Johdatus suurten poikkeamien teoriaan, syksy 2015

Harri Nyrhinen
Helsingin yliopisto

Suurten poikkeamien teoriassa kiinnostuksen kohteena ovat harvinaiset tapahtumat. Vastaavat todennäköisyydet suppenevat tyypillisesti eksponentiaalista vauhtia kohti nollaa jonkin volyymisuureen kasvaessa. Näin ollen ne ovat pieniä. Harvinaisuudestaan huolimatta tarkasteltavat tapahtumat voivat olla kiinnostavia teoreettisesta näkökulmasta mutta myös soveltavasta näkökulmasta seurauustensa takia.

Teoria katsotaan alkaneeksi Harald Cramérin tutkimuksista 1930-luvulla. Voimakasta kehitystä on tapahtunut 1980-luvulta lähtien. Erityisesti teorian perustulokset on pystytty todistamaan yleisille malleille. Samoin sovelluksia on löytynyt runsaasti, esimerkiksi fysiikasta, tilastotieteestä, informaatioteoriasta, riskiteoriasta jne.

Kurssilla käydään läpi teorian kehittely suhteellisen tarkasti vektoriarvoisille stokastisille prosesseille. Sovelluksia tarkastellaan valikoivasti.

Päälähteet:

Dembo, A. and Zeitouni O. (1998). Large deviations techniques and applications (2. edition). Springer.

Rockafellar, T. (1970). Convex analysis. Princeton University Press.

1. Johdanto

Suurten poikkeamien teoriassa ollaan kiinnostuneita harvinaisten tapahtumien todennäköisyyksistä. Asiaa voidaan havainnollistaa keskeisen raja-arvolauseen avulla. Tarkastellaan klaavojen lukumäärää, kun tasapainoista rahaa heitetään n kertaa. Olkoon S_n klaavojen määrä. Keskeisen raja-arvolauseen nojalla

$$(1.1) \qquad \lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n/2}{\sqrt{n}/2} \le x \right) = \Phi(x)$$

tasaisesti joukossa $x \in \mathbb{R}$, missä Φ on standardi normaalijakauman kertymäfunktio. Suppenemisen tasaisuuden nojalla approksimaatiossa päästään pieneen absoluuttiseen virheeseen, kunhan vain n on riittävän suuri. On siis perusteltua approksimoida esimerkiksi

$$(1.2) \qquad \mathbb{P}(S_n \le a) = (1+o(1))\, \Phi\left(\frac{a-n/2}{\sqrt{n}/2}\right)$$

tai

$$(1.3) \qquad \mathbb{P}(S_n \le an) = (1+o(1))\, \Phi\left(\frac{(a-\frac{1}{2})n}{\sqrt{n}/2}\right),$$

missä $a > 0$ on kiinteä. Tämän tyyppisiä todennäköisyyksiä esiintyy usein sovelluksissa.

Kaavan (1.3) mukainen todennäköisyys on luonteeltaan suurten poikkeamien teorian piiriin kuuluva. Oletetaan, että $a \in (0, \frac{1}{2})$. Tällöin oikealla puolella oleva argumentti

$$\frac{(a-\frac{1}{2})n}{\sqrt{n}/2}$$

pienenee kohti miinus ääretöntä.

Approksimaatio on siis lähellä nollaa, kun n on suuri. Liikuttaessa jakauman häntäalueilla on absoluuttisen virheen sijaan suhteellinen virhe usein tärkeämpi. Absoluuttinen virhe (1.3):ssa on

$$(1.4)\quad \left| \mathbb{P}(S_n \leq an) - \phi\left(\frac{(a-\frac{1}{2})n}{\sqrt{n}/2}\right) \right|$$

ja suhteellinen

$$(1.5)\quad \frac{\left| \mathbb{P}(S_n \leq an) - \phi\left(\frac{(a-\frac{1}{2})n}{\sqrt{n}/2}\right) \right|}{\mathbb{P}(S_n \leq an)}.$$

Keskeinen raja-arvo antaa tietoa vain absoluuttisesta virheestä. Kuten myöhemmin tullaan osoittamaan, pätee

$$(1.6)\quad \mathbb{P}(S_n \leq an) = e^{-(1+o(1))n(a\log a + (1-a)\log(1-a) + \log 2)}$$

ja toisaalta

$$(1.7)\quad \phi\left(\frac{(a-\frac{1}{2})n}{\sqrt{n}/2}\right) = e^{-(1+o(1))n \cdot \frac{1}{2}(1-2a)^2}.$$

Suhteelliselta virheeltä (1.5) ei siis voida odottaa pienuutta. Eksponentit kaavoissa (1.6) ja (1.7) voidaan määrätä suurten poikkeamien teorian avulla.

Päähuomio suurten poikkeamien teoriassa on juuri eksponenteissa ($a\log a + (1-a)\log(1-a) + \log 2$ ja $\frac{1}{2}(1-2a)^2$ edellä). Useinkaan ei olla kiinnostuneita todennäköisyyksien tarkasta muodosta, vaan ainoastaan suuruusluokasta.

Todetaan vielä lopuksi, että suurten poikkeamien teorian skaalaus on sama kuin suurten lukujen laissa. Edellä esitetyssä esimerkissähän pätee

$$\frac{S_n}{n} \to \frac{1}{2}, \quad \text{kun } n \to \infty,$$

m.v. Suurten poikkeamien teoriassa tarkastellaan todennäköisyyksiä

$$\mathbb{P}\left(\frac{S_n}{n} \le a \right), \quad a > 0.$$

Suurten lukujen laki sisältyy tarkasteluihin. Esimerkiksi

$$\mathbb{P}\left(\frac{S_n}{n} \le \frac{1}{2} - \varepsilon \right) \to 0, \quad \mathbb{P}\left(\frac{S_n}{n} \le \frac{1}{2} + \varepsilon \right) \to 1,$$

kaikilla $\varepsilon > 0$. Suurten lukujen lakia kutsutaan teoriassa 'tyypilliseksi käyttäytymiseksi'.

2. Summien poikkeamien periaate

Johdannossa esitetyt tyyppiä $\mathbb{P}(S_n/n \le a)$ (tai $\mathbb{P}(S_n/n \ge b)$) olevat todennäköisyydet ovat sovelluksissa usein kiinnostuksen kohteena. Teorian kehittelyn näkökulmasta on kuitenkin hyödyllistä tarkastella asiaa laajemmin.

Palautetaan mieleen eräitä analyysin käsitteitä ja tuloksia. Olkoon a_n, $n = 1, 2, \ldots$, jono laajennetun reaalilukujoukon $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ kuuluvia alkioita. Jonon yläraja-arvo (limes superior, limes supremum, lim sup), $\limsup a_n$, määritellään ehdosta

$$(2.1)\qquad \limsup a_n = \lim_{n\to\infty} \left(\sup_{m \ge n} a_m \right)$$

ja alaraja-arvo (limes inferior, limes infimum, lim inf) ehdosta

$$(2.2)\qquad \liminf a_n = \lim_{n\to\infty} \left(\inf_{m \ge n} a_m \right).$$

Käsitteet ovat määriteltyjä $\forall\, (a_n) \subseteq \overline{\mathbb{R}}$. Muita hyödyllisiä ominaisuuksia ovat esimerkiksi

$$(2.3)\qquad \limsup a_n = -\liminf(-a_n),$$

$$(2.4)\qquad \liminf a_n \le \limsup a_n$$

ja yhtäsuuruus pätee $\Leftrightarrow$ $\lim a_n$ on olemassa. Tällöin $\lim a_n$ yhtyy ylä- ja alaraja-arvoon,

$$(2.5)\qquad \limsup (a_n + b_n) \le \limsup a_n + \limsup b_n,$$

$\forall\, (a_n), (b_n) \subseteq \overline{\mathbb{R}}$, edellyttäen että oikea puoli on määritelty.

Lisäksi $\limsup a_n$ on 'suurin' raja-arvo, joka saadaan (a_n):n osajonoista ts. $\exists$ (a_n):n osajono (a_{n_k}) siten, että (a_{n_k}) suppenee ja

(2.6) $$\lim_{k\to\infty} a_{n_k} = \limsup a_n,$$

ja toisaalta, jos (a_{n_k}) on (a_n):n suppeneva osajono, niin

(2.7) $$\lim_{k\to\infty} a_{n_k} \le \limsup a_n .$$

Analoginen tulos pätee alaraja-arvolle.

Esitetyt käsitteet ja tulokset yleistyvät luonnollisella tavalla muunkin tyyppisiin rajankäynteihin. Jos esimerkiksi $f: \mathbb{R}^d \to \overline{\mathbb{R}}$ on mielivaltainen funktio ja $x \in \mathbb{R}^d$, niin määritellään

(2.8) $$\liminf_{y\to x} f(y) = \lim_{\delta\to 0+}\Big(\inf_{0<|x-y|<\delta} f(y)\Big)$$

jne.

Sanotaan, että funktio $f: \mathbb{R}^d \to \overline{\mathbb{R}}$ on <u>alhaalta puolijatkuva</u> (lyhennys apj), jos

(2.9) $$f(x) \le \liminf_{y\to x} f(y) \quad , \quad \forall x \in \mathbb{R}^d,$$

ja <u>ylhäältä puolijatkuva</u> (ypj), jos

(2.10) $$f(x) \ge \limsup_{y\to x} f(y) \quad , \quad \forall x \in \mathbb{R}^d.$$

Selvästi f on jatkuva s.j.v.s.k. f on alhaalta ja ylhäältä puolijatkuva. Lisäksi f on apj. s.j.v.s.k. <u>tasojoukot</u>

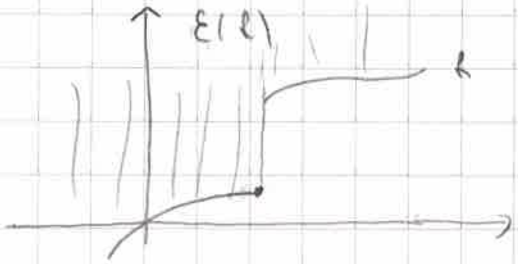
(2.11) $$\Psi_f(\alpha) \doteq \{x \in \mathbb{R}^d \mid f(x) \le \alpha\}$$

ovat suljettuja, $\forall \alpha \in \mathbb{R}$.

Funktion $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ epigraafi $\mathcal{E}(f)$ on joukko

$$\mathcal{E}(f) = \{(x, M) \mid x \in \mathbb{R}^d, M \in \mathbb{R}, M \geq f(x)\} \subseteq \mathbb{R}^{d+1}.$$

Tällöin f on apj sjvsk. $\mathcal{E}(f)$ on suljettu $\mathbb{R}^{d+1}$:n osajoukko.



Perusteluja edellä esitetyille väitteille esitetään harjoituksissa.

Käytetään jatkossa myös merkintää

$$D(f) = \{x \in \mathbb{R}^d \mid f(x) < \infty\}.$$

Useimmiten $-\infty$ ei esiinny f:n arvona tarkasteluissa. Tällöin $D(f)$ on siis f:n äärellisyysalue.

<u>Määritelmä 2.1.</u> Funktio $I : \mathbb{R}^d \to [0, \infty]$ on <u>vauhtifunktio</u>, jos I on a.p.j. Lisäksi I on <u>hyvä</u> vauhtifunktio, jos tasojoukot $\Psi_I(\alpha)$ ovat kompakteja, $\forall \alpha \in \mathbb{R}$.

Vauhtifunktiolta vaaditaan siis alhaalta puolijatkuvuuden lisäksi ei-negatiivisuus. Arvo $+\infty$ on sallittu. Hyvältä vauhtifunktiolta vaaditaan, että tasojoukot ovat paitsi suljettuja myös rajoitettuja.

Vauhtifunktion tehtävänä on liittää harvinaisiin tapahtumiin häviämisvauhteja. Esimerkiksi kaavassa (1.6) esiintyy vauhtifunktio

$$I(x) = \begin{cases} x \log x + (1-x)\log(1-x) + \log 2, & x \in (0,1) \\ \infty, & x < 0 \text{ tai } x > 1. \end{cases}$$

Luonnollisesti on oltava $I \geq 0$, koska (1.6):n vasen puoli on rajoitettu. Esimerkissä I saa arvon nolla pisteessä $x = 1/2$. Tämä vastaa tyypillistä käyttäytymistä.

Suurten poikkeamien periaate on lauseena jakaumaperheestä. Olkoon yleisesti

$$(\Omega_n, \mathcal{F}_n, \mathbb{P}_n), \quad n = 1, 2, \ldots$$

jono todennäköisyyskenttiä ja $\{X_n\}$ jono vastaavissa kentissä määriteltyjä $\mathbb{R}^d$-arvoisia satunnaisvektoreita ts.

$$X_n : (\Omega_n, \mathcal{F}_n) \longrightarrow (\mathbb{R}^d, \mathcal{B}^d)$$

on mitallinen kuvaus, $n = 1, 2, \ldots$ ($\mathcal{B}^d$ = $\mathbb{R}^d$:in Borel-joukot). Merkitään P_n:llä X_n:n jakaumaa ts.

$$(2.12) \quad P_n(B) = \mathbb{P}_n(\omega \in \Omega_n \mid X_n(\omega) \in B), \quad \forall B \in \mathcal{B}^d.$$

Tällöin P_n on todennäköisyysmitta $(\mathbb{R}^d, \mathcal{B}^d)$:llä.

Johdannon esimerkissä on sopivaa valita $(d=1)$

$$(\Omega_n, \mathcal{F}_n, \mathbb{P}_n) = (\Omega, \mathcal{F}, \mathbb{P}),$$

missä $(\Omega, \mathcal{F}, \mathbb{P})$ on tunnetulla tavalla konstruoitu tulokenttä ja

$$X_n = S_n / n,$$

$n = 1, 2, \ldots$ Tällöin tarkastelun kohteena olevat jakaumat P_n ovat S_n/n:n jakaumia.

Vaikka seuraavassa esitettävä määritelmä koskee vain jakaumaperhettä $\{P_n\}$, liitetään näihin aina vastaava satunnaisvektoriperhe $\{X_n\}$ ja tarvittaessa täsmennetään vastaavat todennäköisyyskentät $\{(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)\}$.

<u>Määritelmä 2.2.</u> Olkoon I vauhtifunktio ja $\{P_n\}$ jakaumajono. Sanotaan, että $\{P_n\}$ toteuttaa <u>heikon suurten poikkeamien periaatteen vauhtifunktiolla I</u>, jos

$$(2.13) \quad \liminf n^{-1} \log P_n(G) \geq -\inf_{x \in G} I(x)$$

kaikilla avoimilla joukoilla $G \subseteq \mathbb{R}^d$ ja

$$(2.14) \quad \limsup n^{-1} \log P_n(F) \leq -\inf_{x \in F} I(x)$$

kaikilla kompakteilla joukoilla $F \subseteq \mathbb{R}^d$.

Jos lisäksi (2.14) pätee kaikilla suljetuilla joukoilla $F \subseteq \mathbb{R}^d$, sanotaan, että $\{P_n\}$ toteuttaa <u>suurten poikkeamien periaatteen vauhtifunktiolla I</u>.

Käytetään jatkossa lyhennyksiä HSPP heikosta suurten poikkeamien periaatteesta ja SPP suurten poikkeamien periaatteesta. Otetaan käyttöön myös lyhennysmerkintä

$$I(B) = \inf_{x \in B} I(x), \quad B \in \mathcal{B}^d.$$

Yleisempi versio käsitteestä saataisiin tarkastelemalla todennäköisyyksien $P_n(\cdot)$ muunnoksia

$$a_n^{-1} \log P_n(\cdot),$$

missä $a_n > 0$, $a_n \to \infty$. Teoriaa voitaisiin kehittää tästä lähtökohdasta lähes muuttumattomana. Oletetaan kuitenkin aina, että $a_n = n$, $\forall n$

Määritelmän 2.2 terminologiaa käytetään myös satunnaisvektorijonoista. Jos X_n:in jakauma on P_n, $n=1,2,\ldots$, sanotaan, että $\{X_n\}$ toteuttaa SPP:in (tai HSPP:in), mikäli $\{P_n\}$ toteuttaa SPP:in (HSPP:in).

Usein I:llä on parempiakin jatkuvuusominaisuuksia kuin mitä määritelmä edellyttää. Tarkastellaan määritelmän havainnollistamiseksi tilannetta, jossa I on jatkuva. Olkoon $A \subseteq \mathbb{R}^d$ mielivaltainen suorakulmio, esim. $A = (a_1, b_1) \times \dots \times (a_d, b_d)$. Tällöin

$$\inf_{x \in \mathring{A}} I(x) = \inf_{x \in \bar{A}} I(x),$$

missä $\mathring{A}$ on A:n sisäosa ja $\bar{A}$ sulkeuma. Suurten poikkeamien periaatteen vallitessa on siis

$$\lim_{n \to \infty} n^{-1} \log P_n(A) = - \inf_{x \in A} I(x) = -I(A).$$

Kirjoitettuna ε-muodossa : $\forall \varepsilon > 0, \exists n_\varepsilon$:

$$(2.15) \quad e^{-n(I(A)+\varepsilon)} \leq P_n(A) \leq e^{-n(I(A)-\varepsilon)}$$

aina kun $n > n_\varepsilon$. Todetaan, että approksimaatio

$$P_n(A) \approx e^{-nI(A)}$$

ei ole välttämättä tarkka, koska esimerkiksi $e^{n\varepsilon}$ on suuri niin ollessa suuri. Jos olisi todellisuudessa vaikkapa

$$P_n(A) = n^k e^{-nI(A)},$$

missä $k > 0$ on kiinteä, niin (2.15) pätisi. . .

Tarkastellaan eräitä yleisiä suurten poikkeamien periaatteeseen liittyviä tuloksia.

<u>Lause 2.1.</u> Olkoon jakaumaperhe $\{P_n\}$ mielivaltainen ja

(2.15.1) $$I(x) = -\lim_{\delta \to 0+} \limsup n^{-1} \log P_n(B(x,\delta)), \forall x \in \mathbb{R}^d,$$

missä $B(x,\delta) = \{y \in \mathbb{R}^d : |x-y| < \delta\}$ on avoin pallo. Tällöin I on vauhtifunktio ja (2.14) pätee kaikilla kompakteilla joukoilla F. Lisäksi

(2.15.2) $$\limsup n^{-1} \log P_n(G) \geq -\inf_{x \in G} I(x)$$

kaikilla avoimilla joukoilla G.

<u>Todistus.</u> Selvästi $I(x) \geq 0$, $\forall x$. Olkoon $\alpha \in \mathbb{R}$ ja $I(x) > \alpha$. Tällöin voidaan määrätä δ_0 siten, että

$$-\limsup n^{-1} \log P_n(B(x,\delta_0)) > \alpha.$$

Jos $y \in B(x,\delta_0)$, voidaan määrätä δ_y : $B(y,\delta_y) \subseteq B(x,\delta_0)$, jolloin

$$\limsup n^{-1} \log P_n(B(y,\delta_y)) < -\alpha$$

ja

$$\limsup n^{-1} \log P_n(B(y,\delta)) < -\alpha, \quad \forall \delta \leq \delta_y.$$

Nähdään, että

$$I(y) = -\lim_{\delta \to 0+} \limsup n^{-1} \log P_n(B(y,\delta)) > \alpha.$$

Siis $I(y) > \alpha$, $\forall y \in B(x,\delta_0)$, joten joukko

$\Psi_I(\alpha) = \{x \mid I(x) > \alpha\}$ on avoin ja I on vauhtifunktio.

Todistetaan seuraavaksi (2.14). Olkoon F kompakti. Oletetaan ensin, että $I(F) < \infty$. Olkoon $\varepsilon > 0$. Mielivaltaiselle $x \in F$ määrätään $\delta_x > 0$ siten, että

$$-\limsup n^{-1} \log P_n(B(x, \delta_x)) \geq I(F) - \varepsilon.$$

Heine-Borelin lauseen nojalla voidaan määrätä pisteet $x_1, \dots, x_N \in F$, joille

$$F \subseteq \bigcup_{i=1}^{N} B(x_i, \delta_{x_i}).$$

Seuraavan lemman 2.1 nojalla

$$\limsup n^{-1} \log P_n(F) \leq \max_{i=1}^{N} \limsup n^{-1} \log P_n(x_i, \delta_{x_i})$$

$$\leq -I(F) + \varepsilon,$$

josta (2.14) seuraa. Tapaus $I(F) = \infty$ käsitellään analogisesti.

Olkoon nyt G avoin. Jos $x \in G$, niin $B(x, \delta) \subseteq G$ eräälle $\delta > 0$ ja siis

$$\limsup n^{-1} \log P_n(G) \geq \limsup n^{-1} \log P_n(B(x, \delta)),$$

joten myös

$$\limsup n^{-1} \log P_n(G)$$

$$\geq \lim_{\delta \to 0+} \limsup n^{-1} \log P_n(B(x, \delta)) = -I(x).$$

Alaraja (2.15.2) seuraa ottamalla puolittain supremum yli joukon $x \in G$. $\square$

<u>Lemma 2.1.</u> Olkoot $a_{1n}, \ldots, a_{Nn}$ ei-negatiivisia reaalilukuja, $n = 1, 2, \ldots$, missä $N \in \mathbb{N}$ on kiinteä. Silloin

$$\limsup_{n\to\infty} n^{-1} \log (a_{1n} + \cdots + a_{Nn}) = \max_{i=1}^{N} \limsup_{n\to\infty} n^{-1} \log a_{in}.$$

(sopimus: $\log 0 = -\infty$).

<u>Todistus.</u> Olkoon $\bar{a} = \max_{i=1}^{N} \limsup n^{-1} \log a_{in}$.

Todistetaan väite vain tapauksessa $\bar{a} \in (-\infty, \infty)$. Olkoon $\varepsilon > 0$. Määrätään n_ε siten, että

$$n^{-1} \log a_{in} \le \bar{a} + \varepsilon, \quad \forall n \ge n_\varepsilon, \; i = 1, \ldots, N.$$

Tällöin $a_{1n} + \cdots + a_{Nn} \le N e^{n(\bar{a}+\varepsilon)}$, $\forall n \ge n_\varepsilon$, ja

$$n^{-1} \log (a_{1n} + \cdots + a_{Nn}) \le n^{-1} \log N + \bar{a} + \varepsilon.$$

Nähdään, että

$$\limsup_{n\to\infty} n^{-1} \log (a_{1n} + \cdots + a_{Nn}) \le \bar{a} + \varepsilon,$$

joten

$$\limsup_{n\to\infty} n^{-1} \log (a_{1n} + \cdots + a_{Nn}) \le \bar{a}.$$

Toisaalta

$$n^{-1} \log (a_{1n} + \cdots + a_{Nn}) \ge n^{-1} \log a_{in}, \quad i = 1, \ldots, N,$$

joten

$$\limsup_{n\to\infty} n^{-1} \log (a_{1n} + \cdots + a_{Nn}) \ge \limsup_{n\to\infty} n^{-1} \log a_{in}, \; \forall i.$$

Siis

$$\limsup_{n\to\infty} n^{-1} \log (a_{1n} + \cdots + a_{Nn}) \ge \bar{a}. \quad \square$$

Lause 2.1 viittaa siihen, että suurten poikkeamien periaate toteutuu laajalla luokalla jakaumaperheitä. Ylärajan laajentuminen sopiviin alajoukkoihin automaattisesti suljetuille joukoille.

<u>Määritelmä 2.3.</u> <u>Jakaumaperhe</u> $\{P_n\}$ on <u>eksponentiaalisesti tiukka</u>, jos $\forall \alpha > 0$ on olemassa kompakti joukko F_α, jolle

(2.16) $\limsup n^{-1} \log P_n(F_\alpha^c) \le -\alpha$.

Jos Z_n:n jakauma on P_n, $\forall n$, niin $\{Z_n\}$ on eksp. tiukka, jos $\{P_n\}$ on sitä.

<u>Lause 2.2.</u> Olkoon $\{P_n\}$ eksponentiaalisesti tiukka ja I sellainen vauhtifunktio, että (2.14) pätee kaikille kompakteille joukoille F. Silloin (2.14) pätee kaikille suljetuille joukoille F.

<u>Todistus.</u> Olkoon F suljettu ja $\alpha > 0$. Olkoon F_α sellainen kompakti joukko, että (2.16) pätee. Kirjoitetaan

$$P_n(F) = P_n(F \cap F_\alpha) + P_n(F \cap F_\alpha^c)$$

(A^c tarkoittaa A:n komplementtia ($\mathbb{R}^d$:ssä)). Oletusten ja lemman 2.1 nojalla

$$\begin{aligned} \limsup n^{-1} \log P_n(F) &= \max\{\limsup n^{-1} \log P_n(F \cap F_\alpha), \\ &\qquad \limsup n^{-1} \log P_n(F \cap F_\alpha^c)\} \\ &= \max\{-\inf_{x \in F \cap F_\alpha} I(x), -\alpha\} \\ &\le \max\{-\inf_{x \in F} I(x), -\alpha\}. \end{aligned}$$

Lauseen väite saadaan antamalla $\alpha \to \infty$. $\square$

Esimerkki 2.1. Olkoon

$$\xi_n = (\xi_{n1}, \dots, \xi_{nd}), \quad n = 1, 2, \dots$$

jono riippumattomia ja samoin jakautuneita satunnais-vektoreita ja

$$Y_n = \xi_1 + \dots + \xi_n, \quad \bar{X}_n = \frac{Y_n}{n}.$$

Tällöin $\{Y_n\}$ on satunnaiskulku $\mathbb{R}^d$:ssä. Olkoon P_n vektorin $\bar{X}_n$ jakauma.

Oletetaan, että

$$\mathbb{P}(|\xi_{1i}| \leq a) = 1, \quad i = 1, \dots, d,$$

eräälle $a \in \mathbb{R}$. Tällöin $\{P_n\}$ on triviaalisti eksponentiaalisesti tiukka, joten (2.14) pätee kaikille suljetuille joukoille. Myöhemmin nähdään, että myös (2.13) pätee kaikille avoimille joukoille G, joten $\{P_n\}$ toteuttaa SPP:n. Vauhtifunktio pystytään myös identifioimaan (ns. Cramérin lause, joka esitetään aikanaan).

Tarkastellaan nyt vauhtifunktioiden yksikäsitteisyyttä. 2.12.

Lause 2.3. Oletetaan, että $\{P_n\}$ toteuttaa heikon suurten poikkeamien periaatteen vauhtifunktioilla I_1 ja I_2. Silloin $I_1 = I_2$.

Huomautus. Vauhtifunktioita voi siis olla korkeintaan yksi.

Todistus. Oletetaan, että olisi $I_1(x_0) > I_2(x_0)$ jollekin $x_0 \in \mathbb{R}^d$. Koska I_1 on aps., on

$$\lim_{\delta \to 0+} \left(\inf_{|x-x_0| \le \delta} I_1(x) \right) = I_1(x_0).$$

Toisaalta oletuksen nojalla mielivaltaiselle $\delta > 0$,

$$-\inf_{|x-x_0| \le \delta} I_1(x) \ge \limsup_{n\to\infty} n^{-1} \log P_n(x \,|\, |x-x_0| \le \delta)$$

$$\ge \liminf_{n\to\infty} n^{-1} \log P_n(x \,|\, |x-x_0| < \delta)$$

$$\ge -\inf_{|x-x_0| < \delta} I_2(x) \ge -I_2(x_0).$$

Koska $\delta > 0$ on mielivaltainen, saadaan

$$I_2(x_0) \ge \lim_{\delta \to 0+} \left(\inf_{|x-x_0| \le \delta} I_1(x) \right) = I_1(x_0) \text{ ↯}. \quad \square$$

Yksikäsitteisyys perustuu pelkästään vaatimukseen vauhtifunktion alhaalta puolijatkuvuudesta. Seuraavasta nähdään, että vaatimus ei ole oleellinen rajoitus.

<u>Määritelmä 2.4.</u> Olkoon $f: \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ mielivaltainen funktio. Tällöin f:n <u>alaverho</u> $\underline{f}$ on suurin alhaalta puolijatkuva funktio, jota f majoroi. Toisin sanoen

$$(2.12) \qquad \underline{f}(x) = \sup\{g(x) \mid g \text{ apj}, f(y) \geq g(y), \forall y \in \mathbb{R}^d\}.$$

Helposti nähdään, että kaava (2.12) määrittelee apj-funktion ja että sillä on vaadittu maksimaalisuusominaisuus.

<u>Lause 2.4.</u> Olkoon I mielivaltainen funktio, jolle (2.13) ja (2.14) pätee kaikille avoimille joukoille G ja suljetuille joukoille F. Silloin $\{P_n\}$ toteuttaa SPP:n vauhtifunktiolla $\underline{I}$.

<u>Todistus.</u> Ensinnäkin $\underline{I}$ on apj ja jos F on suljettu, niin

$$\limsup n^{-1} \log P_n(F) \leq -\inf_{x \in F} I(x) \leq -\inf_{x \in F} \underline{I}(x)$$

(koska $I(x) \geq \underline{I}(x), \forall x$). Olkoon nyt G avoin.

Määritellään funktio J ehdosta

$$J(x) = \begin{cases} \inf_{y \in G} I(y), & \text{kun } x \in G \\ 0, & \text{muulloin.} \end{cases}$$

Selvästi J on apj, oletuksesta (2.13) seuraa, että $I(x) \geq 0, \forall x$. Siispä $I(x) \geq J(x), \forall x$, joten myös

$\underline{I}(x) \geq J(x), \forall x$. Oletusten nojalla

$$\liminf n^{-1} \log P_n(G) \geq -\inf_{x \in G} I(x)$$

$$= -\inf_{x \in G} J(x) \geq -\inf_{x \in G} \underline{I}(x). \quad \square$$

Esitetään vielä suurten poikkeamien periaatteen yhteys suurten lukujen lakiin.

<u>Määritelmä 2.5.</u> Satunnaisvektorijonolle $\{X_n\}$ pätee <u>eksponentiaalinen suurten lukujen laki</u>, jos on olemassa sellainen $x_0 \in \mathbb{R}^d$, että

$\forall \varepsilon > 0$, $\exists k_\varepsilon > 0$ ja n_ε :

$$(2.18) \qquad \mathbb{P}_n(|X_n - x_0| \geq \varepsilon) \leq e^{-k_\varepsilon n},$$

$\forall n > n_\varepsilon$. Merkitään: $X_n \xrightarrow{exp} x_0$, kun $n \to \infty$.

<u>Huomautus.</u> Heikossa suurten lukujen laissa vaaditaan, että

$$\lim_{n \to \infty} \mathbb{P}_n(|X_n - x_0| \geq \varepsilon) = 0,$$

$\forall \varepsilon > 0$. Eksponentiaalisesta suurten lukujen laista siis seuraa heikko suurten lukujen laki. Siitä seuraa myös vahva suurten lukujen laki edellyttäen, että X_n:t on määritelty samassa todennäköisyyskentässä ts. $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n) = (\Omega, \mathcal{F}, \mathbb{P})$, $\forall n$.
Tällöin pätee

$$(2.19) \qquad \mathbb{P}(\lim_{n \to \infty} X_n = x_0) = 1.$$

Tämä seuraa Borel-Cantellin lemmasta, sillä (2.18):n nojalla

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - x_0| \geq \varepsilon) < \infty,$$

joten $\forall \varepsilon > 0$,

$$\mathbb{P}(\limsup \{\omega \mid |X_n - x_0| \geq \varepsilon\}) = 0.$$

Tästä seuraa tunnetusti (2.19).

Lause 2.5. Oletetaan, että $\{P_n\}$ toteuttaa suurten poikkeamien periaatteen hyvällä vauhtifunktiolla I. Oletetaan, että $I(x_0) = 0$ ja että $I(x) > 0$, $\forall x \neq x_0$. Silloin $\bar{X}_n \xrightarrow{4.5} x_0$.

Todistus. Olkoon $\varepsilon > 0$. Silloin

$$\limsup n^{-1} \log P_n(|\bar{X}_n - x_0| \geq \varepsilon) \leq -\inf_{|x - x_0| \geq \varepsilon} I(x).$$

Koska I on hyvä, pätee

$$I(x) \to \infty, \quad \text{kun } |x| \to \infty.$$

Olkoon nimittäin $M > 0$. Tällöin

$$\{x \mid I(x) \leq M\}$$

on kompakti, joten se sisältyy erääseen palloon $B(0, R)$. Nähdään, että $I(x) > M$, kun $|x| > R$.

Edellä esitetyn nojalla voidaan määrätä $R > 0$:

$$\inf_{|x - x_0| \geq \varepsilon} I(x) = \inf_{R \geq |x - x_0| \geq \varepsilon} I(x).$$

Alhaalta puolijatkuva funktio saavuttaa miniminsä kompaktissa joukossa, joten on olemassa y i

$$\begin{cases} R \geq |y - x_0| \geq \varepsilon, \\ \inf_{|x - x_0| \geq \varepsilon} I(x) = I(y) > 0 \quad \text{(oletuksen nojalla).} \end{cases}$$

Siispä

$$P_n(|\bar{X}_n - x_0| \geq \varepsilon) \leq e^{-n I(y)/2}, \quad \text{kun } n \text{ suuri.} \ \square$$

3. Gärtner–Ellisin lause

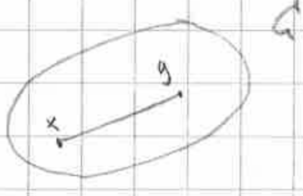
Lauseessa 2.1 johdettiin teoreettiskvantorinen esitys vauhtifunktiolle. Lauseen raja-arvojen määrääminen suoraan on hankalaa, joten on tarvetta kehittää riittäviä ehtoja suurten poikkeamien periaatteelle ja etsiä yksinkertaisempia esityksiä vauhtifunktiolle. Näitä kysymyksiä tarkastellaan tässä luvussa.

Gärtner–Ellisin lause on keskeinen osa konveksia suurten poikkeamien teoriaa (vauhtifunktio tulee olemaan konveksi). Ennen lauseen esittämistä tarkastellaan hieman konveksien funktioiden teoriaa.

Joukko $G \subseteq \mathbb{R}^d$ on <u>konveksi</u>, jos $\forall x, y \in G$, $\lambda \in (0,1)$,

$$(1-\lambda)x + \lambda y \in G.$$

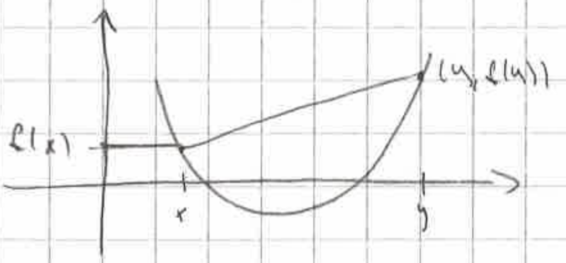
Ts. G on konveksi, jos se sisältää kaikkien pisteidensä väliset yhdysjanat. Erityisesti $\emptyset$, $\mathbb{R}^d$ ja kaikki singleton-joukot ovat konvekseja.



Funktio $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ on <u>konveksi</u>, jos

$$f((1-\lambda)x + \lambda y) \le (1-\lambda) f(x) + \lambda f(y)$$

kaikilla $x, y \in \mathbb{R}^d$, $\lambda \in (0,1)$.



f:n arvot ovat janan alapuolella

Suoraviivaisesti nähdään, että f on konveksi jos ja vain jos epigraafi $\mathcal{E}(f)$ on $\mathbb{R}^{d+1}$:in konveksi osajoukko, missä siis

$$\mathcal{E}(f) = \{(x, M) \mid M \ge f(x)\}.$$

<u>Lemma 3.1.</u> Olkoon (f_n) jono konvekseja funktioita $\mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ ja

$$f(x) = \limsup f_n(x), \quad \forall x \in \mathbb{R}^d.$$

a) jos $f(x) > -\infty$, $\forall x \in \mathbb{R}^d$, niin f on konveksi,

b) jos $f(x_0) \in \mathbb{R}$ ja $f(x) < \infty$, $\forall x \in B(x_0, \delta)$, jollain $\delta > 0$, niin $f(x) > -\infty$, $\forall x \in \mathbb{R}^d$ (ja siis f on konveksi).

<u>Todistus.</u> Selvästi

$$f((1-\lambda)x+\lambda y) = \limsup f_n((1-\lambda)x+\lambda y)$$

$$\leq \limsup\big((1-\lambda)f_n(x)+\lambda f_n(y)\big)$$

$$\leq (1-\lambda)\limsup f_n(x) + \lambda \limsup f_n(y)$$

$$= (1-\lambda)f(x)+\lambda f(y), \quad \forall x,y\in\mathbb{R}^d,\ \lambda\in(0,1)$$

(toisessa epäyhtälössä tarvitaan tieto $f(x) > -\infty$, $\forall x$). Tämä todistaa kohdan a).

Olkoon $y \neq x_0$ mielivaltainen. Määrätään sellainen $\alpha > 1$, että

$$z := (1-\alpha)y + \alpha x_0 \in B(x_0,\delta).$$

Olkoon λ sellainen, että

$$x_0 = (1-\lambda)y + \lambda z,$$

ts. $\lambda = 1/\alpha \in (0,1)$. Tällöin

$$f_n(x_0) \leq (1-\lambda)f_n(y) + \lambda f_n(z), \ \forall n.$$

Koska $f(z) < \infty$, niin on joko $f(y) = \infty$ tai

$$f(x_0) \leq (1-\lambda)f(y) + \lambda f(z).$$

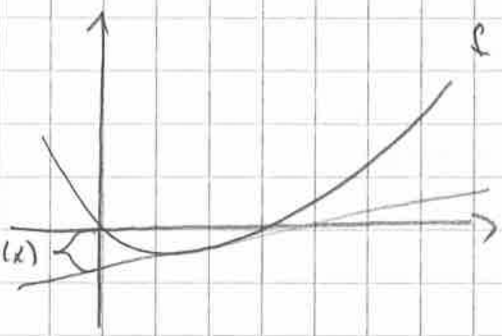
Välttämättä

$$f(y) \geq \frac{1}{1-\lambda}\Big(\underbrace{f(x_0)}_{\in\mathbb{R}} - \lambda \underbrace{f(z)}_{<\infty}\Big) > -\infty. \quad \square$$

Olkoon $f : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$ mielivaltainen.
Määritellään f:n <u>konveksi konjugaatti</u> $f^* : \mathbb{R}^d \to \overline{\mathbb{R}}$ ehdosta

$$f^*(x) = \sup\{\langle t, x\rangle - f(t) \mid t \in \mathbb{R}^d\}. \quad (*)$$

Jatkossa f tulee yleensä olemaan konveksi ja lisäksi $f(0) = 0$. Tällöin $f^*(x)$:llä on seuraava geometrinen tulkinta (tapauksessa $d = 1$).



Mainitaan ohimennen

1) f^* on konveksi konveksien funktioiden pisteittäisenä supremumina (itse asiassa f^* on affiinien kuvausten $x \mapsto \langle t, x\rangle - f(t)$ pisteittäinen supremum; määritelmässä voidaan rajoittua supremumiin yli joukon $t \in D(f)$, missä

$$D(f) = \{t \in \mathbb{R}^d \mid f(t) < \infty\}).$$

2) f^* on aps, koska se on aps-funktioiden pisteittäinen supremum.

$(*)$ $\langle t, x\rangle$ tarkoittaa $\mathbb{R}^d$:n sisätuloa: jos

$t = (t_1, \dots, t_d)$, $x = (x_1, \dots, x_d)$, niin

$$\langle t, x\rangle = t_1 x_1 + \dots + t_d x_d.$$

Todetaan vielä, että jos $f(t) = -\infty$ jollain $t \in \mathbb{R}^d$, niin $f^*(x) = +\infty$, $\forall x \in \mathbb{R}^d$.

Tarkastellaan nyt satunnaisvektorijonoa $\{Z_n\}$ kiinteässä todennäköisyyskentässä $(\Omega, \mathcal{F}, \mathbb{P})$. Aiemmin käytetty merkintä $P_n(A)$ korvataan kaikkialla merkinnällä $\mathbb{P}(Z_n \in A)$ (tämä on yhteensopivaa aiemman kanssa, mutta selkeyttää merkintöjä eikä merkitse alennusta yleisyyden menetystä).

Jonon $\{Z_n\}$ suuriin poikkeamiin kytkeytyvät funktiot

$$\Lambda_n : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\} \quad \text{ja} \quad \Lambda : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\},$$

$$\begin{cases} \Lambda_n(t) = n^{-1} \log \mathbb{E}\{ e^{n \langle t, Z_n \rangle} \}, \\ \\ \Lambda(t) = \limsup_{n \to \infty} \Lambda_n(t). \end{cases}$$

<u>Lemma 3.2.</u> Olkoon ξ mielivaltainen $\mathbb{R}^d$-arvoinen satunnaisvektori ja

$$f(t) = \log \mathbb{E}(e^{\langle t, \xi \rangle}), \quad t \in \mathbb{R}^d.$$

Silloin f on apj ja konveksi. Erityisesti siis Λ_n on aina apj ja konveksi.

(f on ξ:n <u>logaritminen generoiva funktio</u>)

<u>Todistus</u>. Fatoun lemman nojalla kiinteälle $t_0 \in \mathbb{R}^d$,

$$\liminf_{t \to t_0} \mathbb{E}(e^{\langle t, \xi \rangle})$$

$$\geq \mathbb{E}(\liminf_{t \to t_0} e^{\langle t, \xi \rangle}) = \mathbb{E}(e^{\langle t_0, \xi \rangle}).$$

Siis

$$\underline{\lim\inf}_{t \to t_0} f(t) \geq f(t_0),$$

joten f on ap j. Konveksisuus todistetaan Hölderin epäyhtälön nojalla:

jos $p, q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, niin

$$\mathbb{E}(UV) \leq \mathbb{E}(U^p)^{\frac{1}{p}} \, \mathbb{E}(V^q)^{\frac{1}{q}}$$

mielivaltaisille ei-negatiivisille satunnaismuuttujille U, V.

Olkoon $\lambda \in (0,1)$ ja $t, u \in \mathbb{R}^d$. Valitaan $p = \frac{1}{1-\lambda}$, $q = \frac{1}{\lambda}$. Tällöin

$$f((1-\lambda)t + \lambda u) = \log \mathbb{E}\{ e^{(1-\lambda)\langle t, \xi \rangle} e^{\lambda \langle u, \xi \rangle} \}$$

Hölder

$$\leq \log \mathbb{E}(e^{\langle t, \xi \rangle})^{1-\lambda} \mathbb{E}(e^{\langle u, \xi \rangle})^{\lambda}$$

$$= (1-\lambda) f(t) + \lambda f(u).$$

Siis f on konveksi, joten samoin on Λ_n, $\forall n$. $\square$

Tarkastellaan seuraavassa 'kumulatiivista' prosessia

$$Y_n = \xi_1 + \dots + \xi_n$$

ja valitaan

$$X_n = n^{-1} Y_n,$$

missä siis $\xi_1, \xi_2, \dots$ ovat $\mathbb{R}^d$-arvoisia satunnais-vektoreita. Tämä ei merkitse rajoitusta, mutta on luonnollinen erityistapa. Erityisesti siis

$$\Lambda_n(t) = n^{-1} \log \mathbb{E}\{ e^{\langle t, Y_n \rangle} \}.$$

Esimerkki 3.1. Olkoot $\xi, \xi_1, \xi_2, \dots$ riippumattomia ja samoin jakautuneita satunnaisvektoreita. Tällöin

$$\Lambda_n(t) = \log \mathbb{E}\{ e^{\langle t, \xi \rangle} \}, \forall n \in \mathbb{N}, t \in \mathbb{R}^d,$$

ja

$$\Lambda(t) = \log \mathbb{E}\{ e^{\langle t, \xi \rangle} \}, \forall t \in \mathbb{R}^d.$$

Lemma 3.4. Olkoon $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ konveksi ja $\mathring{D}(f) \neq \emptyset$. Silloin f on jatkuva $\mathring{D}(f)$:issä.

Lemman todistukseen palataan myöhemmin.

Lause 3.1 (Gärtner–Ellisin lause, ylärajat).

Mielivaltaiselle satunnaisvektorijonolle $\{Y_n\}$ pätee: Λ^* on konveksi vauhtifunktio ja

(3.1) $$\limsup_{n} n^{-1} \log \mathbb{P}(\bar{Z}_n \in F) \leq -\Lambda^*(F)$$

kaikille kompakteille $F \subseteq \mathbb{R}^d$. Jos $0 \in \mathring{D}(\Lambda)$, niin Λ^* on hyvä vauhtifunktio ja (3.1) pätee kaikille suljetuille $F \subseteq \mathbb{R}^d$.

Todistus. Koska $\Lambda(0) = 0$, niin

$$\Lambda^*(x) \geq \langle 0, x \rangle - \Lambda(0) = 0.$$

Jos $\Lambda(t) > -\infty$, $\forall t$, on Λ^* konveksi ja apj aiemmin todetun nojalla. Siis Λ^* on myös vauhtifunktio. Jos $\Lambda(t) = -\infty$ jollain t, on $\Lambda^*(x) = +\infty$ $\forall x$, siis konveksi ja apj.

Olkoon $F \subseteq \mathbb{R}^d$ kompakti. Esitetään (3.1):n todistus vain tapauksessa $\Lambda^*(F) < \infty$. Olkoon $\varepsilon > 0$. Mielivaltaiselle $x \in F$ kiinnitetään $t_x \in \mathbb{R}^d$:

$$\Lambda^*(x) \geq \langle t_x, x \rangle - \Lambda(t_x) > \Lambda^*(F) - \varepsilon.$$

Tällöin

$$x \in \{y \in \mathbb{R}^d \mid \langle t_x, y \rangle > \Lambda(t_x) + \Lambda^*(F) - \varepsilon\} \doteq H(t_x).$$

Ilmeisesti $H(t_x)$ on avoin, $\forall x$, ja

$$F \subseteq \bigcup_{x \in F} H(t_x).$$

Heine-Borelin lauseen nojalla voidaan valita $x_1, \dots, x_N \in F$:

$$(3.2) \qquad F \subseteq \bigcup_{i=1}^{N} H(t_{x_i}).$$

Tshebyshevin epäyhtälön nojalla

$$e^{n\Lambda_n(t_{x_i})} = \mathbb{E}\{ e^{\langle t_{x_i}, Y_n \rangle} \}$$

$$\geq \mathbb{E}\{ e^{\langle t_{x_i}, Y_n \rangle} 1(\langle t_{x_i}, Y_n \rangle > n(\Lambda(t_{x_i}) + \Lambda^*(F) - \varepsilon)\}$$

$$\geq e^{n(\Lambda(t_{x_i}) + \Lambda^*(F) - \varepsilon)} \mathbb{P}(\overline{X}_n \in H(t_{x_i})).$$

Siis

$$\limsup n^{-1} \log \mathbb{P}(\overline{X}_n \in H(t_{x_i})) \leq -\Lambda^*(F) + \varepsilon.$$

Lemman 2.1 ja (3.2):n nojalla

$$\limsup n^{-1} \log \mathbb{P}(\overline{X}_n \in F) \leq -\Lambda^*(F) + \varepsilon.$$

Näin ollen (3.1) pätee kompakteille joukoille.

Oletetaan nyt lisäksi, että $0 \in \mathring{\mathcal{D}}(\Lambda)$. Lemmojen 3.1 ja 3.2 nojalla Λ on konveksi. Lemman 3.4 nojalla voidaan määrätä $\delta > 0$:

$$\Lambda(t) \leq \alpha, \quad \forall t \in B(0, \delta).$$

Olkoon $\varepsilon = \delta/2$ ja $x \neq 0$. Silloin

$$\Lambda^*(x) \geq \left\langle \varepsilon \frac{x}{|x|}, x \right\rangle - \Lambda\left(\varepsilon \frac{x}{|x|}\right)$$

$$\geq \varepsilon |x| - \alpha.$$

Nähdään, että $\Lambda^*(x) \to \infty$, kun $|x| \to \infty$, joten Λ^*:in tasojoukot ovat kompakteja. Siis Λ^* on hyvä vauhtifunktio.

Todistetaan lopuksi (3.1) suljetuille joukoille. Sovelletaan lausetta 2.2. Merkitään komponenteittain

$$X_n = (X_{n1}, \ldots, X_{nd}), \quad Y_n = (Y_{n1}, \ldots, Y_{nd}).$$

Riittää todistaa, että $\{X_{ni}\}$ on eksponentiaalisesti tiukka, (harj.). Olkoon

$$\Lambda^{(i)}(t) = \limsup_{n \to \infty} n^{-1} \log \mathbb{E}\{e^{tY_{ni}}\}, \quad t \in \mathbb{R}.$$

Tällöin $0 \in \mathring{\mathcal{D}}(\Lambda^{(i)})$. Olkoon $t_0 > 0$ sellainen, että

$$\Lambda^{(i)}(t_0) < \infty.$$

Tšebyševin epäyhtälön nojalla mielivaltaiselle $M > 0$,

$$\mathbb{E}(e^{t_0 Y_{ni}}) \ge \mathbb{E}(e^{t_0 Y_{ni}} 1(Y_{ni} > nM)) \ge e^{t_0 nM}\, \mathbb{P}(\bar{X}_{ni} > M),$$

joten

$$\limsup_{n\to\infty} n^{-1} \log \mathbb{P}(\bar{X}_{ni} > M) \le \Lambda^{(i)}(t_0) - t_0 M.$$

Tämä saadaan mielivaltaisen pieneksi antamalla $M \to \infty$.

Samanlainen tulos saadaan todennäköisyyksille $\mathbb{P}(\bar{X}_{ni} < -M)$.

Lemman 2.1 nojalla $\{\bar{X}_{ni}\}$ on eksponentiaalisesti tiukka. $\square$

Suurten poikkeamien periaatteeseen pääsemiseksi on vielä tarkasteltava alarajoja. Tähän tarvitaan melko paljon konveksia analyysiä, jota tarkastellaan seuraavassa.