Todennäköisyyslaskennan ja tilastotieteen peruskurssi

Esimerkkikokoelma 2

Aiheet:

Satunnaismuuttujat ja todennäköisyysjakaumat
Kertymäfunktio, pistetodenäköisyysfunktio ja tiheysfunktio
Jakaumien tunnusluvut
Tärkeimmät diskreetit jakaumat
Tärkeimmät jatkuvia jakaumat

Satunnaismuuttujat ja todennäköisyysjakaumat

Satunnaismuuttujat, -luvut ja -vektorit

Satunnaismuuttuja X on kuvaus, joka liittää jokaiseen perusjoukon S pisteeseen $s \in S$ alkion $X(s) \in S'$. Alkio $X(s)$ on satunnaismuuttujan X **realisaatio** ja S' on satunnaismuuttujan X mahdollisten realisaatioiden joukko eli X:n **tilajoukko**. Yleisen satunnaismuuttujan määritelmässä vaaditaan lisäksi, että kuvaus X on sopivassa mielessä mitallinen, eli että muotoa $\{s \in S: X(s) \in B\}$ oleville tapahtumille voidaan määrittää todennäköisyys aina kun B on riittävän säännöllinen tilajoukon S' osajoukko. Kaikki tällä kurssilla käsiteltävät satunnaismuuttujat toteuttavat tämän teknisen ehdon, ja jatkossa kaikki tilajoukon osajoukot oletetaan ilman erillistä mainintaa riittävän säännöllisiksi.

Jos tilajoukko S':

- on reaalilukujen osajoukko, on X **reaaliarvoinen satunnaismuuttuja** eli **satunnaisluku**,
- on n-ulotteisen Euklidisen avaruuden osajoukko, on X **satunnaisvektori**,
- kattaa lukujonot $(x_1,x_2,x_3,...)$, on X **satunnaisjono**,
- on funktioavaruus, on X **satunnaisfunktio**.

Tällä kurssilla käsitellään lähes yksinomaan satunnaislukuja ja 2-ulotteisia satunnaisvektoreita. Ellei erikseen mainita, kaikki satunnaismuuttujat ovat jatkossa reaaliarvoisia.

Satunnaismuuttujan jakauma

Satunnaismuuttujan X **jakauma** on kuvaus, joka liittää jokaiseen tilajoukon osajoukkoon $B \subset S'$ luvun $P_X(B) = \Pr(X \in B)$, joka kertoo millä todennäköisyydellä X:n realisaatio kuuluu joukkoon B. Satunnaismuuttujan jakauma on tilajoukon S' **todennäköisyysjakauma**, eli kuvaus P_X toteuttaa todennäköisyyden aksioomat ja sille pätee samat laskusäännöt kuin otosavaruuden S todennäköisyysjakaumalle Pr.

Diskreetti satunnaismuuttuja

Satunnaismuuttuja X on **diskreetti**, jos sen tilajoukko S' on **numeroituva**, eli X:n mahdolliset realisaatiot voidaan numeroida joko muodossa $S' = \{x_1,x_2,...,x_n\}$ tai $S'=\{x_1,x_2,x_3,...\}$. Diskreetin satunnaismuuttujan **pistetodennäköisyysfunktio** on funktio

$$f(x_i) = \Pr(X = x_i),$$

joka kertoo millä todennäköisyydellä X saa arvon x_i. Diskreetin satunnaismuuttujan pistetodennäköisyysfunktio toteuttaa ehdot

$$f(x_i) \geq 0 \text{ kaikilla } i$$

ja

$$\sum_i f(x_i) = 1.$$

Todennäköisyys, että diskreetti satunnaismuuttuja saa arvon joukossa B voidaan laskea pistetodennäköisyysfunktion avulla muodossa

$$\Pr(X \in B) = \sum_{x_i \in B} x_i \,.$$

Näin ollen diskreetin satunnaismuuttujan X pistetodennäköisyysfunktio määrää X:n jakauman yksikäsitteisesti.

Jatkuva satunnaisluku

Satunnaisluku eli reaaliarvoinen satunnaismuuttuja X on *jatkuva*, jos sen jakauma voidaan esittää **tiheysfunktion** $f(x) \geq 0$ avulla muodossa

$$\Pr(X \in B) = \int_B f(x)\,dx$$

kaikilla reaaliakselin osajoukoilla B. Jatkuvan satunnaisluvun tiheysfunktio toteuttaa ehdon

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1.$$

Jatkuvan satunnaisluvun *pistetodennäköisyysfunktio on identtisesti nolla*, eli $\Pr(X = a) = 0$ kaikilla reaaliluvuilla a. Tämä paradoksaaliselta kuulostava fakta voidaan selittää niin, että $\{X=a\}$ tarkoittaa tapahtumaa, että X:n arvo on reaaliluku a äärettömän monen desimaalin tarkkuudella. Toisaalta $\Pr(a\text{-}h \leq X \leq a+h) \approx 2\,h\,f(a) > 0$ mielivaltaisen (mutta ei äärettömän) pienillä h>0 kaikissa niisä pisteissä a, missä $f(a) > 0$ ja f on jatkuva.

Kertymäfunktio

Satunnaisluvun eli reaaliarvoisen satunnaismuuttujan **kertymäfunktio**

$$F(x) = \Pr(X \leq x)$$

kertoo, millä todennäköisyydellä satunnaisluvun X arvo on enintään x.

Kertymäfunktiolle pätee

(1) $\lim_{x\to-\infty} F(x) = 0$

(2) $\lim_{x\to+\infty} F(x) = 1$

(3) F on *ei-vähenevä*:

$F(x_1) \le F(x_2)$, jos $x_1 \le x_2$

(4) F on *jatkuva oikealta*:

$\lim_{h\to 0+} F(x+h) = F(x)$

(5) $\Pr(X > x) = 1 - F(x)$

(6) $\Pr(a < X \le b) = F(b) - F(a)$

Diskreetin satunnaisluvun kertymäfunktio

Diskreetin satunnaisluvun kertymäfunktio X saadaan laskettua pistetodennäköisyysfunktion $f(x_i)$ avulla kaavasta

$$F(x) = \Pr(X \le x) = \sum_{i: xi \le x} f(x_i).$$

Jatkuvan satunnaisluvun kertymäfunktio

Jatkuvan satunnaisluvun X, jolla on tiheysfunktio f, kertymäfunktio F saadaan integroimalla tiheysfunktiota

$$F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(t)\,dt.$$

Tällöin tiheysfunktio määräytyy kertymäfunktiosta derivoimalla

$$f(x) = F'(x) = \frac{d}{dx} F(x).$$

Lisäksi X saa arvon väliltä (a,b] todennäköisyydellä

$$\Pr(a < X \le b) = F(b) - F(a) = \int_{a}^{b} f(x)\,dx.$$

Koska jatkuvalle satunnaisluvulle $\Pr(X=x) = 0$ kaikilla x, ei välin $(a,b]$ päätepisteillä ole merkitystä, ja pätee

$$\Pr(a < X \le b) = \Pr(a \le X < b) = \Pr(a < X < b) = \Pr(a \le X \le b).$$

Jakaumien tunnusluvut

Diskreetin satunnaisluvun odotusarvo

Diskreetin satunnaisluvun X, jonka pistetodennäköisyysfunktio on f, *odotusarvo* on ei-satunnainen vakio

$$E(X) = \sum_i x_i f(x_i) = \sum_i x_i \Pr(X = x_i).$$

Odotusarvon määritelmässä oletetaan, että yo. yhtälön oikealla puolella oleva summa suppenee ja tällöin sanotaan, että odotusarvo on olemassa.

Jatkuvan satunnaisluvun odotusarvo

Jatkuvan satunnaisluvun *X,* jolla on tiheysfunktio *f,* odotusarvo on ei-satunnainen vakio

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x f(x)\,dx.$$

Yllä oletetaan, että oikealla puolella oleva integraali on hyvin määritelty.

Odotusarvon ominaisuuksia

Satunnaisluvun X odotusarvo on X:n todennäköisyysjakauman massakeskipiste. Kaikille vakioille $a_1,a_2,\ldots$ ja kaikille satunnaisluvuille $X_1,X_2,\ldots$ pätee

$$\mathrm{E}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i\,\mathrm{E}(X_i).$$

Lisäksi vakion a odotusarvo on $\mathrm{E}(a) = a$.

Satunnaismuuttujan muunnoksen odotusarvo

Jos X on *diskreetti satunnaismuuttuja*, jonka pistetodennäköisyysfunktio on $f(x_i)=\Pr(X=x_i)$, ja g on deterministinen funktio X:n tilajoukosta reaaliluvuille, niin satunnaisluvun $Y=\mathrm{g}(X)$ odotusarvo saadaan kaavasta

$$\mathrm{E}(g(X)) = \sum_i g(x_i) f(x_i).$$

Jos X on *jatkuva satunnaisluku*, jolla on tiheysfunktio f, ja g on deterministinen funktio X:n tilajoukosta reaaliluvuille, niin satunnaisluvun $Y=\mathrm{g}(X)$ odotusarvo saadaan kaavasta

$$\mathrm{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x)\,dx.$$

Varianssi ja keskihajonta

Varianssi ja keskihajonta kuvaavat todennäköisyysjakauman hajontaa odotusarvonsa ympärillä. Satunnaisluvun X, jolla on odotusarvo μ_X, **varianssi** on ei-satunnainen vakio

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu_X)^2].$$

Varianssi voidaan laskea myös kaavalla

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mu_X^2,$$

missä $\mathrm{E}(X^2)$ on X:n toinen momentti. X:n **keskihajonta** on varianssin neliöjuuri

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Jos X on *diskreetti* satunnaisluku, jonka pistetodennäköisyysfunktio on $f(x_i)=\Pr(X=x_i)$, niin sen varianssi saadaan kaavasta

$$\mathrm{Var}(X) = \sum_i (x_i-\mu_X)^2 f(x_i).$$

Jos X on *jatkuva* satunnaisluku, jolla on tiheysfunktio f, niin varianssi saadaan kaavasta

$$\operatorname{Var}(X) = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f(x)\,dx.$$

Varianssin ominaisuuksia

Olkoon a vakio. Tällöin

$$\operatorname{Var}(a) = 0.$$

Olkoot X_i *riippumattomia* satunnaislukuja ja a_i vakioita. Tällöin

$$\operatorname{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \operatorname{Var}(X_i).$$

Kvantiilit ja mediaani

Satunnaisluvun X:n ja sen todennäköisyysjakauman **kvantiili** kertalukua $p \in (0,1)$ on reaaliluku x_p, joka toteuttaa yhtälön

$$F(x_p) = p,$$

missä F on X:n kertymäfunktio. Kvantiili x_p toteuttaa yhtälöt

$$\Pr(X \le x_p) = p \quad \text{ja} \quad \Pr(X > x_p) = 1 - p,$$

jakaa satunnaisluvun todennäköisyysmassan kahteen osaan niin, että massasta $p\times100$ % on kvantiilin x_p alapuolella ja $(1-p)\times100$ % aidosti x_p:n yläpuolella. Jos kertymäfunktiolla F on olemassa käänteisfunktio F^{-1}, niin $x_p = F^{-1}(p)$ kaikilla $0<p<1$. Muussa tapauksessa kvantiilit eivät ole yksikäsitteisiä.

Kertaluvun $p = 0.5$ kvantiilia kutsutaan **mediaaniksi**. Mediaani jakaa todennäköisyysmassan kahteen osaan niin, että massasta 50 % on mediaanin alapuolella ja 50 % aidosti sen yläpuolella.

Diskreettejä jakaumia

Diskreetti tasajakauma

Olkoon X diskreetti satunnaismuuttuja, jonka mahdolliset arvot ovat $x_1, x_2, \ldots, x_n$. Jos X saa jokaisen mahdollisen arvonsa samalla todennäköisyydellä, pätee

$$\Pr(X = x_i) = \frac{1}{n},\ i = 1,2,\ldots,n$$

ja tällöin satunnaismuuttuja X noudattaa **diskreettiä tasajakaumaa**, jonka pistetodennäköisyysfunktio on

$$f(x_i) = \frac{1}{n},\ i = 1,2,\ldots,n.$$

Tällöin X:n odotusarvo ja varianssi saadaan kaavoista

$$\operatorname{E}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{ja} \quad \operatorname{Var}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

Bernoulli-jakauma

Diskreetti satunnaismuuttuja X noudattaa Bernoulli-jakaumaa parametrilla p, merkitään $X \sim \mathrm{Ber}(p)$, jos sen arvojoukko on $\{0,1\}$ ja sen pistetodennäköisyysfunktio on

$$f(k) = p^k (1-p)^{1-k},\ k = 0,1.$$

Tällöin X:n odotusarvo ja varianssi ovat

$$\mathrm{E}(X) = p \quad \text{ja} \quad \mathrm{Var}(X) = p(1-p).$$

Jokainen $\{0,1\}$-arvoinen satunnaismuuttuja X noudattaa Bernoulli-jakaumaa parametrilla $p = \Pr(X=1)$.

Tapahtuman A **indikaattori** on satunnaismuuttuja 1_A, joka määritellään kaavalla:

$$1_A = \begin{cases} 1, & \text{jos tapahtuma } A \text{ sattuu,} \\ 0, & \text{jos tapahtuma } A \text{ ei satu.} \end{cases}$$

Tällöin 1_A on Bernoulli-jakautunut parametrilla $p=\Pr(A)$.

Binomijakauma

Diskreetti satunnaisluku X noudattaa **binomijakaumaa** parametrilla n ja p, merkitään $X \sim \mathrm{Bin}(n,p)$, jos sen arvojoukko on $\{0,1,...,n\}$ ja sen pistetodennäköisyysfunktio on

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0,1,2,\ldots,n.$$

Tällöin X:n odotusarvo ja varianssi ovat

$$\mathrm{E}(X) = np \quad \text{ja} \quad \mathrm{Var}(X) = np(1-p).$$

Tehdään n riippumatonta havaintoa satunnaisilmiöstä ja tarkastellaan kuinka monta kertaa tapahtuma A sattuu kyseisessä havaintosarjassa. Tällöin A:n esiintymiskertojen lukumäärä on diskreetti satunnaisluku $X \sim \mathrm{Bin}(n,p)$, missä $p = \Pr(A)$.

Binomijakauman ja Bernoulli-jakauman yhteys

Jos $X_1, X_2,\ldots, X_n$ ovat *riippumattomia* $\mathrm{Ber}(p)$-jakautuneita satunnaislukuja, niin summa

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

noudattaa binomijakaumaa parametrein n ja p.

Geometriset jakaumat

Diskreetti satunnaismuuttuja X noudattaa **lukujoukon {1,2,3...} geometrista jakaumaa** parametrinaan p, jos sen tilajoukko on $\{1,2,3...\}$ ja sen pistetodennäköisyysfunktio on

$$f(k) = (1-p)^{k-1} p,\ k = 1,2,3,\ldots$$

Tällöin X:n odotusarvo ja varianssi ovat

$$\mathrm{E}(X) = \frac{1}{p} \quad \text{ja} \quad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

Toistetaan tilastollista koetta, missä tehdään riippumattomia havaintoja satunnaisilmiöstä niin kauan, kunnes ennalta valittu tapahtuma A sattuu ensimmäisen kerran. Tällöin tarvittavien toistojen lukumäärä X noudattaa lukujoukon $\{1,2,3,...\}$ geometrista jakaumaa parametrinaan $p = \Pr(A)$.

Diskreetti satunnaismuuttuja Y noudattaa **lukujoukon {0,1,2,3...} geometrista jakaumaa** parametrinaan p, jos sen tilajoukko on $\{0,1,2,...\}$ ja sen pistetodennäköisyysfunktio on

$$f(k) = (1-p)^k p\,,\ k = 0,1,2,\ldots$$

Tällöin Y:n odotusarvo ja varianssi ovat

$$\mathrm{E}(X) = \frac{1-p}{p} \quad \text{ja} \quad \mathrm{Var}(Y) = \frac{1-p}{p^2}.$$

Jos X noudattaa lukujoukon $\{1,2,3,...\}$ geometrista jakaumaa parametrilla p, niin Y=X-1 noudattaa lukujoukon $\{0,1,2,...\}$ geometrista jakaumaa samalla parametrilla.

Hypergeometrinen jakauma

Diskreetti satunnaisluku X noudattaa **hypergeometrista jakaumaa** parametreinaan N,K, ja n, jos sillä on pistetodennäköisyysfunktio

$$f(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}},\ \max(0, n+K-N) \le k \le \min(n,K).$$

Tällöin X:n odotusarvo ja varianssi ovat

$$\mathrm{E}(X) = n\frac{K}{N} \quad \text{ja} \quad \mathrm{Var}(X) = n\frac{K}{N}\left(1-\frac{K}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Hypergeometrinen jakauma ilmenee tilastollisessa kokeessa, jossa halutaan selvittää kuinka suuri osuus äärellisen perusjoukon S alkioista kuuluu perusjoukon osajoukkoon A. Suoritetaan perusjoukosta n:n alkion satunnaisotanta *ilman palautusta* ja merkitään X:llä niiden havaintojen lukumäärää, jotka kuuluvat joukkoon A. Tällöin X noudattaa hypergeometrista jakaumaa parametreinaan N=n(S), K=n(A) ja n.

Hypergeometrisen jakauman arvioiminen binomijakaumalla

Jos perusjoukon koko N on suuri suhteessa otoskokoon n, niin hypergeometrisen jakauman todennäköisyydet ovat lähellä binomijakauman Bin(n,p) todennäköisyyksiä, missä

$$p = K/N$$

on joukon A tarkka koko suhteessa perusjoukkoon. Siten hypergeometrista jakaumaa voidaan approksimoida binomijakaumalla, kun **otantasuhde** n/N on riittävän pieni.

Satunnaisotanta ilman palautusta ja palauttaen

Satunnaisotanta on tilastollinen menetelmä selvittää, miten suuri osuus tietyn äärellisen perusjoukon S alkioista kuuluu perusjoukon osajoukkoon A.

Satunnaisotannassa **ilman palautusta** perusjoukosta S poimitaan satunnaisesti n alkiota ja tutkitaan, kuinka moni niistä kuuluu osajoukkoon A. Tällaisessa satunnaisotoksessa joukkoon A kuuluvien alkioiden lukumäärä on diskreetti satunnaisluku, joka noudattaa *hypergeometrista jakaumaa* parametreinaan N=n(S), K=n(A) ja n.

Satunnaisotannassa **palauttaen** perusjoukosta S poimitaan alkioita yksi kerrallaan tutkittavaksi ja jokainen tutkittu alkio palautetaan perusjoukkoon takaisin. Tällöin sama alkio voi tulla poimituksi monta kertaa. Diskreetti satunnaisluku, joka n:n toiston jälkeen kertoo kuinka monta kertaa on havaittu alkio, joka kuuluu joukkoon A, noudattaa *binomijakaumaa* Bin(n,p) parametreinaan n ja p = n(A)/n(S).

Jos satunnaisotannassa ilman palautusta otoskoko n on pieni suhteessa perusjoukon kokoon N, voidaan yllämainittua hypergeometrista jakaumaa approksimoida binomijakaumalla Bin(n,p), jonka parametrit ovat samat kuin satunnaisotannassa palauttaen. Tässä mielessä satunnaisotanta ilman palautusta antaa likimain samoja tuloksia kuin satunnaisotanta palauttaen, kun otantasuhde n/N on pieni.

Poisson-jakauma

Diskreetti satunnaisluku X noudattaa **Poisson-jakaumaa** parametrinaan $\lambda > 0$, merkitään $X \sim$ Poi(λ), jos sen pistetodennäköisyysfunktio on

$$f(k) = \frac{e^{-\lambda}\lambda^k}{k!}, k = 0,1,2,\ldots$$

Tällöin X:n odotusarvo ja varianssi ovat

$$\mathrm{E}(X) = \lambda \quad \text{ja} \quad \mathrm{Var}(X) = \lambda.$$

Binomijakauman arvioiminen Poisson-jakaumalla

Jos Bin(n,p)-jakauman parametri n on suuri ja parametri p lähellä nollaa niin, että jakauma odotusarvo on likimain $np \approx \lambda$, kyseistä binomijakaumaa voidaan approksimoida Poi(λ)-jakaumalla. Tässä mielessä Poisson-jakauma kuvaa harvinaisten riippumattomien tapahtumien lukumäärän jakaumaa pitkissä toistokoesarjoissa.

Jatkuvia jakaumia

Jatkuva tasajakauma

Satunnaisluku X noudattaa välin (a,b) **jatkuvaa tasajakaumaa**, jos sillä on tiheysfunktio

$$f(x) = \begin{cases} 0, & x < a \\ \dfrac{1}{b-a}, & a \le x \le b \\ 0, & x > a \end{cases}$$

Tällöin X:n odotusarvo ja varianssi ovat

$$\mathrm{E}(X) = \frac{a+b}{2} \quad \text{ja} \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

Eksponenttijakauma

Jatkuva satunnaismuuttuja X noudattaa *eksponenttijakaumaa* parametrinaan $\lambda > 0$, merkitään $\mathrm{Exp}(\lambda)$, jos sillä on tiheysfunktio $f(x)$, missä

$$f(x) = \lambda \exp(-\lambda x), \quad x \geq 0,$$

ja $f(x) = 0$, kun $x < 0$. Tällöin X:n odotusarvo ja varianssi ovat

$$\mathrm{E}(X) = \frac{1}{\lambda} \quad \text{ja} \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Eksponenttijakautunut satunnaisluku X toteuttaa muistittomuusominaisuuden:

$$\Pr(X > s+t \mid X > s) = \Pr(X > t) \quad \text{kaikilla s,t} > 0.$$

Normaalijakauma

Satunnaisluku X noudattaa **normaalijakaumaa** parametreinaan μ ja σ^2, merkitään $X \sim \mathrm{N}(\mu,\sigma^2)$, jos sillä on tiheysfunktio

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

Tällöin X:n odotusarvo ja varianssi ovat

$$\mathrm{E}(X) = \mu \quad \text{ja} \quad \mathrm{Var}(X) = \sigma^2.$$

Normaalijakauman tiheysfunktio tunnetaan nimellä **Gaussin kellokäyrä**. Tiheysfunktio on yksihuippuinen ja saa maksiminsa pisteessä μ. Lisäksi tiheysfunktio symmetrinen pisteen $x = \mu$ suhteen.

Normaalijakauman kertymäfunktio $F(x)$ saadaa integroimalla vastaavaa tiheysfunktiota $f(x)$. Kertymäfunktiota vastaavalle integraalille *ei ole tiedossa sievennettyä suljettua muotoa*. Tämän vuoksi normaalijakauman kertymäfunktion numeeriset arvot tulee katsoa taulukoista tai laskea tietokoneella.

Standardoitu normaalijakauma

Normaalijakauma N(0,1), jonka odotusarvo on 0 ja varianssi on 1, on **standardoitu normaalijakauma.** Mikä tahansa normaalijakautunut satunnaisluku voidaan *muuntaa standardoituun muotoon* vähentämällä ensin odotusarvo ja jakamalla sitten keskihajonnalla:
Jos $X \sim \mathrm{N}(\mu,\sigma^2)$, niin

$$Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$$

ja

$$\Pr(a \le X \le b) = \Pr\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) = \Pr\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right).$$

Siten normaalijakaumaan $\mathrm{N}(\mu,\sigma^2)$ liittyvät todennäköisyydet voidaan aina määrätä standardoidun normaalijakauman $\mathrm{N}(0,1)$ avulla.

Esimerkki

Määrää todennäköisyys $\Pr(1.5 \le X \le 3)$, kun X noudattaa normaalijakaumaa N(2,1/4).

Käytetään tehtävän ratkaisemisessa standardoidua normaalijakaumaa N(0,1), jonka kertymäfunktion arvoja on kurssilla jaettavissa taulukoissa. Koska X:n odotusarvo on $\mu = 2$ ja keskihajonta $\sigma = 1/2$,

$$\Pr(1.5 \le X \le 3) = \Pr\left(\frac{1.5-2}{1/2} \le \frac{X-2}{1/2} \le \frac{3-2}{1/2}\right) = \Pr(-1 \le Z \le 2),$$

missä

$$Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1).$$

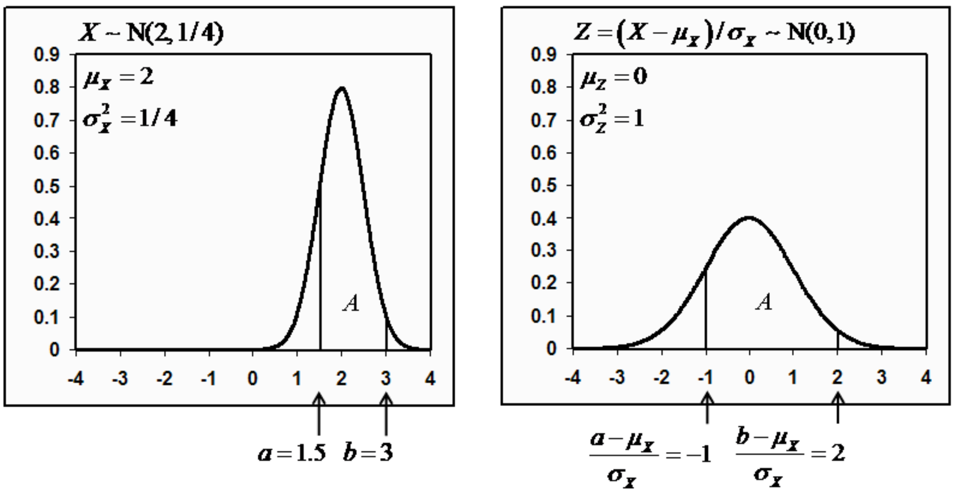
Edelleen

$$\Pr(-1 \le Z \le 2) = \Phi(2) - \Phi(-1),$$

missä Φ on standardoidun normaalijakauman kertymäfunktio, joka arvot löytyvät tilastollisista taulukoista tai tietokoneohjelmistoista. Sijoittamalla arvot $\Phi(-1)=0.1587$ ja $\Phi(2)=0.9772$ ylläolevaan kaavaan, havaitaan että

$$\Pr(1.5 \le X \le 3) = \Phi(2) - \Phi(-1) = 0.9772 - 0.1587 = 0.8185$$

Todennäköisyyden $\Pr(1.5 \le X \le 3)$ määrääminen vastaa alueen A pinta-alan määräämistä alla.



Riippumattomien normaalijakautuneiden satunnaislukujen summa

Jos satunnaisluvut $X_i \sim \mathrm{N}(\mu_i,\sigma_i^2)$ ovat keskenään riippumattomat, niin tällöin niiden summa

$$Y_n = \sum_{i=1}^{n} X_i$$

noudattaa normaalijakaumaa $N(\mu,\sigma^2)$ parametrein $\mu = \mu_1 + ... + \mu_n$ ja $\sigma^2 = \sigma_1^2 + ... + \sigma_n^2$.

Keskeinen raja-arvolause

Olkoot $X_1,...,X_n$ riippumattomia ja samoin jakautuneita satunnaislukuja, joiden odotusarvo on μ varianssi σ^2. Tällöin summan

$$Y_n = \sum_{i=1}^{n} X_i$$

odotusarvo on $n\mu$ varianssi $n\sigma^2$. Vähentämällä odotusarvo ja jakamalla keskihajonnalla voidaan Y_n standardoida muotoon

$$Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}}.$$

Kun $n \to \infty$, *keskeisen raja-arvolauseen* perusteella satunnaisluvun Z_n kertymäfunktio suppenee kohti standardoidun normaalijakauman N(0,1) kertymäfunktiota $\Phi(z)$, eli pätee

$$\lim_{n\to+\infty} \Pr\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le z \right) = \Phi(z).$$

Keskeisestä raja-arvolauseesta seuraa, että summan Y_n jakaumaa voidaan suurilla n arvioida normaalijakaumalla $N(n\mu,n\sigma^2)$.

Binomijakauman arvioiminen normaalijakaumalla tai Poisson-jakaumalla

Suurillä n:n arvoilla voidaan binomijakaumaa Bin(n,p) approksimoida tietyin ehdoin joko normaalijakaumalla tai Poisson-jakaumalla.

1. Jos $X \sim \text{Bin}(n,p)$ noudattaa binomijakaumaa, missä n on suuri ja p ei ole kovin lähellä nollaa eikä ykköstä, niin X:n kertymäfunktiota voidaan approksimoida normaalijakaumalla $N(np,np(1-p))$. Tällöin siis pätee

$$\Pr(a < X \le b) \approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right),$$

   missä Φ on standardoidun normaalijakauman N(0,1) kertymäfunktio.

2. Jos X ~ Bin(n,p), noudattaa binomijakaumaa, missä n on suuri ja p on lähellä nollaa niin, että binojakauman odotusarvo on likimain $np \approx \lambda$, kyseistä binomijakaumaa voidaan approksimoida Poi(λ)-jakaumalla.

Voidaan siis todeta, että suuressa satunnaisotoksessa *tyypillisiä* tapahtumia on havaintojen lukumäärä likimain normaalijakautunut, kun taas suuressa satunnaisotoksessa *harvinaisia* tapahtumia vastaava luku on likimain Poisson-jakautunut.

Hypergeometrisen jakauman arvioiminen normaalijakaumalla tai Poisson-jakaumalla

Hypergeometrinen jakauma ilmenee tilastollisessa kokeessa, jossa halutaan selvittää kuinka suuri osuus äärellisen perusjoukon S alkioista kuuluu perusjoukon osajoukkoon A. Suoritetaan perusjoukosta n:n alkion satunnaisotanta *ilman palautusta* ja merkitään X:llä niiden havaintojen lukumäärää, jotka kuuluvat joukkoon A. Tällöin X noudattaa hypergeometrista jakaumaa parametreinaan $N=\mathrm{n}(S)$, $K=\mathrm{n}(A)$ ja n.

Koska hypergeometrinen jakauma parametrein N,K,n on lähellä Bin($n,K/N$)-jakaumaa, kun otantasuhde n/N on pieni, voidaan binomijakauman arviota hyödyntämällä tehdä seuraavat johtopäätökset.

1. Kun otantasuhde n/N on pieni ja osajoukon A alkioiden osuus K/N on *tyypillinen* eli riittävän kaukana nollasta ja ykkösestä, voidaan hypergeometrista jakaumaa parametrein N,K,n approksimoida normaalijakaumalla N(μ,σ^2), missä

$$\mu = n\frac{K}{N},$$
$$\sigma^2 = n\frac{K}{N}\left(1-\frac{K}{N}\right).$$

2. Kun otantasuhde n/N on pieni ja osajoukon A alkioiden osuus K/N on *pieni* eli lähellä nollaa niin että $nK/N \approx \lambda$, voidaan hypergeometrista jakaumaa parametrein N,K,n approksimoida Poisson-jakaumalla Poi(λ), missä

$$\lambda = nK/N.$$

Esimerkki 3.3

Heitetään symmetristä kolikkoa 3 kertaa, jolloin siis Pr(Kruuna) = Pr(Klaava) = 1/2 ja olkoon satunnaismuuttuja X = Kruunien lukumäärä 3:ssa heitossa.

(a) Määrää todennäköisyydet tapahtumille $X = 0, 1, 2, 3$ ja määrittele niiden avulla satunnaismuuttujan X pistetodennäköisyysfunktio. Hahmottele funktion kuvaaja paperille.

(b) Määrää satunnaismuuttujan X kertymäfunktio. Hahmottele funktion kuvaaja paperille.

(c) Mikä on tapahtuman $X = 1.5$ todennäköisyys?

(d) Määrää tapahtuman $X > 1$ todennäköisyys sekä satunnaismuuttujan pistetodennäköisyys- että kertymäfunktion avulla.

Esimerkki 3.3 – Mitä opimme?

Esimerkissä 3.3. tarkastellaan *diskreetin satunnaismuuttujan* määrittelemistä sekä ko. satunnaismuuttujan *pistetodennäköisyysfunktion* ja *kertymäfunktion* määrittämistä puuverkon avulla.

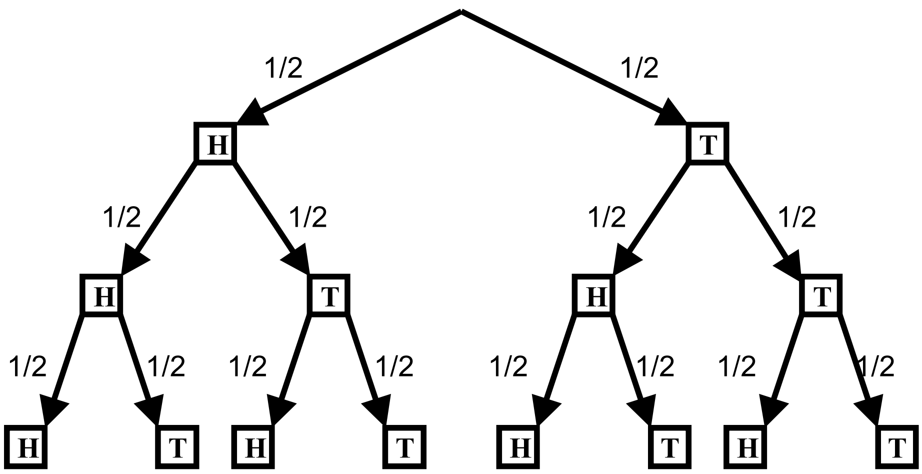
Esimerkki 3.3 – Ratkaisu

(a) Merkitään

H = Kruuna (engl. *head*),

T = Klaava (engl. *tail*).

Tehtävän satunnaisilmiön tulosvaihtoehdoista voidaan rakentaa alla oleva puuverkko:



Todennäköisyydet erilaisille kruunien ja klaavojen kombinaatioille voidaan laskea.

Tehtävän tapauksessa minkä tahansa kolmen kombinaation todennäköisyys on

$$\left(\frac{1}{2}\right)^3 = \frac{1}{8},$$

koska heitot ovat toisistaan riippumattomia.

Kombinaatioita, joissa on 0 kpl kirjainta H:ta, on 1 kpl. Siten $\Pr(\text{TTT}) = \frac{1}{8}$.

Kombinaatioita, joissa on 1 kpl kirjainta H, on 3 kpl. Siten

$$\Pr(\text{HTT tai THT tai TTH}) = \frac{3}{8}$$

käyttämällä poissulkevien tapahtumien yhteenlaskusääntöä.

Kombinaatioita, joissa on 2 kpl kirjainta H, on 3 kpl. Siten

$$\Pr(\text{HHT tai HTH tai THH}) = \frac{3}{8}.$$

Kombinaatioita, joissa on 3 kpl kirjainta H, on 1 kpl. Siten $\Pr(\text{HHH}) = \frac{1}{8}$.

Siten voimme esittää satunnaismuuttujan

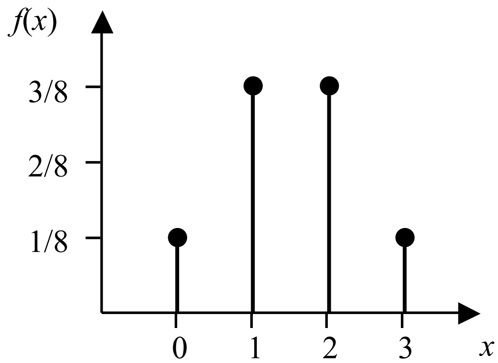
X = Kruunien lukumäärä 3:ssa kolikonheitossa

pistetodennäköisyysfunktion

$f(x) = \Pr(X = x), \quad x = 0, 1, 2, 3,$

alla vasemmalla olevana taulukkona. Alla oikealla on satunnaismuuttujan X pistetodennäköisyysfunktion kuvaaja.

x	$f(x) = \Pr(X = x)$
0	1/8
1	3/8
2	3/8
3	1/8



(b) Diskreetin satunnaismuuttujan *kertymäfunktio* voidaan laskea kaavalla

$$F(x) = \sum_{i \mid x_i \le x} \Pr(X = x_i).$$

Summassa lasketaan yhteen kaikki pistetodennäköisyydet

$$\Pr(X = x_i),$$

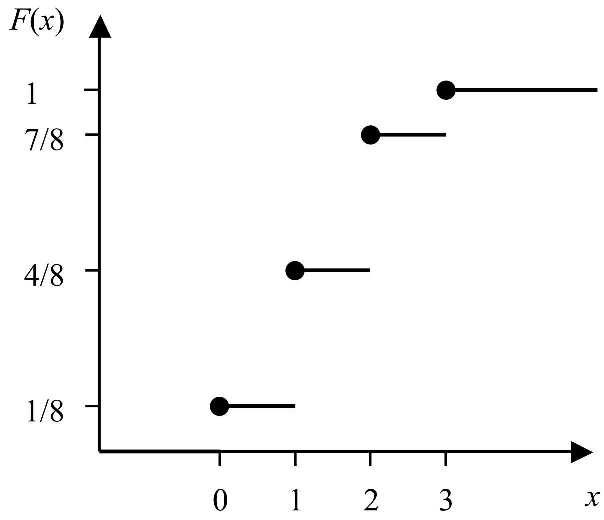
joille pätee $x_i \le x$.

Siten voimme esittää satunnaismuuttujan

X = Kruunien lukumäärä 3:ssa kolikonheitossa

kertymäfunktion alla vasemmalla olevana taulukkona. Alla oikealla on satunnaismuuttujan X kertymäfunktion kuvaaja.

x	$F(x)$
$x < 0$	0
$0 \le x < 1$	1/8
$1 \le x < 2$	4/8
$2 \le x < 3$	7/8
$3 \le x$	1



(c) Koska $X = 1.5$ on tapahtumana *mahdoton*, niin

$$\Pr(X = 1.5) = 0.$$

(d) Pistetodennäköisyysfunktiosta:

$$\Pr(X > 1) = \Pr(X = 2) + \Pr(X = 3) = 3/8 + 1/8 = 4/8 = 1/2.$$

Kertymäfunktiosta:

$$\Pr(X > 1) = 1 - \Pr(X \le 1) = 1 - F(1) = 1 - 4/8 = 4/8 = 1/2.$$

Esimerkki 3.4

Jatkuvalla satunnaismuuttujalla X on tiheysfunktio muotoa

$$f(x) = \begin{cases} x + b, \text{ kun } 0 \le x \le 1, \\ 0, \text{ muulloin.} \end{cases}$$

(a) Määrää vakion b arvo.

(b) Määrää tapahtuman $X = 0.5$ todennäköisyys.

(c) Määrää tapahtuman $0 \le X \le 0.5$ todennäköisyys.

(d) Määrää satunnaismuuttujan X kertymäfunktio.

Esimerkki 3.4 – Mitä opimme?

Esimerkissä 3.4 tarkastellaan jatkuvan satunnaismuuttujan *tiheysfunktiota* ja *kertymäfunktiota*.

Esimerkki 3.4 – Ratkaisu

(a) Koska kaikille tiheysfunktioille $f(x)$ pätee

$$\int_{-\infty}^{+\infty} f(x)dx = 1,$$

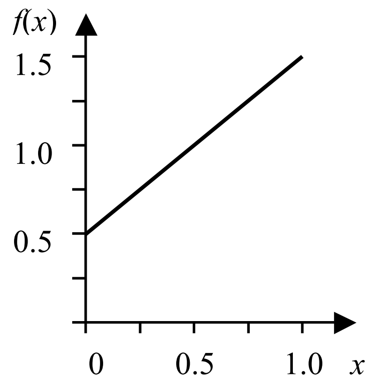
saadaan vakio b määrätyksi yhtälöstä

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_0^1 (x+b)dx = \left[\frac{1}{2}x^2 + bx\right]_0^1 = \frac{1}{2} + b,$$

ja ratkaisuksi saadaan

$b = 1/2.$

Oikealla on satunnaismuuttujan X tiheysfunktion $f(x)$ kuvaaja välillä [0, 1].



(b) Koska jatkuvilla jakaumilla jokaisen yksittäisen pisteen todennäköisyys on nolla, niin

$\Pr(X = 0.5) = 0.$

(c) Jos satunnaismuuttujalla X on tiheysfunktio on $f(x)$, niin

$$\Pr(a \le X \le b) = \int_a^b f(x)dx.$$

Siten

$$\Pr(0 \le X \le 0.5) = \int_0^{0.5} \left(x + \frac{1}{2}\right)dx = \left[\frac{1}{2}x^2 + \frac{1}{2}x\right]_0^{0.5} = \frac{3}{8}.$$

(d) Kun satunnaismuuttujalla X on tiheysfunktio on $f(x)$, sen kertymäfunktio saadaan kaavalla

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Siten tehtävän satunnaismuuttujan X kertymäfunktioksi saadaan välillä [0, 1]:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_0^x (t + \frac{1}{2})dt = \left[\frac{1}{2}t^2 + \frac{1}{2}t\right]_0^x = \frac{1}{2}x(x+1).$$

Tämän välin ulkopuolella

$F(x) = 0$, kun $x \le 0$,

$F(x) = 1$, kun $x \ge 1$.

Esimerkki 3.5

Osallistut rahapeliin, jossa heitetään kolmea symmetristä kolikkoa (ks. esimerkki 3.3). Peliin osallistumisesta pitää maksaa panos ja pelaaja saa voittona kruunien verran euroja.

(a) Mikä on korkein panos, mikä sinun kannattaa maksaa osallistumisesta peliin?

Ohje: Määrää satunnaismuuttujan

X = Kruunien lukumäärä 3:ssa rahanheitossa

odotusarvo.

(b) Mikä on voittosumman keskihajonta?

Esimerkki 3.5 – Mitä opimme?

Esimerkissä 3.5. tarkastellaan *odotusarvon*, *varianssin* ja *keskihajonnan* määräämistä esimerkin 3.3. diskreetille satunnaismuuttujalle.

Esimerkki 3.5 – Ratkaisu

Esimerkin 3.3 mukaan satunnaismuuttujan

X = Kruunien lukumäärä 3:ssa rahanheitossa

pistetodennäköisyysfunktio

$f(x) = \Pr(X = x)$, $x = 0, 1, 2, 3,$

voidaan esittää seuraavana taulukkona:

x	$f(x) = \Pr(X = x)$
0	1/8
1	3/8
2	3/8
3	1/8

(a) Diskreetin satunnaismuuttujan X, jonka pistetodennäköisyysfunktio on $f(x)$, odotusarvo saadaan kaavalla:

$$\mathrm{E}(X) = \sum_i x_i \Pr(X = x_i) = \sum_i x_i f(x_i) .$$

Tehtävän tapauksessa

$$\mathrm{E}(X) = \sum_{x=0}^{3} x \Pr(X = x) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{12}{8} = \frac{3}{2} = 1.5 .$$

Siten sinun kannattaa maksaa peliin osallistumisesta korkeintaan 1.5 €, koska se on odotettavissa oleva voitto.

Muista, että odotusarvo on satunnaismuuttujan todennäköisyysjakauman massakeskipiste. Huomaa, että tässä tapauksessa

$\Pr(X = \mathrm{E}(X)) = 0.$

(b) Satunnaismuuttujan *keskihajonta* on satunnaismuuttujan varianssin neliöjuuri. Määrätään siksi ensin satunnaismuuttujan X *varianssi*. Käytetään varianssin laskemiseen kaavaa

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \left[\mathrm{E}(X)\right]^2,$$

missä

$\mathrm{E}(X)$ = satunnaismuuttujan X odotusarvo,

$\mathrm{E}(X^2)$ = satunnaismuuttujan X toinen momentti.

Satunnaismuuttujan X odotusarvoksi saatiin (a)-kohdassa

$\mathrm{E}(X) = 3/2.$

Lasketaan satunnaismuuttujan X *toinen momentti*

$$\mathrm{E}(X^2) = \sum_{x=0}^{3} x^2 \Pr(X = x) = 0^2 \times \frac{1}{8} + 1^2 \times \frac{3}{8} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{8} = \frac{24}{8} = 3.$$

Siten

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \left[\mathrm{E}(X)\right]^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4} = 0.75\,.$$

Satunnaismuuttujan X keskihajonnaksi saadaan

$$\frac{\sqrt{3}}{2} = 0.866025\,.$$

Varianssi ja keskihajonta kuvaavat satunnaismuuttujan
hajontaa jakauman odotusarvon eli massakeskipisteen suhteen.

Esimerkki 3.6

Määrää esimerkin 3.4 todennäköisyysjakauman odotusarvo ja keskihajonta.

Esimerkki 3.6 – Mitä opimme?

Esimerkissä 3.6. tarkastellaan *odotusarvon*, *varianssin* ja *keskihajonnan* määräämistä esimerkin 3.4. jatkuvalle satunnaismuuttujalle.

Esimerkki 3.6 – Ratkaisu

Kun satunnaismuuttujalla X on *tiheysfunktio* $f(x)$, sen *odotusarvo* saadaan kaavalla

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x f(x)\,dx\,.$$

Esimerkin 3.4 tiheysfunktio:

$$f(x) = \begin{cases} x + 1/2\,, \text{ kun } 0 \le x \le 1, \\ 0\,, \text{ muulloin.} \end{cases}$$

Odotusarvo on siis

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_0^1 x\left(x + \frac{1}{2}\right)dx = \int_0^1 \left(x^2 + \frac{1}{2}x\right)dx = \left[\frac{1}{3}x^3 + \frac{1}{4}x^2\right]_0^1 = \frac{7}{12}.$$

Keskihajonta on satunnaismuuttujan varianssin neliöjuuri. Määrätään siksi ensin satunnaismuuttujan X *varianssi*. Käytetään varianssin laskemiseen kaavaa

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \left[\mathrm{E}(X)\right]^2,$$

missä

$\mathrm{E}(X)$ = satunnaismuuttujan X odotusarvo,

$\mathrm{E}(X^2)$ = satunnaismuuttujan X *toinen momentti.*

Satunnaismuuttujan X odotusarvoksi saatiin edellä

$\mathrm{E}(X) = 7/12.$

Määrätään satunnaismuuttujan X *toinen momentti*:

$$\mathrm{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_0^1 x^2\left(x + \frac{1}{2}\right)dx = \int_0^1\left(x^3 + \frac{1}{2}x^2\right)dx = \left[\frac{1}{4}x^4 + \frac{1}{6}x^3\right]_0^1 = \frac{5}{12}.$$

Siten *varianssiksi* saadaan

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \left[\mathrm{E}(X)\right]^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144} = 0.076389\,.$$

Satunnaismuuttujan X *keskihajonnaksi* saadaan

$$\frac{\sqrt{11}}{12} = 0.276385\,.$$

Esimerkki 3.8

Olkoon X diskreetti satunnaisluku, jonka pistetodennäköisyysfunktio on

$$f(x_i) = \Pr(X = x_i) = \frac{1}{n},\ i = 1,2,\ldots,n.$$

Satunnaismuuttuja X noudattaa joukon $\{x_1, x_2, x_3, x_4\}$ *diskreettiä tasajakaumaa*.

(a) Määrää satunnaismuuttujan X odotusarvo.

(b) Määrää satunnaismuuttujan X varianssi.

Esimerkki 3.8 – Mitä opimme?

Esimerkissä 3.8 johdetaan *diskreetin tasajakauman odotusarvo* ja *varianssi*.

Esimerkki 3.8 – Ratkaisu

Suoraan diskreetin jakauman odotusarvon määritelmän mukaan

$$\mathrm{E}(X) = \mu_X = \sum_{i=1}^n x_i f(x_i) = \sum_{i=1}^n x_i \frac{1}{n} = \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}\,,$$

mikä on lukujen $x_1, x_2, \ldots, x_n$ *aritmeettinen keskiarvo*.

Suoraan *diskreetin jakauman varianssin* määritelmän mukaan

$$\mathrm{Var}(X) = \sum_{i=1}^{n} (x_i - \mu_X)^2 f(x_i) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 ,$$

mikä on lukujen $x_1, x_2, \ldots , x_n$ *otovarianssi*.

Huomautus:

Aritmeettinen keskiarvo ja *otosvarianssi* ovat tärkeimpiä välimatka- ja suhdeasteikollisten muuttujien havaittujen arvojen tunnuslukuja.

Esimerkki 3.9

Jatkuva satunnaismuuttuja X noudattaa *eksponenttijakaumaa* parametrinaan $b>0$, kun sillä on tiheysfunktio $f(x)$, missä

$$f(x) = be^{-bx} , x > 0,$$

ja $f(x) = 0$ kun $x \leq 0$.

Määrää satunnaismuuttujan X odotusarvo nojaten suoraan odotusarvon määritelmään.

Opastus: Integroinneissa kannattaa käyttää osittaisintegrointia.

Esimerkki 3.9 – Mitä opimme?

Esimerkissä 3.9 johdetaan eksponenttijakauman odotusarvo.

Esimerkki 3.9 – Ratkaisu

Todetaan ensin, että yllä määritelty funktio $f(x)$ kelpaa tiheysfunktioksi, koska

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{0}^{\infty} be^{-bx}\,dx = \left[-e^{-bx}\right]_0^{\infty} = 1.$$

Jatkuvan jakauman odotusarvon määritelmästä seuraa soveltamalla *osittais-integrointia*:

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} xf(x)\,dx = \int_{0}^{\infty} xbe^{-bx}\,dx = \left[-xe^{-bx}\right]_0^{\infty} + \int_{0}^{\infty} e^{-bx}\,dx = \left[-\beta e^{-bx}\right]_0^{\infty} = 1/b.$$

Esimerkki 4.1

Pelaaja heittää symmetristä nelisivuista noppaa 12 kertaa. Oletetaan, että nopan sivut on merkitty silmäluvuilla 1, 2, 3 ja 4. Tällöin jokaisella silmäluvulla on sama todennäköisyys realisoitua.

(a) Laske silmälukujen summan odotusarvo, varianssi ja keskihajonta.

(b) Pelaaja saa voittonaan silmälukujen summan euroina kymmenkertaisena. Mikä on voiton odotusarvo ja standardipoikkeama? Kannattaako peliin osallistua, jos osallistuminen maksaa 400 euroa?

Esimerkki 4.1 – Mitä opimme?

Esimerkissä 4.1. sovelletaan *diskreettiä tasajakaumaa*.

Esimerkki 4.1 – Ratkaisu

Koska noppa on oletettu symmetriseksi, voidaan olettaa, että nopanheiton tulos X on diskreetti satunnaismuuttuja, joka noudattaa joukon $\{1,2,3,4\}$ tasajakaumaa.

Satunnaismuuttujan X pistetodennäköisyysfunktio on

$$f(x) = \Pr(X = x) = 1/4, \quad x = 1, 2, 3, 4.$$

Satunnaismuuttujan X odotusarvo on

$$\mathrm{E}(X) = \sum_{x=1}^{4} x\Pr(X = x) = 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 3\cdot\frac{1}{4} + 4\cdot\frac{1}{4} = \frac{1+2+3+4}{4} = \frac{10}{4} = 2.5\,.$$

Satunnaismuuttujan X toinen momentti on

$$\mathrm{E}(X^2) = \sum_{x=1}^{4} x^2\Pr(X = x)$$

$$= 1^2\cdot\frac{1}{4} + 2^2\cdot\frac{1}{4} + 3^2\cdot\frac{1}{4} + 4^2\cdot\frac{1}{4} = \frac{1^2+2^2+3^2+4^2}{4} = \frac{30}{4} = 7.5$$

Satunnaismuuttujan X *varianssi* on

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \left[\mathrm{E}(X)\right]^2 = \frac{30}{4} - \left[\frac{10}{4}\right]^2 = \frac{20}{16} = 1.25\,.$$

Satunnaismuuttujan X *keskihajonta* on varianssin neliöjuuri eli

$$\sqrt{1.25} = 1.118034\,.$$

Kun tetraedrin muotoista noppaa heitetään n kertaa, jokaisen heiton tulos X_i , $i = 1, 2, \ldots , n$ on satunnaismuuttuja, joka noudattaa edellä määriteltyä diskreettiä tasaista jakaumaa. Lisäksi voimme olettaa, että heittojen tulokset ovat toisistaan *riippumattomia*.

(a) Heittotulosten *summa*

$$Z = \sum_{i=1}^{12} X_i$$

on diskreetti satunnaismuuttuja, joka on *riippumattomien*, *samaa* (edellä määriteltyä) *diskreettiä tasajakaumaa noudattavien satunnaismuuttujien* X_i , $i = 1, 2, \ldots , 12$ *summa*.

Summan Z *odotusarvo*:

$$\mathrm{E}(Z) = \mathrm{E}\left(\sum_{i=1}^{12} X_i\right) = \sum_{i=1}^{12}\mathrm{E}(X_k) = 12\times 2.5 = 30$$

Huomaa, että satunnaismuuttujien summan odotusarvo on *aina* satunnaismuuttujien odotusarvojen summa – myös silloin kun summattavat ovat riippuvia.

Summan Z *varianssi*:

$$Var(Z) = Var\left(\sum_{i=1}^{12} X_i\right) = \sum_{i=1}^{12} Var(X_i) = 12 \times 1.25 = 15$$

Huomaa, että satunnaismuuttujien summan varianssi on satunnaismuuttujien varianssien summa yleensä *vain silloin*, kun summattavat ovat ovat riippumattomia.

Summan Z *keskihajonta* on

$$\mathrm{D}(Z) = \sqrt{15} = 3.873$$

Huomaa, että

$$\mathrm{D}(Z) \neq 12 \times \mathrm{D}(X_i)$$

ts. satunnaismuuttujien summan keskihajonta *ei ole* satunnaismuuttujien keskihajontojen summa. Tämä johtuu tietysti siitä, että summan neliöjuuri ei ole ko. lukujen neliöjuurien summa.

(b) Pelaajan saama *voitto*

$$Y = 10 \times Z$$

on *diskreetti satunnaismuuttuja*.

Voiton Y *odotusarvo*:

$$\mathrm{E}(Y) = 10 \times \mathrm{E}(Z) = 300 \text{ €}$$

Voiton Y *varianssi*:

$$\mathrm{D}^2(Y) = 10^2 \times \mathrm{D}^2(Z) = 1500 \text{ €}^2$$

Voiton Y *keskihajonta*:

$$\mathrm{D}(Y) = 38.73 \text{ €}$$

Koska peliin osallistuminen maksaa 400 €, pelaajat kärsivät jokaisessa pelissä keskimäärin *tappion*, jonka suuruus on

$$400 - \mathrm{E}(Y) = 400 - 300 = 100 \text{ €}$$

Esimerkki 4.2

Ruuveja valmistava kone tekee viallisia ruuveja todennäköisyydellä 1/10. Poimitaan koneen valmistamien ruuvien joukosta 20 ruuvia tarkastettavaksi satunnaisesti yksi kerrallaan.

Oletetaan, että koneen valmistamien ruuvien lukumäärä on niin suuri otoskokoon 20 verrattuna, että voimme approksimoida viallisten ruuvien lukumäärän jakaumaa otoksessa *binomijakaumalla*.

(a) Mikä on todennäköisyys, että viallisia ruuveja löydetään *täsmälleen* 2 kpl?

(b) Mikä on todennäköisyys, että viallisia ruuveja löydetään *vähintään* 1 kpl?

(c) Mikä on odotusarvo viallisten ruuvien lukumäärälle?

Esimerkki 4.2 – Mitä opimme?

Esimerkissä 4.2 sovelletaan *binomijakaumaa*.

Esimerkki 4.2 – Ratkaisu

Ruuveja valmistava kone tekee viallisia ruuveja todennäköisyydellä 1/10. Poimitaan satunnaisesti 20 ruuvia tarkastettavaksi yksi kerrallaan. Oletetaan, että koneen valmistamien ruuvien lukumäärä on hyvin suuri otoskokoon 20 verrattuna. Tällöin viallisten ruuvien lukumäärä X tarkastettavaksi poimittujen 20 ruuvin joukossa on *diskreetti satunnaismuuttuja*, jonka jakaumaa voidaan approksimoida *binomijakaumalla*:

$$X \sim_a \mathrm{Bin}(n, p)$$

jossa

$$n = 20$$

$$p = 0.1$$

Tämä perustuu seuraavaan:

Tarkastettavaksi poimittujen 20 ruuvin joukko muodostaa *yksinkertaisen satunnaisotoksen* valmistettujen ruuvien joukosta. Otosta poimittaessa on käytetty otantaa *ilman palautusta*. Tällöin viallisten ruuvien lukumäärä otoksessa noudattaa *hypergeomeristä jakaumaa*.

Koska valmistettujen ruuvien lukumäärä on oletettu hyvin suureksi otoskokoon 20 verrattuna, on *otantasuhde* niin pieni, että voimme approksimoida hypergeometrista jakaumaa binomijakaumalla.

Satunnaismuuttujan X pistetodennäköisyysfunktio on likimain

$$f(x) = \Pr(X = x) = \binom{20}{x} p^x (1-p)^{20-x}\ ,\ p = 0.1\ ,\ x = 0,1,2,\ldots,20$$

(a) Todennäköisyys, että viallisia ruuveja löydetään *täsmälleen* 2 kpl, on

$$\Pr(X = 2) = \binom{20}{2} \times 0.1^2 \times 0.9^{18} \approx 0.285$$

(b) Se, että viallisia ruuveja löydetään *vähintään* 1 kpl, voidaan esittää tapahtumana seuraavassa muodossa:

$$\{X > 0\} = \{X = 1\} \cup \{X = 2\} \cup \mathrm{L} \ \cup \{X = 20\}$$

Määrätään todennäköisyys tälle tapahtumalle soveltamalla *vastakohdan todennäköisyyden kaavaa*:

$$\Pr(X > 0) = 1 - \Pr(X = 0) = 1 - \binom{20}{0} \times 0.1^0 \times 0.9^{20} = 1 - 0.122 = 0.878$$

(c) Odotusarvo viallisten ruuvien lukumäärälle on

$$\mathrm{E}(X) = np = 20 \times 0.1 = 2$$

Esimerkki 4.3

Tehdas valmistaa tuotetta, jolla on erittäin korkeat laatukriteerit. Keskimäärin vain 60 % tuotteista täyttää kriteerit. Poimitaan palauttaen tuotteita tarkastettavaksi satunnaisesti yksi kerrallaan.

(a) Mikä on todennäköisyys sille, että joudumme tarkastamaan *vähintään* 4 tuotetta *ensimmäisen* viallisen tuotteen löytämiseksi?

(b) Oletetaan, että olemme tarkastaneet 3 tuotetta löytämättä yhtään viallista tuotetta. Mikä on todennäköisyys, että joudumme tarkastamaan *vähintään* 4 tuotetta *lisää* ensimmäisen viallisen tuotteen löytämiseksi?

(c) Mikä on odotettavissa oleva lukumäärä tuotteille, jotka joudumme tarkastamaan *ensimmäisen* viallisen tuotteen löytämiseksi?

Esimerkki 4.3 – Mitä opimme?

Esimerkissä 4.3 sovelletaan lukujoukon {1,2,3,...} *geometrista jakaumaa*.

Esimerkki 4.3 – Ratkaisu

Poimitaan tuotteita tarkastettavaksi satunnaisesti yksi kerrallaan. Ensimmäisen viallisen tuotteen järjestysnumero X on diskreetti satunnaismuuttuja, joka noudattaa lukujoukon {1,2,3,...} geometrista jakaumaa parametrilla

$$p = 1 - 0.6 = 0.4,$$

joka on todennäköisyys löytää viallinen tuote.

Satunnaismuuttujan X pistetodennäköisyysfunktio on

$$f(x) = \Pr(X = x) = q^{x-1}p\,,\ p = 0.4\,,\ q = 1 - p = 0.6\,,\ x = 1,2,3,\ldots$$

Satunnaismuuttujan X kertymäfunktio kokonaislukupisteissä on

$$F(k) = \Pr(X \le k) = 1 - (1-p)^k, k = 1,2,3\ldots,$$

(a) Todennäköisyys sille, että joudumme tarkastamaan *vähintään* 4 tuotetta *ensimmäisen* viallisen tuotteen löytämiseksi, on

$$\begin{aligned}
\Pr(X \ge 4) &= \Pr(X > 3) \\
&= 1 - \Pr(X \le 3) \\
&= 1 - F(3) \\
&= 1 - [1 - (1-p)^3] \\
&= (1-p)^3 \\
&= 0.6^3 = 0.216
\end{aligned}$$

(b) Jatkoa kohdalle (a).

Oletetaan, että olemme tarkastaneet 3 tuotetta löytämättä yhtään viallista tuotetta. Tällöin ensimmäisen viallisen tuotteen järjestysnumeron on oltava 4 tai suurempi. Se, että joudumme tarkastamaan *vähintään* 4 tuotetta *lisää* ensimmäisen viallisen tuotteen löytämiseksi, merkitsee sitä, että joudumme kaikkiaan tarkastamaan *vähintään* 7 tuotetta.

Siten kysytty todennäköisyys on ehdollinen todennäköisyys

$$\begin{aligned}\Pr(X \geq 7 \mid X \geq 4) &= \frac{\Pr(X \geq 7 \ ja\ X \geq 4)}{\Pr(X \geq 4)} \\ &= \frac{\Pr(X \geq 7)}{\Pr(X \geq 4)} \\ &= \frac{\Pr(X > 6)}{\Pr(X > 3)} \\ &= \frac{1-F(6)}{1-F(3)} \\ &= \frac{0.6^6}{0.6^3} = 0.6^3 = 0.216\end{aligned}$$

Toisaalta todennäköisyys, että joudumme tarkastamaan *vähintään* 4 tuotetta *ensimmäisen* viallisen tuotteen löytämiseksi, on (a)-kohdan mukaan

$$\Pr(X \geq 4) = 1 - F(3) = 0.6^3 = 0.216$$

Se, että olemme saaneet (a)- ja (b)-kohdassa saman tuloksen ei ole sattumaa, vaan tulos voidaan yleistää seuraavaan muotoon: Jos satunnaismuuttuja X noudattaa geometrista jakaumaa, niin aina pätee

$$\Pr(X \geq a+b \mid X \geq a) = \Pr(X \geq 1+b)\,,\ a = 1,2,3,\ldots\,,\ b = 0,1,2,3,\ldots$$

Tulos merkitsee sitä, että geometrisella jakaumalla on ns. *unohtamisominaisuus*:

Todennäköisyys joutua tarkastamaan vähintään b tuotetta *lisää* ensimmäisen viallisen tuotteen löytämiseksi *ei riipu* siitä, kuinka monta tuotetta on aikaisemmin jouduttu tarkastamaan löytämättä yhtään viallista. Voimme ilmaista tämän sanomalla, että tarkastusprosessi *unohtaa* oman "historiansa".

(c) Jatkoa kohdalle (a).

Odotettavissa oleva lukumäärä tuotteille, jotka joudumme tarkastamaan kunnes löydämme *ensimmäisen* viallisen tuotteen, on

$$E(X) = 1/p = 1/0.4 = 2.5$$

Esimerkki 4.4

Pakkauksessa on 100 tuotetta, joista 20 on viallisia.

(a) Poimitaan pakkauksesta 5 tuotetta tarkastettavaksi otannalla *ilman palautusta*.
Mikä on todennäköisyys, että otokseen tulee täsmälleen 1 viallinen tuote?
Mikä on odotettavissa olevien viallisten tuotteiden lukumäärä otoksessa?

(b) Poimitaan pakkauksesta 5 tuotetta tarkastettavaksi otannalla *palauttaen*.
Mikä on todennäköisyys, että otokseen tulee täsmälleen 1 viallinen tuote?
Mikä on odotettavissa olevien viallisten tuotteiden lukumäärä otoksessa?

Esimerkki 4.4 – Mitä opimme?

Esimerkissä 4.4 sovelletaan *binomijakaumaa* ja *hypergeometrista jakaumaa* satunnaisotantaan *palautuksella* ja *ilman palautusta*.

Esimerkki 4.4 – Ratkaisu

Tehtävän tapauksessa perusjoukon S koko on

$$N = n(S) = 100$$

viallisten tuotteiden joukon $A \subset S$ koko on

$$r = n(A) = 20$$

ja otoksen $B \subset S$ koko on

$$n = n(B) = 5$$

Määritellään *diskreetti satunnaismuuttuja*

X = ”Viallisten tuotteiden lukumäärä otoksessa”

Satunnaismuuttujan X jakauma riippuu siitä poimitaanko otos *ilman palautusta* tai *palauttaen*: Jos otos poimitaan *ilman palautusta*, X noudattaa *hypergeometrista jakaumaa*. Jos otos poimitaan *palauttaen*, X noudattaa *binomijakaumaa*.

Huomaa, että tehtävän tapauksessa otantasuhde

$$n/N = 0.05$$

joten binomijakauman pitäisi melko hyvin *approksimoida* hypergeometrista jakaumaa.

(a) Jos otoksen poiminta tapahtuu *ilman palautusta*, niin

$$X \sim \mathrm{HyperGeom}(N, r, n)$$

jossa

$$N = 100$$

$$r = 20$$

$$n = 5$$

Siten todennäköisyys, että otokseen poimittujen joukossa on 1 viallinen, on

$$f(1) = \Pr(X = 1) = \frac{\binom{20}{1}\binom{100-20}{5-1}}{\binom{100}{5}} = 0.420$$

Satunnaismuuttujan X *odotusarvo* on

$$\mathrm{E}(X) = n\frac{r}{N} = 5 \times \frac{20}{100} = 1$$

Vertaa saatuja tuloksia (b)-kohdan tuloksiin.

(b) Jos otanta suoritetaan *palauttaen*, viallisten lukumäärä tarkastettujen joukossa on

$X \sim \mathrm{Bin}(n, p)$

jossa

$n = 5$

$p = r/N = 0.2$

Siten todennäköisyys, että otokseen poimittujen joukossa on täsmälleen 1 viallinen, on

$$f(1) = \Pr(X = 1) = \binom{5}{1} \times 0.2 \times 0.8^4 = 0.410$$

Satunnaismuuttujan X *odotusarvo* on

$$\mathrm{E}(X) = np = 5 \times 0.2 = 1$$

Vertaa saatuja tuloksia (a)-kohdan tuloksiin.

Esimerkki 4.5

Taksikeskuksen palvelujonoon tulevien asiakkaiden lukumäärä minuutin aikana noudattaa Poisson-jakaumaa niin, että keskimäärin jonoon tulee 4 asiakasta minuutissa.

(a) Mikä on todennäköisyys, että minuutin kuluessa ei tule yhtään asiakasta?

(b) Mikä on todennäköisyys, että minuutissa tulee *korkeintaan* 4 asiakasta?

(c) Mikä on odotettavissa olevien asiakkaiden lukumäärä minuutin aikana?

Esimerkki 4.5 – Mitä opimme?

Esimerkissä 4.5 sovelletaan *Poisson-jakaumaa*.

Esimerkki 4.5 – Ratkaisu

Oletetaan, että minuutin aikana jonoon tulevien asiakkaiden lukumäärä X on *diskreetti satunnaismuuttuja*, joka noudattaa *Poisson-jakaumaa*:

$X \sim \mathrm{Poisson}(\lambda)$

jossa

λ = minuutissa jonoon tulevien asiakkaiden keskimääräinen lukumäärä = 4.

Siten satunnaismuuttujan X *pistetodennäköisyysfunktio* on

$$f(k) = \Pr(X = k) = \frac{e^{-4} 4^k}{k!}, \; x = 0, 1, 2, \ldots$$

(a) Todennäköisyys, että minuutissa ei tule *yhtään* asiakasta, on

$$\Pr(X = 0) = \frac{e^{-4}(4)^0}{0!} = 0.018.$$

(b) Todennäköisyys, että minuutissa tulee *korkeintaan* 4 asiakasta, on

$$
\begin{aligned}
\Pr(X \le 4) &= \sum_{x=0}^{4} \Pr(X = x) \\
&= \sum_{x=0}^{4} \frac{e^{-\lambda}(\lambda)^x}{x!} \\
&= e^{-4}\left(\frac{4^0}{0!} + \frac{4^1}{1!} + \frac{4^2}{2!} + \frac{4^3}{3!} + \frac{4^4}{4!}\right) \\
&= 0.018316 \times (1 + 4 + 8 + 10.667 + 10.667) \\
&= 0.018316 \times 34.333 \\
&= 0.62884
\end{aligned}
$$

(c) Koska Poisson-jakauman odotusarvo sama kuin sen parametri, kysytty odotusarvo on

$$\mathrm{E}(X) = \lambda = 4.$$

Esimerkki 4.7

Binomijakauman pistetodennäköisyysfunktio on

$$f(x) = \Pr(X = x) = \binom{n}{x} p^x q^{n-x}\ , 0 < p < 1\ , q = 1 - p\ , x = 0, 1, 2, \ldots, n$$

(a) Todista, että

$$\sum_{x=0}^{n} f(x) = 1$$

(b) Todista, että

$$\mathrm{E}(X) = np$$

Esimerkki 4.7 – Mitä opimme?

Esimerkissä 4.7 tarkastellaan *binomijakauman* ominaisuuksia.

Esimerkki 4.7 – Ratkaisu

(a) Suoraan *binomikaavan* mukaan

$$\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1^n = 1$$

(b) Suoraan laskemalla saadaan

$$
\begin{aligned}
\mathrm{E}(X) &= \sum_{x=0}^{n} x f(x) = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} \\
&= np
\end{aligned}
$$

missä viimeinen yhtälö perustuu siihen, että

$$\sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} = 1.$$

Tämä seuraa siitä, että summassa lasketaan yhteen *kaikki* binomijakauman

$$\mathrm{Bin}(n-1, p)$$

pistetodennäköisyydet

$$f(x) = \frac{(n-1)!}{x!(n-1-x)!} p^x (1-p)^{n-1-x}, \; x = 0, 1, 2, \ldots, n-1$$

Binomijakauman odotusarvo saadaan myös helposti johdetuksi käyttämällä hyväksi sitä, että *riippumattomien*, *samaa Bernoulli-jakaumaa noudattavien satunnaismuuttujien summa noudattaa binomijakaumaa*:

Olkoot $X_1, X_2, \ldots, X_n$ riippumattomia, samaa *Bernoulli-jakaumaa* $\mathrm{Ber}(p)$ noudattavia *diskreettejä satunnaismuuttujia*:

Tällöin *diskreetti satunnaismuuttuja*

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

noudattaa *binomijakaumaa* parametrein n ja p:

$$X \sim \mathrm{Bin}(n, p)$$

Odotusarvon ominaisuuksien perusteella

$$\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} p = np$$

koska Bernoulli-jakauman ominaisuuksien perusteella

$$\mathrm{E}(X_i) = p, \; i = 1, 2, \ldots, n$$

Esimerkki 5.1

Sähkölampun elinikä X (yksikkönä 1000 h) noudattaa jakaumaa, jonka tiheysfunktio on

$$f(x) = c/x^3, \; x \geq 1$$

missä c on vakio.

(a) Määrää vakion c arvo.

(b) Millä todennäköisyydellä lamppu kestää yli 5000 h?

(c) Mikä on lampun keskimääräinen elinikä?

(d) Määrää lampun eliniän *mediaani* eli määrää aika x, jolla $\Pr(X \le x) = 0.5$.

Esimerkki 5.1 – Mitä opimme?

Esimerkissä 5.1 tarkastellaan erään *jatkuvan jakauman* ominaisuuksia.

Esimerkki 5.1 – Ratkaisu

(a) Jatkuvan satunnaismuuttujan tiheysfunktio $f(x)$ toteuttaa aina ehdon

$$\int_{-\infty}^{+\infty} f(x)dx \equiv 1$$

Siten vakio c saadaan määrätyksi yhtälöstä

$$\int_{-\infty}^{+\infty} f(x)dx = c\int_{1}^{\infty} \frac{1}{x^3}\,dx = c \cdot \left[-\frac{1}{2} \times \frac{1}{x^2} \right]_1^{\infty} = c\left(\frac{1}{2} - 0 \right) = \frac{c}{2} = 1$$

joten

$$c = 2$$

(b) Tapahtuman

$$\{\text{Lampun elinikä } X > 5000 \text{ h}\}$$

todennäköisyys saadaan *integroimalla* satunnaismuuttujan X *tiheysfunktio* välillä $(5, \infty)$:

$$\Pr(X > 5) = \int_{5}^{\infty} f(x)dx = \int_{5}^{\infty} \frac{2}{x^3}\,dx = \left[-\frac{1}{x^2} \right]_5^{\infty} = \frac{1}{25} - 0 = \frac{1}{25} = 0.04$$

(c) Lampun *keskimääräinen elinikä* on lampun eliniän X *odotusarvo*

$$\mathrm{E}(X) = \int_{1}^{\infty} xf(x)dx = \int_{1}^{\infty} x \cdot \frac{2}{x^3}\,dx = \int_{1}^{\infty} \frac{2}{x^2}\,dx = \left[-\frac{2}{x} \right]_1^{\infty} = 2 - 0 = 2$$

Siten lampun keskimääräinen elinikä on tunteina 2000 h.

(d) Lampun eliniän *mediaani* saadaan ehdosta

$$\Pr(X \le x) = \int_{1}^{x} f(t)dt = \int_{1}^{x} \frac{2}{t^3}\,dt = \left[-\frac{1}{t^2} \right]_1^{x} = 1 - \frac{1}{x^2} = 0.5$$

joten mediaaniksi saadaan

$$x = \sqrt{2} \approx 1.412$$

Siten lampun eliniän mediaani on tunteina n. 1412 h.

Esimerkki 5.2

Eräässä laitteessa on komponentti, jonka elinikä X (yksikkönä kuukausi) noudattaa eksponenttijakaumaa parametrinaan 1/4.

(a) Mikä on komponentin keskimääräinen elinikä?

(b) Määrää komponentin eliniän *mediaani* eli määrää ikä x siten, että $\Pr(X \le x) = 0.5$.

(c) Määrää todennäköisyys, että komponentti kestää kauemmin kuin 6 kuukautta.

(d) Millä todennäköisyydellä komponentti toimii vielä vähintään yhden kuukauden, jos se on jo toiminut kuukauden?

(e) Millä todennäköisyydellä komponentti toimii vielä vähintään yhden kuukauden, jos se on jo toiminut kaksi kuukautta?

Esimerkki 5.2 – Mitä opimme?

Esimerkissä 5.2. sovelletaan *eksponenttijakaumaa*.

Esimerkki 5.2 – Ratkaisu

Tehtävän satunnaismuuttuja

$$X = \text{komponentin elinikä kuukausina}$$

noudattaa eksponenttijakaumaa parametrilla $\lambda = 1/4$:

$$X \sim \text{Exp}(1/4).$$

(a) Komponentin *keskimääräinen* elinikä on komponentin eliniän X *odotusarvo*

$$\text{E}(X) = 1/\lambda = 4 \text{ kuukautta.}$$

(b) Eksponenttijakautuneelle satunnaismuuttujalle X pätee

$$\Pr(X > x) = 1 - \text{P}(X \le x) = 1 - F(x) = \exp(-\lambda x)$$

jossa $F(x)$ on eksponenttijakauman *kertymäfunktio*. Siten

$$\Pr(X > x) = 0.5 \Leftrightarrow \exp(-\lambda x) = 0.5 \Leftrightarrow x = \log(2)/\lambda = 4\log(2) \approx 2.773$$

Satunnaismuuttujan X *mediaani* on siten n. 2.773 kuukautta.

(c) Kohdassa (b) mainitusta aputuloksesta seuraa, että

$$\Pr(X > 6) = \exp(-\,6\lambda) = \exp(-3/2) \approx 0.223$$

(d) ja (e)

Koska esponenttijakaumalla on ns. *unohtamisominaisuus*, kohdissa (d) ja (e) saadaan sama vastaus:

$$\begin{aligned}
&\Pr(\text{“Toimii vähintään vielä kuukauden”} \mid \text{”On toiminut jo } a \text{ kuukautta”}) \\
&\quad = \Pr(X > a + 1 \mid X > a) \\
&\quad = \Pr(X > a + 1)/\Pr(X > a) \\
&\quad = \exp(-\lambda(a + 1))/\exp(-\lambda a) \\
&\quad = \exp(-\lambda) \\
&\quad = \Pr(X > 1) \\
&\quad = \Pr(\text{“Toimii vähintään vielä kuukauden”})
\end{aligned}$$

Siten kohdassa (b) mainitusta aputuloksesta seuraa, että

$$\Pr(X > 1) = \exp(-\lambda) = \exp(-1/4) \approx 0.779$$

Esimerkki 5.3

Olkoon satunnaismuuttuja $Z \sim N(0,1)$.

(a) Määrää satunnaismuuttujan Z mediaani eli piste z siten, että $\Pr(Z \le z) = 0.5$.

(b) Määrää $\Pr(Z > 1.85)$.

(c) Määrää $\Pr(Z \le -1.85)$.

(d) Määrää z siten, että $\Pr(Z \le z) = 0.2$.

(e) Määrää z siten, että $\Pr(Z \ge z) = 0.8$.

(f) Määrää $\Pr(|Z| \le 2)$.

(g) Määrää z siten, että $\Pr(|Z| \ge z) = 0.1$.

Olkoon satunnaismuuttuja $X \sim N(-3,9)$.

(h) Määrää $\Pr(X \le -2)$.

(i) Määrää x siten, että $\Pr(X \ge x) = 0.05$.

Esimerkki 5.3 – Mitä opimme?

Esimerkissä 5.3. tarkastellaan *normaalijakaumaa* ja *todennäköisyyksien määräämistä normaalijakaumasta*.

Esimerkki 5.3 – Ratkaisu

(a) Koska satunnaismuuttujan Z jakauma on *symmetrinen* jakauman painopisteen 0 suhteen, niin

$$\Pr(Z \le 0) = 0.5 = \Pr(Z \ge 0)$$

Tämä näkyy myös standardoidun normaalijakauman taulukoissa.

(b) Vastakohdan todennäköisyyden kaavasta ja taulukoista nähdään, että

$$\Pr(Z > 1.85) = 1 - \Pr(Z \le 1.85) = 1 - 0.9678 = 0.0322$$

(c) Taulukoista:

$$\Pr(Z \le -1.85) = 0.0322$$

Tulos saadaan myös (b)-kohdasta, koska *standardoitu normaalijakauma* N(0,1) *on symmetrinen* painopisteensä 0 suhteen:

$$\Pr(Z \le -1.85) = \Pr(Z \ge 1.85) = \Pr(Z > 1.85) = 0.0322$$

(d) Taulukoista:

$$\Pr(Z \le z) = 0.2 \Rightarrow z \approx -0.84$$

(e) Todetaan ensin, että komplementtitapahtuman todennäköisyyden kaavasta seuraa, että

$$\Pr(Z \ge z) = 0.8 \Rightarrow \Pr(Z \le z) = 1 - \Pr(Z \ge z) = 1 - 0.8 = 0.2$$

Siten (d)-kohdan mukaan

$$\Pr(Z \le z) = 0.2 \Rightarrow z \approx -0.84$$

(f) Todetaan ensin, että standardoidun normaalijakauman N(0,1) symmetrian takia

$$\begin{aligned}\Pr(|Z| \le 2) &= \Pr(-2 \le Z \le +2) \\ &= \Pr(Z \le +2) - \Pr(Z \le -2) \\ &= \Pr(Z \le +2) - (1 - \Pr(Z \le +2)) \\ &= 2\times\Pr(Z \le +2) - 1\end{aligned}$$

Taulukoista:

$$\Pr(Z \le +2) = 0.9772$$

Siten

$$\Pr(|Z| \le 2) = 2\times\Pr(Z \le +2) - 1 = 2\times 0.9772 - 1 = 0.9544$$

(g) Todetaan ensin, että standardoidun normaalijakauman N(0,1) symmetrian takia

$$\Pr(|Z| \ge z) = 2\times\Pr(Z \ge z)$$

Siten

$$\Pr(|Z| \ge z) = 2\times\Pr(Z \ge z) = 0.1 \Rightarrow \Pr(Z \ge z) = 0.05 \Rightarrow \Pr(Z \le z) = 0.95$$

Taulukoista:

$$z \approx 1.64$$

Olkoon satunnaismuuttuja $X \sim \mathrm{N}(-3,9)$, jolloin

$$\mathrm{E}(X) = \mu = -3$$

$$\mathrm{Var}(X) = \sigma^2 = 9$$

$$\sigma = 3$$

Tällöin standardoitu satunnaismuuttuja

$$Z = \frac{X-\mu}{\sigma} = \frac{X+3}{3} : \ \mathrm{N}(0,1)$$

ja

$$X = \sigma Z + \mu = 3Z - 3 : \ \mathrm{N}(-3,9)$$

(h) Todetaan ensin, että

$$\Pr(X \le -2) = \Pr\left(\frac{X+3}{3} \le \frac{-2+3}{3}\right) = \Pr(Z \le 1/3)$$

jossa

$$Z = \frac{X+3}{3} : \ \mathrm{N}(0,1)$$

Taulukoista:

$$\Pr(Z \le 1/3) = 0.6293 = \Pr(X \le -2)$$

(i) Taulukoista:

$$\Pr(Z \ge z) = 0.05 \Rightarrow \Pr(Z \le z) = 0.95 \Rightarrow z \approx 1.64$$

Siten

$$x = 3z - 3 \approx 3\times 1.64 - 3 = 1.92$$

Esimerkki 5.4

Olkoon satunnaismuuttuja $X \sim \mathrm{N}(4,4)$.

(a) Määrää $\mathrm{P}(X = 1)$.

(b) Määrää satunnaismuuttujan X mediaani eli x siten, että $\Pr(X \le x) = 0.5$.

(c) Määrää $\Pr(X \le 3)$.

(d) Määrää x siten, että $\Pr(X \le x) = 0.99$.

(e) Määrää x siten, että $\Pr(X \le x) = 0.01$.

(f) Määrää satunnaismuuttujan X odotusarvon μ suhteen symmetriset pisteet $\mu - x$ ja $\mu + x$ niin, että niiden ulkopuolelle jää todennäköisyysmassasta 5%.

Esimerkki 5.4 – Mitä opimme?

Esimerkissä 5.4. tarkastellaan *normaalijakaumaa* ja *todennäköisyyksien määräämistä normaalijakaumasta*.

Esimerkki 5.4 – Ratkaisu

Olkoon satunnaismuuttuja $X \sim \mathrm{N}(4,4)$, jolloin

$$\mathrm{E}(X) = \mu = 4$$

$\mathrm{Var}(X) = \sigma^2 = 4$

$\mathrm{D}(X) = \sigma = 2$

Tällöin standardoitu satunnaismuuttuja

$$Z = \frac{X-\mu}{\sigma} = \frac{X-4}{2} : \quad \mathrm{N}(0,1)$$

ja

$$X = \sigma Z + \mu = 2Z + 4 : \quad \mathrm{N}(4,4)$$

(a) Kaikilla jatkuvilla jakaumilla yhden pisteen todennäköisyys = 0. Siten

$\Pr(X = 1) = 0$

(b) Koska *normaalijakauma on symmetrinen painopisteensä* suhteen, niin

$\Pr(X \le 4) = 0.5 = \Pr(X \ge 4)$

(c) Todetaan ensin, että

$$\Pr(X \le 3) = \Pr\left(\frac{X-4}{2} \le \frac{3-4}{2}\right) = \Pr(Z \le -0.5)$$

jossa

$$Z = \frac{X-4}{2} : \quad \mathrm{N}(0,1)$$

Taulukoista:

$\Pr(Z \le -0.5) = 0.3085 = \Pr(X \le 3)$

(d) Taulukoista:

$\Pr(Z \le z) = 0.99 \Rightarrow z \approx 2.33$

Siten

$x = 2 \times z - 4 \approx 2 \times (2.33) + 4 = 8.66$

(e) Taulukoista:

$\Pr(Z \le z) = 0.01 \Rightarrow z \approx -2.33$

Siten

$x = 2z - 4 \approx 2 \times (-2.33) + 4 = -0.66$

Kommentti (d)- ja (e)-kohtiin:

Pisteet –0.66 ja 8.66 sijaitsevat symmetrisesti normaalijakauman N(4,4) odotusarvon 4 molemmilla puolilla. Piste –0.66 erottaa jakauman *vasemmalle* hännälle todennäköisyysmassan 0.01 ja piste 8.66 erottaa jakauman *oikealle* hännälle todennäköisyysmassan 0.01. Pisteiden –0.66 ja 8.66 *väliin* jäävä todennäköisyysmassa on kooltaan

$$1 - \Pr(X \leq -0.66) - \Pr(X \geq +8.66) = 1 - 0.01 - 0.01 = 0.98$$

(f) Todetaan ensin, että standardoidun normaalijakauman N(0,1) symmetrian takia

$$\Pr(|Z| \geq z) = 2\times\Pr(Z \geq z)$$

Siten

$$\Pr(|Z| \geq z) = 2\times\Pr(Z \geq z) = 0.05 \Rightarrow \Pr(Z \geq z) = 0.025$$

Koska

$$\Pr(Z \geq z) = 1 - \Pr(Z \leq z)$$

voimme ratkaista z:n yhtälöstä

$$\Pr(Z \leq z) = 1 - \Pr(Z \geq z) = 1 - 0.025 = 0.975$$

Taulukoista:

$$z \approx 1.96$$

Siten

$$\mu + x = \mu + \sigma z \approx 4 + 2\times1.96 = 7.92$$

$$\mu - x = \mu - \sigma z \approx 4 - 2\times1.96 = 0.08$$

Esimerkki 5.5

Tehtaalla on naulakone, jonka tekemien naulojen painot vaihtelevat satunnaisesti ja toisistaan riippumatta noudattaen normaalijakaumaa odotusarvonaan 10 g ja varianssinaan 0.05 g^2.

Naulat pakataan laatikoihin niin, että yhteen laatikkoon tulee aina 1000 naulaa. Valitaan satunnaisesti joukko naulalaatikoita ja punnitaan laatikoiden sisällöt. Mikä osuus laatikoista on sellaisia, joiden sisältö painaa vähemmän kuin 9.99 kg?

Esimerkki 5.5 – Mitä opimme?

Esimerkissä 5.5. sovelletaan *normaalijakaumaa*.

Esimerkki 5.5 – Ratkaisu

Laatikkoon pakattujen naulojen painot ovat *riippumattomia* satunnaismuuttujia

$$X_i,\ i = 1, 2, \ldots, 1000,$$

jotka noudattavat *normaalijakaumaa* parametrein $\mu_X = 10$ g ja $\sigma_X^2 = 0.05$ g^2:

Naulalaatikon sisällön paino

$$Y = \sum_{i=1}^{1000} X_i$$

on normaalijakautunut satunnaismuuttuja, jonka parametrit ovat

$$\mu_Y = 1000 \times 10 \text{ g} = 10000 \text{ g}$$
$$\sigma_Y^2 = 1000 \times 0.05 \text{ g} = 50 \text{ g}$$

eli

$$Y : \ \mathrm{N}(10000, 50)$$

Siten

$$Z = \frac{Y - 10000}{\sqrt{50}} : \ \mathrm{N}(0,1)$$

ja

$$\Pr(Y \le 9990) = \Pr\left(\frac{Y - 10000}{\sqrt{50}} \le \frac{9990 - 10000}{\sqrt{50}}\right) = \Pr(Z \le -1.41) = 0.0793$$

Suurten lukujen lain mukaan

$$100 \times 0.0793 = 7.93 \ \%$$

naulalaatikoiden sisällöistä isossa otoksessa painaa alle 9.99 kg.

Esimerkki 5.6

Eräässä maassa miesten pituudet vaihtelevat satunnaisesti ja toisistaan riippumatta noudattaen normaalijakaumaa odotusarvonaan 180 cm ja varianssinaan 25 cm^2.

Kerätään joukko 200 miehen satunnaisotoksia ja määrätään jokaisesta otoksesta miesten pituuksien aritmeettinen keskiarvo. Mikä osuus otoksista määrätyistä keskiarvoista on välin (179.5, 180.5) ulkopuolella?

Esimerkki 5.6 – Mitä opimme?

Esimerkissä 5.6. sovelletaan *normaalijakaumaa*.

Esimerkki 5.6 – Ratkaisu

Miesten pituudet otoksessa ovat *riippumattomia* satunnaismuuttujia

$$X_i \, , \, i = 1, 2, \ldots , 200$$

jotka noudattavat *normaalijakaumaa* parametrein $\mu_X = 180$ cm ja $\sigma_X^2 = 25$ cm^2.

Otokseen poimittujen miesten pituuksien aritmeettinen keskiarvo

$$\bar{X} = \frac{1}{200} \sum_{i=1}^{200} X_i$$

on normaalijakautunut satunnaismuuttuja, jonka parametrit ovat

$$\mu_Y = 180 \text{ cm}$$
$$\sigma_Y^2 = \frac{25}{200} \text{ cm}^2 = 0.125 \text{ cm}^2$$

eli

$$\bar{X} : \ \mathrm{N}(180, 0.125)$$

Tällöin

$$Z = \frac{\bar{X} - 180}{\sqrt{0.125}} : \ \mathrm{N}(0,1)$$

Edellä esitetystä seuraa, että

$$
\begin{aligned}
&\Pr(\bar{X} \le 179.5 \text{ tai } \bar{X} \ge 180.5) \\
&\quad = 1 - \Pr(179.5 \le \bar{X} \le 180.5) \\
&\quad = 1 - \Pr\left(\frac{179.5-180}{\sqrt{0.125}} \le \frac{\bar{X}-180}{\sqrt{0.125}} \le \frac{180.5-180}{\sqrt{0.125}}\right) \\
&\quad = 1 - \Pr(-1.41 \le Z \le +1.41) \\
&\quad = 1 - \{\Pr(Z \le +1.41) - \Pr(Z \le -1.41)\} \\
&\quad = 1 - \{\Pr(Z \le +1.41) - [1 - \Pr(Z \le +1.41)]\} \\
&\quad = 2 \times \{1 - \Pr(Z \le +1.41)\} \\
&\quad = 2 \times (1 - 0.9207) \\
&\quad = 2 \times 0.0793 \\
&\quad = 0.1586
\end{aligned}
$$

Todennäköisyyden *frekvenssitulkinnan* mukaan

$$100 \times 0.1586 = 15.86\ \%$$

otoksien aritmeettisista keskiarvoista on välin (179.5,180.5) ulkopuolella.

Esimerkissä 5.7 tarkastellaan *keskeisen raja-arvolauseen* soveltamista binomi- ja Poisson-todennäköisyyksien laskemiseen.

Esimerkki 5.7

Heität virheetöntä noppaa 60000 kertaa.

(a) Mikä on odotettavissa oleva kuutosten lukumäärä?

(b) Mikä on arviolta todennäköisyys, että kuutosten lukumäärä on suljetulla välillä [9900, 10150]?

Ohje: Käytä (b)-kohdassa keskeiseen raja-arvolauseeseen perustuvaa normaalijakauma-approksimaatiota.

Esimerkki 5.7 – Mitä opimme?

Esimerkissä 5.7 tarkastellaan *keskeisen raja-arvolauseen soveltamista binomitodennäköisyyksien approksoimiseen*.

Esimerkki 5.7 – Ratkaisu

(a) Kuutosten lukumäärä X on satunnaismuuttuja, joka noudattaa binomijakaumaa Bin(n, p), jossa

$$n = 60000$$

$p = 1/6$

Siten *odotettavissa oleva kuutosten lukumäärä* on

$\mathrm{E}(X) = np = 10000$

(b) *Keskeisen raja-arvolauseen* mukaan standardoitu satunnaismuuttuja

$$Z = \frac{X - \mathrm{E}(X)}{\mathrm{D}(X)} \approx \mathrm{N}(0,1)$$

missä

$\mathrm{E}(X) = np = 10000$

$\mathrm{D}^2(X) = \mathrm{Var}(X) = np(1-p) = 8333.33$

$\mathrm{D}(X) = 91.2871$

Standardoimalla satunnaismuuttuja X ja käyttämällä standardoidun normaalijakauman N(0,1) taulukoita saadaan kysytyn todennäköisyyden approksimaatioksi:

$$\begin{aligned}
&\Pr(9900 \le X \le 10150) \\
&= \Pr\left(\frac{9900-10000}{91.2871} \le \frac{X-10000}{91.2871} \le \frac{10150-10000}{91.2871}\right) \\
&\approx \Pr(-1.10 \le Z \le 1.64) \\
&= \Pr(Z \le 1.64) - \Pr(Z \le -1.10) \\
&= 0.9495 - 0.1357 \\
&= 0.8138.
\end{aligned}$$