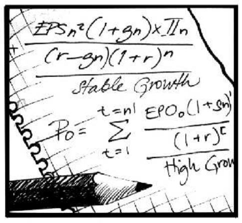


Sattuman matematiikkaa III

Kolmogorovin aksioomat ja frekvenssitulkinta

Tommi Sottinen
Tutkija
Matematiikan ja tilastotieteen laitos, Helsingin yliopisto
Laboratoire de Probabilités et Modèles Aléatoires, Université de Paris VI

Solmun numerossa 2/2002 aloitettiin todennäköisyyslaskentaa käsittelevä kirjoitussarja. Osassa I käsiteltiin todennäköisyyslaskennan historiaa ja muutamia todennäköisyyden tulkintoja: klassista, frekventististä ja geometrista. Osassa II (Solmu 1/2003) esitettiin modernin todennäköisyyslaskennan perusta: Kolmogorovin [2] aksioomat.

Tässä kirjoitussarjan kolmannessa osassa emme mene tarinassa eteenpäin vaan syvemmälle. Osoitamme, että Kolmogovin aksioomat ovat siinä mielessä sopiva matemaattinen malli todennäköisyyslaskennalle, että frekvenssitulkinta voidaan johtaa niistä. (On itsestään selvää, että klassinen ja geometrinen tulkinta seuraavat Kolmogorovin aksioomista.)

Kolmogorovin aksioomat

Kertaamme lyhyesti kirjoitussarjan osassa II esitetyt aksioomat, eli kolmikon $(\Omega, \mathcal{F}, \mathbf{P})$.

Ω on *perusjoukko*, josta kohtalon jumalatar, Lady Fortuna, valitsee satunnaiskokeen tuloksen ω.

$\mathcal{F}$ on kokoelma Ω:n osajoukkoja, joka on suljettu numeroituvan monien joukko-operaatioiden suhteen (siis $\mathcal{F}$ on σ-algebra, ks. osa II). Kutsumme $\mathcal{F}$:n jäseniä *tapahtumiksi*. Välttämättä kaikki Ω:n osajoukot eivät siis ole tapahtumia. Syy tähän valitettavaan seikkaan löytyy mittateorian syvistä vesistä. Emme käsittele tätä aihetta enempää. Lukija voi lohduttautua sillä, että käytännössä on vaikeaa keksiä osajoukkoa, joka ei ole tapahtuma.

$\mathbf{P}$ on *todennäköisyys*, siis kuvaus tapahtumajoukolta $\mathcal{F}$ reaalilukujoukolle $\mathbb{R}$, joka toteuttaa ehdot

(TN_1) $\mathbf{P}(A) \geq 0$ kaikilla tapahtumilla A,

(TN_2) $\mathbf{P}(\Omega) = 1$,

(TN_3) jos $A_1, A_2, \ldots$ ovat tapahtumia, joista korkeintaan yksi voi sattua kerrallaan, niin

$$\mathbf{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)\,.$$

Kohdat (TN_1) ja (TN_2) ovat luonnollisia. Kohta (TN_3), *täysadditiivisuus*, on vähemmän viaton. Siitä seuraa esimerkiksi, ettemme voi valita luonnollista lukua umpimähkään (siis siten, että jokaisella luvulla on yhtä suuri todennäköisyys tulla valituksi). Jokainen varmasti hyväksyy, että todennäköisyys on *additiivinen*:

(TN_3') jos tapahtumat A_1 ja A_2 eivät voi molemmat sattua samalla kertaa, niin

$$\mathbf{P}\big(A_1 \cup A_2\big) \;=\; \mathbf{P}\big(A_1\big) + \mathbf{P}\big(A_2\big).$$

Jos lisäksi hyväksymme, että todennäköisyys on *jatkuva*:

(TN_3'') jos tapahtumien jono $A_1, A_2, \ldots$ on laskeva, toisin sanoen $A_1 \supset A_2 \supset \cdots$, niin

$$\mathbf{P}\Big(\bigcap_{n=1}^{\infty} A_n\Big) \;=\; \lim_{n\to\infty} \mathbf{P}\big(A_n\big),$$

niin joudumme hyväksymään täysadditiivisuuden. Nimittäin (TN_3') yhdessä (TN_3''):n kanssa on yhtäpitävä (TN_3):n kanssa. Emme perustele tätä tässä, vaikkakaan perustelu ei ole erityisen hankala. Joka tapauksessa additiivisuus täydessä muodossaan on välttämätön frekvenssitulkinnan kannalta.

Huomautamme lopuksi, että täysadditiivisuudesta seuraa, että olivatpa joukot $A_1, A_2, \ldots$ erillisiä tai eivät, niin joka tapauksessa

$$\mathbf{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) \;\le\; \sum_{i=1}^{\infty} \mathbf{P}(A_i)\,. \tag{1}$$

Epäyhtälön (1) oikealla puolella on liikaa joukkojen $A_1, A_2, \ldots$ mahdolliset päällekkäisyydet. Kahden joukon tapauksessa tämä päällekkäisyys on helppo nähdä:

$$\begin{aligned}\mathbf{P}(A_1 \cup A_2) &= \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cap A_2)\\ &\le \mathbf{P}(A_1) + \mathbf{P}(A_2).\end{aligned}$$

Toistokoe: riippumattomien toistojen satunnaiskoe

Frekvenssitulkinnassa on kyse *toistokokeesta*, eli yhdestä ja samasta satunnaiskokeesta, jota toistetaan loputtomasti. Tällöin Ω:n alkiot ovat jonoja

$$\omega \;=\; (\omega_1, \omega_2, \ldots)\,.$$

Tässä ω_i on se alkio, jonka Lady Fortuna valitsee toistossa i. Lisäksi toistot ovat *riippumatomia*: jos A ja B ovat tapahtumia, jotka määräytyvät erillisten toistokertojen perusteella, niin

$$\mathbf{P}\big(A \cap B\big) \;= \mathbf{P}\big(A\big)\mathbf{P}\big(B\big).$$

Toistokokeella ei siis ole muistia: aikaisemmat tapahtumat eivät vaikuta tulevien tapahtumien todennäköisyyksiin.

Tyypillinen esimerkki toistokokeesta on kolikon heitto. Jos kolikko on joka heitolla samanlainen, se ei siis esimerkiksi kulu heitossa, niin toistot ovat riippumattomia.

Olkoon nyt A jokin yksittäiseen satunnaiskokeeseen liittyvä tapahtuma. Esimerkiksi kolikon heitossa se voisi olla ”kolikko laskeutuu klaavapuoli ylöspäin”. Koska kyse on toistokokeesta, merkitsemme

$$A_i \;=\; \{A \text{ sattuu toistossa } i\}\,.$$

Tapahtuma A_i riippuu ω:sta vain koordinaatin ω_i kautta. Siten A_i:t ovat riippumattomia.

Frekvenssitulkinta ja binomimuuttuja

Olkoon n luonnollinen luku. Tapahtuman A *frekvenssi*

$$\begin{aligned}F_n[A] &= \#\{i \,:\, A_i, i \le n\}\\ &= \big\{\text{niiden } i \le n \text{ lukumäärä, joilla } A_i\big\}\\ &= \big\{\text{niiden toistojen } i \le n \text{ lukumäärä,}\\ &\qquad \text{joilla } A_i \text{ sattuu}\big\}\end{aligned}$$

ja sen *suhteellinen frekvenssi*

$$f_n[A] \;=\; \frac{F_n[A]}{n}.$$

Jos $f_n[A]$ suppenee jossakin mielessä kohti jotain lukua p, niin tällöin *tulkitsemme*, että $p = \mathbf{P}(A)$.

Koska tapahtuma A on jatkossa aina sama, niin kirjoitamme lyhyesti $F_n = F_n[A]$ ja $f_n = f_n[A]$.

Käsittelemme nyt hieman suppenemista

$$f_n \;\to\; p. \tag{2}$$

Ongelma tämän suppenemisen ymmärtämisessä on se, että f_n ei ole mikään kiinteä luku. Se on *satunnaismuuttuja*, eli funktio perusjoukolta Ω reaaliluvuille $\mathbb{R}$.

Kiinnittämällä $\omega \in \Omega$ voimme tarkastella tavallista reaalilukujonojen suppenemista ja yrittää osoittaa esimerkiksi, että

$$f_n(\omega) \;\to\; p \qquad \text{kaikilla } \omega \in \Omega.$$

Tämä siis vastaa funktioiden pisteittäistä suppenemista. Emme kuitenkaan voi toivoa mitään näin hienoa tulosta. Tämän näemme tarkastelemalla kolikon heittoa. Olkoon A_i tapahtuma ”i:nnellä heitolla tulee klaava”. Jos $\omega = (\text{klaava}, \text{klaava}, \ldots)$, niin $f_n(\omega) = 1$. Toisaalta jos $\omega = (\text{kruuna}, \text{kruuna}, \ldots)$, niin $f_n(\omega) = 0$.

Määrittelemme suppenemisen (2) seuraavassa osiossa kahdella eri tavalla. Sitä ennen käsittelemme satunnaismuuttujia F_n ja f_n.

Oletamme nyt, että tapahtumilla A_i on todennäköisyys Kolmogorovin aksiomaattisessa mielessä. Merkitsemme $p = \mathbf{P}(A_i)$. Tässä $\mathbf{P}$ on todennäköisyys jonoavaruudessa Ω, vaikkakin itse tapahtuma liittyy vain yksittäiseen toistoon i. Tällöin F_n siis laskee "onnistuneiden" tapahtumien lukumäärän n:n toiston sarjassa, kun yksittäisen "onnistumisen" todennäköisyys on p. Satunnaismuuttuja F_n saa siis jonkin arvon joukosta $\{0, 1, \ldots, n\}$. Etsimme nyt satunnaismuuttujan F_n *jakauman*, toisin sanoen kuvauksen

$$k \ \mapsto\ \mathbf{P}\big(F_n = k\big).$$

Tarkastelkaamme tapahtumaa $\{F_n = k\}$. Tällöin siis A on sattunut k kertaa ja jäänyt sattumatta $n-k$ kertaa. Näin voi käydä mm. silloin, kun A sattuu aluksi k kertaa "putkeen", eli tapahtumat A_1, A_2, $\ldots$, A_k sattuvat, ja tämän jälkeen ei A enää satu, eli tapahtumat A^c_{k+1}, A^c_{k+2}, $\ldots$, A^c_n sattuvat. Koska $\mathbf{P}(A_i) = p$, niin $\mathbf{P}(A^c_i) = 1 - p$. Siten juuri kuvatun tapahtuman todennäköisyys on riippumattomuuden nojalla

$$p^k \cdot (1-p)^{n-k}. \tag{3}$$

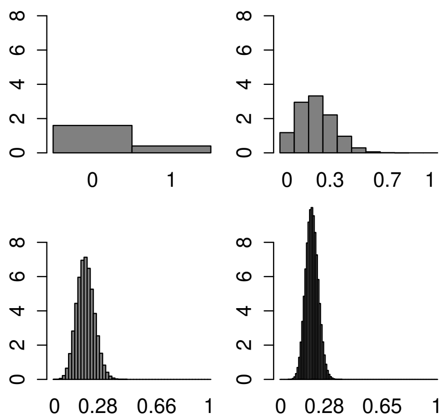
Yleisesti ottaen "onnistumisien" A_i ei tarvitse tapahtua aluksi "putkeen", vaan ne voivat tapahtua missä tahansa kohtaa n:ssä toistossa. Kuitenkin jokaisen yksittäisen n toiston tapahtuman, jossa on k kappaletta "onnistumisia", todennäköisyys on (3). Näitä yksittäisiä tapahtumia on, kuten kirjoitussarjan osassa I todettiin,

$$\binom{n}{k} \ =\ \frac{n!}{k!(n-k)!}$$

eri kappaletta. Siten, aksiooman (TN_3) nojalla,

$$\mathbf{P}(F_n = k) \ =\ \binom{n}{k} p^k (1-p)^{n-k}.$$

Sanomme, että F_n on *binomijakautunut* parametrein n ja p, ja käytämme merkintää $F_n \sim \mathrm{Bin}(n, p)$.



KUVA 1. Satunnaismuuttujan $f_n = F_n/n$ jakauma, kun $p = 0{,}2$ ja $n = 1, 10, 50, 100$.

Suurten lukujen lait

Tarkastelemme, missä mielessä raja-arvo (2), ja siten frekvenssitulkinta, voidaan ymmärtää. Jo aikaisemmin huomasimme, että funktioiden pisteittäinen suppeneminen on liian vahva käsite tässä yhteydessä.

Heikon suurten lukujen lain tapauksessa ymmärrämme suppenemisen $f_n \to p$ niin, että

$$\mathbf{P}\big(|f_n - p| \geq \varepsilon\big) \ \to\ 0 \tag{4}$$

millä tahansa luvulla $\varepsilon > 0$. Suppeneminen kaavassa (4) tarkoittaa tietysti tavallista reaalilukujonon suppenemista. Heikko suurten lukujen laki tarkoittaa siis sitä, että *todennäköisyys sille, että f_n poikkeaa luvusta p menee kohti nollaa, kun n kasvaa.* Sanomme myös, että f_n suppenee kohti lukua p *stokastisesti.*

Vahva suurten lukujen laki on lähellä funktioiden pisteittäistä suppenemista: ymmärrämme suppenemisen $f_n \to p$ niin, että

$$\mathbf{P}(f_n \to p) \ =\ \mathbf{P}\big(\{\omega \in \Omega : f_n(\omega) \to p\}\big) \ =\ 1. \tag{5}$$

Kyse on siis siitä, että *funktiot f_n suppenevat pisteittäin kohti lukua p paitsi ehkä jossakin poikkeuksellisessa pistejoukossa, jonka todennäköisyys on nolla.* Tällöin sanomme myös, että f_n suppenee kohti lukua p *melkein varmasti.*

Ensimmäisen version suurten lukujen laeista todisti Jakob Bernoulli [1]. Hänen kunniakseen satunnaiskoetta, jossa on kaksi tulosmahdolisuutta, kutsutaan *Bernoulli-kokeeksi* ja siten $\mathrm{Bin}(1, p)$-jakautuneesta satunnaismuuttujasta käytetään myös nimitystä *Bernoulli-muuttuja.*

Mainittokoon vielä, että kirjoittajan mielestä nimitys "suurten lukujen laki" ei ole erityisen onnistunut. Parempi nimitys olisi "loputtomien toistojen laki". Onneton nimitys lienee Siméon Poisson'n peruja.

Suurten lukujen lakien perustelu

Tämä on kirjoituksen tekninen osio, sen matemaattinen pihvi. Todennäköisyyslaskennan teoriasta vähemmän kiinnostunut lukija halunnee siirtyä suoraan osioon "Varoituksen sanoja".

Heikko tapaus

Tehtävänämme on löytää sellainen yläraja

$$(6) \qquad r(n,\varepsilon) \;\geq\; \mathbf{P}\big(|f_n - p| \geq \varepsilon\big)\,,$$

että $r(n,\varepsilon) \to 0$ kaikilla positiivisilla ε. Tämä ei itse asiassa ole erityisen vaikeaa. Ennakoimme kuitenkin vahvan tapauksen ja etsimme sellaisen ylärajan, joka suppenee riittävän nopeasti. Tämä on jo hieman hankalaa. Käytämme luennoissa [3] esitettyä tekniikkaa.

Tarkastelemme aluksi tapahtumassa

$$\big\{|f_n - p| \geq \varepsilon\big\} \;=\; \big\{|F_n - np| \geq n\varepsilon\big\}$$

itseisarvon positiivista puolta. Olkoon $r \geq 1$ ja $a \in (p, \varepsilon{+}p]$ sellainen luku, että $an \in \mathbb{N}$ (tällainen luku löytyy, kunhan n on riittävän iso). Koska F_n on $\mathrm{Bin}(n,p)$-jakautunut, niin

$$\begin{aligned}
&\mathbf{P}\big(F_n \geq (\varepsilon{+}p)n\big) \\
&\quad\leq\; \mathbf{P}\big(F_n \geq an\big) \\
&\quad=\; \sum_{k=an}^{n} \binom{n}{k} p^k (1{-}p)^{n-k} \\
&\quad=\; \frac{1}{r^{an}} \sum_{k=an}^{n} \binom{n}{k} r^m p^k (1{-}p)^{n-k} \\
&\quad\leq\; \frac{1}{r^{an}} \sum_{k=an}^{n} \binom{n}{k} (rp)^k (1{-}p)^{n-k} \\
&\quad\leq\; \frac{1}{r^{an}} \sum_{k=0}^{n} \binom{n}{k} (rp)^k (1{-}p)^{n-k}.
\end{aligned}$$

Binomiteoreeman, siis sen joka kertoo miten sulut avataan, nojalla

$$\sum_{k=0}^{n} \binom{n}{k} (rp)^k (1{-}p)^{n-k} \;=\; \big(rp + (1{-}p)\big)^n.$$

Siten

$$(7) \qquad \mathbf{P}\big(F_n \geq an\big) \;\leq\; \frac{1}{r^{an}}\big(rp - (1{-}p)\big)^n.$$

Epäyhtälön (7) vasen puoli ei riipu parametrin $r \geq 1$ valinnasta. Etsimme siten optimaalisen arvon r:lle. Optimikohta löytyy tavalliseen tapaan derivoimalla. Jätämme nämä työläät, mutta suoraviivaiset yksityiskohdat lukijalle. Toteamme vain, että minimikohta on

$$r_{\min} \;=\; \frac{1{-}p}{p} \cdot \frac{a}{1{-}a} \;>\; 1.$$

Sijoittamalla $r_{\min}$:n kaavaan (7) saamme ylärajan

$$\begin{aligned}
\mathbf{P}\big(f_n - p \geq \varepsilon\big) \;&\leq\; C_{a,p}\,(r_{\min})^{-an} \\
&=\; g_+(n,\varepsilon).
\end{aligned}$$

Tässä on tärkeää, että yläraja $g_+(n,\varepsilon)$ suppenee kohti nollaa eksponentiaalista vauhtia.

Tarkastelemme nyt itseisarvon negatiivista puolta. Vaihtamalla onnistumiset epäonnistumisiksi huomaamme, että satunnaismuuttuja $n - F_n$ on binomijakautunut parametrein n ja $1{-}p$. Koska

$$\{-F_n \geq na\} \;=\; \{n - F_n \geq (1{-}a)n\}\,,$$

niin voimme päätellä, kuten edellä, että

$$\mathbf{P}\big(f_n - p \leq -\varepsilon\big) \;\leq\; g_-(n,\varepsilon),$$

missä $g_-(n,\varepsilon)$ suppenee nollaan eksponentiaalista vauhtia.

Yhdistämällä saadut ylärajat olemme todistaneet heikon suurten lukujen lain. Voimme nimittäin valita ylärajaksi (6)

$$g(r,\varepsilon) \;=\; g_+(r,\varepsilon) + g_-(r,\varepsilon).$$

Vahva tapaus

Käytämme eksponentiaalista ylärajaa (6) ja seuraavaa tulosta.

Borel–Cantellin lemma. *Olkoot $A_1, A_2, \ldots$ sellaisia tapahtumia, että sarja*

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n)$$

suppenee. Tällöin A_n sattuu, melkein varmasti, vain äärellisen monella indeksillä n. Toisin sanoen

$$\mathbf{P}\big(A_n \text{ äärettömän usein}\big) \;=\; 0.$$

Tässä $\{A_n$ äärettömän usein$\}$ on niiden $\omega \in \Omega$ joukko, joilla $\omega \in A_n$ äärettömän usealla ideksillä n.

Perustelemme nyt Borel–Cantellin lemman. Merkitsemme aluksi

$$B_n \;=\; \bigcup_{i=n}^{\infty} A_i.$$

Toisin sanoen $B_n = \{A_i$ jollakin $i \geq n\}$. Siten

$$\big\{A_n \text{ äärettömän usein}\big\} \;=\; \bigcap_{n=1}^{\infty} B_n.$$

Joukot B_n ovat laskevia: $B_{n+1} \subset B_n$. Siten todennäköisyyden jatkuvuudesta (aksiooma (TN_3'')) seuraa, että

$$\mathbf{P}\Big(\bigcap_{n=1}^{\infty} B_n\Big) \;=\; \lim_{n\to\infty} \mathbf{P}(B_n).$$

Toisaalta epäyhtälöstä (1) seuraa, että

$$\mathbf{P}(B_n) \;\leq\; \sum_{i=n}^{\infty} \mathbf{P}(A_i).$$

Lemman väite seuraa kokoamalla yllä luettelemamme (epä)yhtälöt:

$$\begin{aligned}
&\mathbf{P}\big(A_n \text{ äärettömän usein}\big)\\
&\quad= \mathbf{P}\Big(\bigcap_{n=1}^{\infty} B_n\Big)\\
&\quad= \lim_{n\to\infty} \mathbf{P}(B_n)\\
&\quad\le \lim_{n\to\infty} \sum_{i=n}^{\infty} \mathbf{P}(A_i)\\
&\quad= 0.
\end{aligned}$$

Vahva suurten lukujen laki seuraa nyt suoraan Borel–Cantellin lemmasta. Nimittäin, jos

$$A_{n,k} \;=\; \left\{ \left|f_n - p\right| \;\ge\; \tfrac{1}{k} \right\},$$

niin $f_n \not\to p$ tarkoittaa, että $A_{n,k}$ sattuu jollakin $k \in \mathbb{N}$ äärettömän usein. Siis

$$\{f_n \not\to p\} \;=\; \bigcup_{k=1}^{\infty} \Big\{ A_{n,k} \text{ äärettömän usein} \Big\}.$$

Toisaalta ylärajan (6) nojalla $\mathbf{P}(A_{n,k}) \;\le\; g\big(n, \tfrac{1}{k}\big)$, missä $\big(g(n, 1/k)\big)_{n=1}^{\infty}$ suppenee sarjana. Siten, Borel–Cantellin lemman nojalla,

$$\mathbf{P}(A_{n,k} \text{ äärettömän usein}) = 0.$$

Lopulta väite seuraa epäyhtälöstä (1):

$$\begin{aligned}
\mathbf{P}(f_n \not\to p) &= \mathbf{P}\Big(\bigcup_{k=1}^{\infty} \{A_{n,k} \text{ äärettömän usein}\}\Big)\\
&\le \sum_{k=1}^{\infty} \mathbf{P}(A_{n,k} \text{ äärettömän usein})\\
&= \sum_{k=1}^{\infty} 0\\
&= 0.
\end{aligned}$$

Vahva suurten lukujen laki seurasi siis siitä, että kaikilla $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathbf{P}\big(\left|f_n - p\right| \ge \varepsilon\big) \;<\; \infty. \tag{8}$$

Satunnaismuuttujajono, joka toteuttaa ehdon (8), suppenee kohti lukua p *nopeasti*. Esitettyjen kolmen suppenemisen välinen suhde on:

$$\begin{gathered}
\text{nopea}\\
\Downarrow\\
\text{melkein varma}\\
\Downarrow\\
\text{stokastinen.}
\end{gathered}$$

Nämä implikaatiot ovat siinä mielessä aitoja, ettei niitä voida kääntää.

Varoituksen sanoja

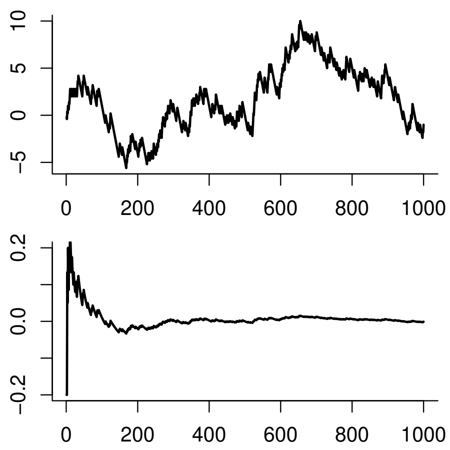
Frekvensitulkinnan mukaan *suhteellinen erotus*

$$\left|f_n - p\right| \;=\; \frac{\left|F_n - np\right|}{n}$$

suppenee kohti nollaa, kun n kasvaa. Absoluuttisen *erotuksen* tapauksessa kuitenkin

$$\left|F_n - np\right| \;\to\; \infty,$$

vieläpä niin että $F_n - np$ saa mielivaltaisen suuria ja pieniä arvoja. Todennäköisyys ei siis ole mikään kuminauha, jonka kohtalo pakottaa kohti keskiarvoa. Se ei vastaa näkemystä ”kosmisesta oikeudenmukaisuudesta”, jonka mukaan onnistumisien jälkeen on seurattava epäonnistumisia ja että jokainen on keskimäärin yhtä hyvä. Kohtalo voi toki muistaa aikaisemmat epäonnistumiset, mutta satunnainen riippumaton toistokoe ei niitä muista.



Kuva 2. Simuloidut polut $F_n - np$ ja $f_n - p$, kun $p = 0{,}2$ ja $n = 1, \ldots, 1000$.

Jos siis pelaat rulettia ja olet havainnut 9 punaista ja 1 mustan, niin ei kannata ruveta pelaamaan mustaa sen takia, että ”pitäähän niitä mustiakin tulla, kun on tullut niin paljon punaisia”. Itse asiassa nyt kannattaa pelata punaista! Syyn tähän kerromme seuraavissa kirjoituksissa.

Viitteet

[1] Bernoulli, Jakob: *Ars Conjectandi,* Basel, 1713.

[2] Kolmogorov, Andrei Nikolaevitš: *Grundbegriffe der Wahrscheinlichkeitsrechnung,* Berlin, 1933.

[3] Nummelin, Esa: *Todennäköisyysteoria,* Luennot, Helsingin yliopisto, Matematiikan laitos, 2003.