Egy kis kiegészítés a statisztika tételhez...

Általában az a nagy-nagy feladat, hogy a rengeteg adatsokaságból szeretnénk valamit megjósolni a jövőre vonatkozóan néhány jellemző adatból. Ezek a statisztikai jellemzők pl. a darabszám (n), az átlag ($\overline{x}$), a szórás ($D = \sqrt{\frac{(x_i - \overline{x})^2}{n}}$), stb.

A Markov-tulajdonság. Ez arról szól, hogy egy nem-negatív adatsokaságban szeretnénk megmondani, hogy hány olyan elem van, mely nagyobb egy bizonyos értéknél.
Legyen az adatsokaság x_1, x_2, x_3, $\dots x_n$, ahol $x_i \geq 0$. Azt állítjuk, hogy legfeljebb $\frac{n \cdot \overline{x}}{A}$ darab van köztük, melyek A.nál nagyobbak. (Tehát csupán a darabszám (n) és az átlag ($\overline{x}$) ismeretében szeretnénk az A-nál nagyobb elemek számára becslést adni.)

A **bizonyítás**hoz rendezzük növekvő sorrendbe a x_1, x_2, x_3, $\dots x_n$ értékeket. Ekkor igaz a következő:

$$n\overline{x} = x_1 + x_2 + x_3 + \cdots + x_n$$

Az elemek közül legyen k darab, mely A-nál nagyobb. Ezek helyett írjuk be a legkisebb értékét (A), a többi helyére pedig 0-t.

$$n\overline{x} = \underbrace{x_1 + x_2 + x_3 + \cdots}_{<A} + \underbrace{\overbrace{\cdots + x_n}^{k\,\text{darab}}}_{\geq A}$$

$$n\overline{x} \geq 0 + 0 + 0 + \cdots + \overbrace{A + A + \cdots + A}^{k\,\text{darab}}$$

$$n\overline{x} \geq k \cdot A$$
$$k \leq \frac{n\overline{x}}{A}$$

A Csebisev-tulajdonság. Ha még egy „kicsikét” többet ismerünk az adathalmazról, akkor jobb becslést is tudunk adni...
Legyen az adatssokaság x_1, x_2, x_3, $\dots x_n$, ahol $x_i \geq 0$. Most az állítjuk, hogy legfeljebb $\frac{n}{B^2}$ darab van köztük, melyek $\overline{x}$-tól való eltérésük legalább BD. (Tehát most ismerjük az adatsokaság szórását is, s azt fogjuk tudni megmondani, hogy a B-szeres szórási intervallumon kívül legfeljebb hány elem lehet.)

Bizonyítás. Ismét rendezzük sorrendbe az adatsokaságot, de most az átlagtól valoó eltérésük szerinti növekvő sorrendbe. Ekkor fennáll, hogy

$$n \cdot D^2 = (x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2$$

Legyen k azon elemek darabszáma, melyek $\overline{x}$-tól való eltérése legalább $B \cdot D$. Ekkor

$$n \cdot D^2 = \underbrace{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2 + \cdots}_{<B^2D^2} + \underbrace{\overbrace{\cdots (x_n - \overline{x})^2}^{k\,\text{darab}}}_{\geq B^2D^2}$$

$$n \cdot D^2 \geq 0 + 0 + 0 + \cdots + \overbrace{B^2D^2 + B^2D^2 + \cdots + B^2D^2}^{k\,\text{darab}}$$

$$nD^2 \geq k \cdot B^2D^2$$

$$k \leq \frac{n}{B^2}$$

Tehát legfeljebb k darab olyan elem van, mely eltérése az átlagtól nagyobb, mint a szórás B-szerese. A többi elemre fennáll, hogy $|x_i - \overline{x}| < BD$.

Jó munkát kívánok!

Peti bá´