите во оптек се во апоени од 10 денари, што е приближно $7\,190\,000$ евра