6 Epäyhtälöitä

- Epäyhtälöt ovat yksi matemaatikon voimakkaimmista työvälineistä.
- Yhtälö $a = b$ kertoo sen, että kaksi ehkä näennäisesti erilaista asiaa ovat samoja.
- Epäyhtälö $a \leq b$ saattaa antaa keinon analysoida monimutkaista asiaa paljon helpommin ymmärrettävän asian avulla.
- Todennäköisyyslaskennassa epäyhtälöitä käytetään tyypillisesti teorian apuna (esim. raja-arvotulosten johtaminen), mutta niille löytyy myös konkreettisempia sovelluksia (esim. algoritmien ominaisuuksien analysointi).

Epäyhtälöiden käytöstä

Kiinnitä huomiota seuraaviin asioihin:

- Millä oletuksilla epäyhtälö on voimassa?
- Kumpi suuruusjärjestys asioiden välillä vallitsee?

Usein matemaattinen argumentointi perustuu siihen, että muotoa $a \leq b$ oleva tulos johdetaan pitkän epäyhtälöketjun

$$a \leq a_1 \leq \cdots \leq a_k \leq b$$

avulla. Jos tässä ketjussa yksikin $\leq$-merkki on kirjoitettu vahingossa väärin päin, niin koko argumentilta putoaa pohja pois.

6.1 Markovin ja Tšebyševin epäyhtälöt

Lause (Markovin epäyhtälö)

Olkoon $X \geq 0$ sm, jonka odotusarvo on olemassa. Tällöin

$$P(X \geq a) \leq \frac{EX}{a}, \qquad \forall a > 0$$

Todistus. Määritellään sm Y kaavalla

$$Y = a\,1_{[a,\infty)}(X) = \begin{cases} 0, & \text{kun } 0 \leq X < a, \\ a, & \text{kun } X \geq a. \end{cases}$$

Koska $Y \leq X$, on

$$EX \geq EY = 0 \cdot P(Y = 0) + a \cdot P(Y = a) = a\,P(X \geq a).$$

Tšebyševin epäyhtälö

Lause (Tšebyševin epäyhtälö)

Olkoon X sm, jonka odotusarvo on μ ja varianssi σ^2 on olemassa
Tällöin

$$P[|X-\mu| \geq t] \leq \frac{\sigma^2}{t^2} \qquad \forall t > 0.$$

Erityisesti, jos $\sigma^2 > 0$, niin

$$P[|X-\mu| \geq k\,\sigma] \leq \frac{1}{k^2} \qquad \forall k > 0.$$

Todistus. Sovelletaan Markovin epäyhtälöä sm:aan $Y = (X-\mu)^2$.

Heikko suurten lukujen laki

- Engl. *weak law of large numbers, WLLN*.
- Lauseessa esiintyvää satunnaismuuttujien suppenemisen lajia kutsutaan stokastiseksi suppenemiseksi.

Lause (Heikko suurten lukujen laki)

Olkoon $X_1, X_2, \ldots$ jono riippumattomia sm:ia, joilla on sama odotusarvo $\mu = EX_i$ ja varianssi $\sigma^2 = \operatorname{var} X_i < \infty$. Tällöin keskiarvojen

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

muodostama jono $(\bar{X}_n)$ konvergoi stokastisesti kohti arvoa μ, eli

$$P(|\bar{X}_n - \mu| \geq \epsilon) \xrightarrow[n \to \infty]{} 0, \qquad \text{kaikilla } \epsilon > 0.$$

Heikon suurten lukujen lain todistus

Olkoon $\epsilon > 0$ annettu. Koska

$$E\bar{X}_n = \mu, \qquad \operatorname{var}\bar{X}_n = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n},$$

niin Tšebyševin epäyhtälön nojalla

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \xrightarrow[n\to\infty]{} 0.$$

6.2 Konveksit funktiot ja Jensenin epäyhtälö

- Funktio on konveksi, jos sen kuvaaja jää jokaisen jänteensä alapuolelle.
- Funktio on konkaavi, jos sen kuvaaja on jokaisen jänteensä yläpuolella.
- Yleensä funktiot eivät ole konvekseja eikä konkaaveja.

Konveksisuuden tarkka määritelmä

- Seuraavassa määritelmässä sana väli tarkoittaa mitä tahansa suljettua, avointa tai puoliavointa äärellistä tai ääretöntä väliä.
- Jos I on väli ja $x, y \in I$, niin myös pisteitä x ja y yhdistävän janan pisteet $\lambda x + (1-\lambda)y$, $0 \leq \lambda \leq 1$ kuuluvat välille I.

Määritelmä (Konveksi funktio, konkaavi funktio)

Olkoon $I \subset \mathbb{R}$ väli. Funktio $g : I \to \mathbb{R}$ on konveksi, jos

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y),$$
$$\text{kaikilla } x, y \in I \text{ ja kaikilla } 0 \leq \lambda \leq 1.$$

Funktio $g : I \to \mathbb{R}$ on konkaavi, jos $-g$ on konveksi.

Esimerkkejä

- Funktiot x^2, e^x ja $|x|$ ovat konvekseja koko reaaliakselilla.
- Funktio $1/x$ on konveksi, kun $x > 0$.
- Funktio $\log x$ on konkaavi funktio, kun $x > 0$.
- Affiini funktio $a + bx$ on sekä konveksi että konkaavi koko reaaliakselilla.

Konveksin funktion jänteiden kulmakertoimet

Seuraavassa lauseessa esiintyy erotusosamääriä, jotka kannattaa tulkita geometrisesti g:n jänteiden kulmakertoimien avulla. Piirrä kuva.

Lause

Olkoon I avoin väli. Funktio $g : I \to \mathbb{R}$ on konveksi, jos ja vain jos kaikilla $a < b < c$, joille $a, b, c \in I$ pätee epäyhtälö

$$\frac{g(b) - g(a)}{b - a} \leq \frac{g(c) - g(b)}{c - b}. \tag{1}$$

Konveksisuuden tarkistaminen derivaatan avulla

Lause (Konveksisuustarkistimia)

Olkoon $I \subset \mathbb{R}$ avoin väli, ja $g : I \to \mathbb{R}$ funktio.

(a) Jos g on jatkuvasti derivoituva ja g' on kasvava (ja molemmat ominaisuudet ovat voimassa koko I:llä), niin g on konveksi.

(b) Jos g on kahdesti derivoituva I:llä, ja $g''(x) \geq 0$ kaikilla $x \in I$, niin g on konveksi.

Todistus. a-kohta käydään läpi taululla.
b-kohta: Jos $g'' > 0$ välillä I, niin g' on kasvava välillä I, joten a-kohdan nojalla g on konveksi.

Diskreetti versio Jensenin epäyhtälöstä

- Jos $g : I \to \mathbb{R}$ on konveksi, niin konveksin funktion määritelmästä saadaan helposti induktiolla johdettua epäyhtälö
$$g\Big(\sum_{i=1}^{n} p_i\, x_i\Big) \le \sum_{i=1}^{n} p_i\, g(x_i), \qquad (2)$$
joka pitää paikkansa, kun kukin $x_i \in I$ ja kukin $p_i \ge 0$ ja $\sum_i p_i = 1$.
- Jos X on diskreetti sm siten, että $P(X = x_i) = p_i$, niin tämä voidaan kirjoittaa
$$g(EX) \le Eg(X)$$
- Lineaarikombinaatiota $x = \sum_i p_i x_i$, jossa painot p_i ovat ei-negatiivisia ja summautuvat ykköseksi, kutsutaan pisteiden $x_1, \dots, x_n$ konveksiksi kombinaatioksi. Tietenkin myös konveksi kombinaatio $x \in I$.

Konveksin funktion kuvaaja on tangenttinsa yläpuolella

Lause

Jos $I \subset \mathbb{R}$ on avoin väli ja $g : I \to \mathbb{R}$ on konveksi funktio, ja $m \in I$, niin voidaan valita luku k siten, että

$$g(x) \geq g(m) + k(x - m), \qquad \forall x \in I.$$

Todistuksesta nähdään, että mikäli g on derivoituva pisteessä m, niin pitää valita $k = g'(m)$, jolloin epäyhtälön oikea puoli on on tangentin yhtälö pisteessä m.

Jensenin epäyhtälö

Lause (Jensenin epäyhtälö)

Olkoon $I \subset \mathbb{R}$ avoin väli, ja $g : I \to \mathbb{R}$ konveksi funktio. Jos X on sm, jonka arvot ovat I:ssä (tn:llä yksi), ja EX sekä $Eg(X)$ ovat olemassa, niin

$$g(EX) \leq Eg(X) \tag{3}$$

Todistus. Koska $\mu = EX \in I$, niin voidaan valita luku k siten, että

$$g(x) \geq g(\mu) + k(x - \mu), \qquad \forall x \in \mu.$$

Siis todennäköisyydellä yksi

$$g(X) \geq g(\mu) + k(X - \mu),$$

ja Jensenin epäyhtälö seuraa ottamalla odotusarvo puolittain.

Esimerkki Jensenin epäyhtälön soveltamisesta

- Funktio $g(x) = x^2$ on konveksi, joten Jensenin epäyhtälön mukaan
$$g(EX) = (EX)^2 \leq E(X^2) = Eg(X),$$
mikäli kyseiset odotusarvot ovat olemassa.
- Tästä seuraa arvio
$$\operatorname{var} X = E(X^2) - (EX)^2 \geq 0,$$
minkä tietenkin tiesimme ennestään varianssin ominaisuuksien perusteella.

6.3 Hölderin epäyhtälö

Lause

Olkoot $p, q > 1$ sellaisia lukuja, että

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Tällöin

$$E|XY| \leq \left[E|X|^p\right]^{\frac{1}{p}} \left[E|Y|^q\right]^{\frac{1}{q}}$$

6.4 Cauchyn-Schwarzin epäyhtälö, kovarianssi ja korrelaatio

- Cauchyn-Schwarzin epäyhtälö on Hölderin epäyhtälön se erikoistapaus, jossa $p = q = 2$:

$$E|XY| \leq \sqrt{EX^2}\sqrt{EY^2}.$$

- Tämä epäyhtälö voi joissakin tapauksissa olla voimassa muodossa $E|XY| \leq \infty$, mikä ei kerro mitään mielenkiintoista.
- Jos sekä $EX^2 < \infty$ että $EY^2 < \infty$, niin silloin yläraja on äärellinen, ja saadaan tulos

$$|E(XY)| \leq E|XY| \leq \sqrt{EX^2}\sqrt{EY^2}. \tag{4}$$

Cauchy-Schwarz ja kovarianssi

- Oletetaan nyt, että satunnaismuuttujilla X ja Y on äärelliset varianssit, $\sigma_X^2 = \operatorname{var} X$ ja $\sigma_Y^2 = \operatorname{var} Y$.
- Niiden **kovarianssi** $\operatorname{cov}(X, Y)$ määritellään odotusarvona

$$\operatorname{cov}(X, Y) = E[(X - EX)(Y - EY)].$$

- Tämä odotusarvo on äärellinen, sillä kun Cauchyn-Schwarzin epäyhtälöä sovelletaan satunnaismuuttujien $X - EX$ ja $Y - EY$ tuloon, niin saadaan epäyhtälö

$$|\operatorname{cov}(X, Y)| \le \sigma_X \, \sigma_Y .$$

Tässä ylärajana on muuttujien keskihajontojen tulo.

Korrelaatiokerroin

- Mikäli $\sigma_X > 0$ ja $\sigma_Y > 0$, niin X:n ja Y:n korrelaatiokerroin (jota merkitään usein $\operatorname{corr}(X, Y)$ tai ρ_{XY}) määritellään kaavalla

$$\operatorname{corr}(X, Y) = \rho_{XY} = \rho = \frac{\operatorname{cov}(X, Y)}{\sigma_X \, \sigma_Y}. \tag{5}$$

- Cauchyn-Scwarzin epäyhtälön nojalla

$$|\operatorname{corr}(X, Y)| \le 1.$$

- Jos $\operatorname{corr}(X, Y) > 0$, jolloin myös $\operatorname{cov}(X, Y) > 0$, niin sanotaan että X ja Y ovat *positiivisesti korreloituneita*.
- Jos $\operatorname{corr}(X, Y) < 0$, jolloin myös $\operatorname{cov}(X, Y) < 0$, niin sanotaan että X ja Y ovat *negatiivisesti korreloituneita*.
- Jos $\operatorname{corr}(X, Y) = 0$, jolloin myös $\operatorname{cov}(X, Y) = 0$, niin sanotaan, että X ja Y eivät korreloi (lineaarisesti) tai että ne ovat (lineaarisesti) korreloimattomia.