TODENNÄKÖISYYSLASKUN KERTAUS

Peruskäsitteitä

Otosavaruus $\mathcal{S}$

$\mathcal{S}$ on satunnaiskokeen E kaikkien mahdollisten alkeistapahtumien e joukko.

Esim. 1. Noppaa heitettäessä $\mathcal{S}$ on $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$.

Esim. 2. Hehkulampun kestoikä $\mathcal{S} = \{x \in \mathcal{R} \mid x > 0\}$.

Tapahtuma

Satunnaiskokeessa tapahtuma on otosavaruuden osajoukko. Tapahtumaa merkitään yleensä isolla kirjaimella $A,\ B, \ldots$

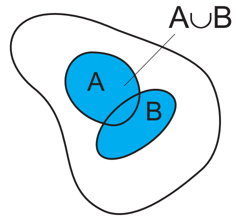
Esim. 1. Nopanheitossa saadan parillinen luku: $A = \{2, 4, 6\} \subset \mathcal{S}$.

Esim. 2. Hehkulampun kestoikä on vähintään 3000 h: $A = \{x \in \mathcal{R} \mid x > 3000\} \subset \mathcal{S}$.

<u>Varma tapahtuma</u>: Otosavaruus $\mathcal{S}$ itse.

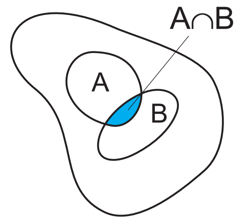
<u>Mahdoton tapahtuma</u>: Otosavaruuden tyhjä osajoukko ϕ.

Tapahtumien yhdistely



Yhdiste (unioni) “A tai B”.

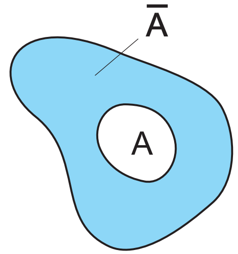
$$A \cup B = \{e \in \mathcal{S} \mid e \in A \text{ tai } e \in B\}$$



Leikkaus “A ja B”.

$$A \cap B = \{e \in \mathcal{S} \mid e \in A \text{ ja } e \in B\}$$

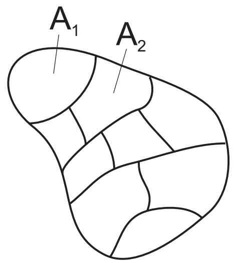
Tapahtumat A ja B ovat toisensa poissulkevat, jos $A \cap B = \phi$.



Komplementti “ei A”.

$$\bar{A} = \{e \in \mathcal{S} \mid e \notin A\}$$

Otosavaruuden ositus



Tapahtumajoukko $A_1, A_2, \ldots$ muodostaa otosavaruuden $\mathcal{S}$ osituksen, kun

1. Tapahtumat ovat pareittain toisensa poissulkevia, $A_i \cap A_j = \phi$, kun $i \neq j$.
2. Ne yhdessä muodostavat koko otosavaruuden, $\cup_i A_i = \mathcal{S}$.

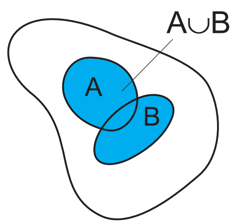
Todennäköisyys

Kuhunkin tapahtumaan A liitetään todennäköisyys $\mathrm{P}\{A\}$.

Empiirisesti todennäköisyys $\mathrm{P}\{A\}$ tarkoittaa A:n suhteellisen esiintymisfrekvenssin $N(A)/N$ raja-arvoa toistokokeessa

$$\mathrm{P}\{A\} = \lim_{N\to\infty} N(A)/N \qquad \begin{cases} N & = \text{toistojen lukumäärä} \\ N(A) & = A\text{:n esiintymiskertojen lukumäärä} \end{cases}$$

Todennäköisyyden ominaisuuksia



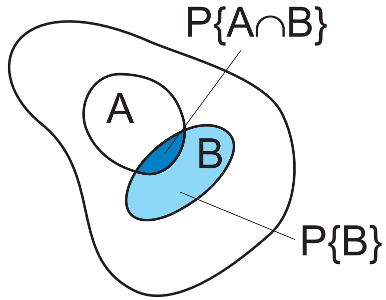
1. $0 \leq \mathrm{P}\{A\} \leq 1$
2. $\mathrm{P}\{\mathcal{S}\} = 1 \qquad \mathrm{P}\{\phi\} = 0$
3. $\mathrm{P}\{A \cup B\} = \mathrm{P}\{A\} + \mathrm{P}\{B\} - \mathrm{P}\{A \cap B\}$
4. Jos $A \cap B = 0$, niin $\mathrm{P}\{A \cup B\} = \mathrm{P}\{A\} + \mathrm{P}\{B\}$

   Jos $A_i \cap A_j = 0$, kun $i \neq j$, niin $\mathrm{P}\{\cup_i A_i\} = \mathrm{P}\{A_1 \cup \ldots \cup A_n\} = \mathrm{P}\{A_1\} + \ldots \mathrm{P}\{A_n\}$
5. $\mathrm{P}\{\bar{A}\} = 1 - \mathrm{P}\{A\}$
6. Jos $A \subseteq B$ niin $\mathrm{P}\{A\} \leq \mathrm{P}\{B\}$

Ehdollinen todennäköisyys

Tapahtuman A todennäköisyys ehdolla, että B on jo tapahtunut.

$$\boxed{\mathrm{P}\{A\,|\,B\} = \frac{\mathrm{P}\{A \cap B\}}{\mathrm{P}\{B\}}} \qquad \Rightarrow \mathrm{P}\{A \cap B\} = \mathrm{P}\{A\,|\,B\}\mathrm{P}\{B\}$$

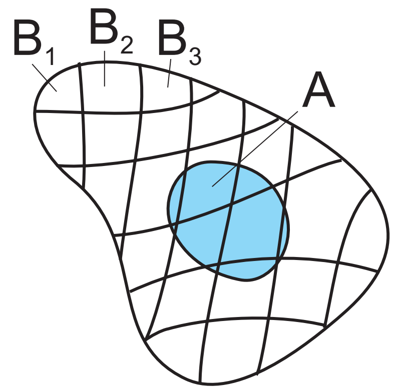


Kokonaistodennäköisyyden kaava

Olkoon $\{B_1, \ldots, B_n\}$ täydellinen joukko toisensa poissulkevia tapahtumia eli ns. otosavaruuden $\mathcal{S}$ ositus,

1. $\cup_i B_i \;=\; \mathcal{S}$ varma tapahtuma $\mathrm{P}\{\cup_i B_i\} \;=\; 1$
2. $B_i \cap B_j \;=\; \phi$ kun $i \neq j$ $\mathrm{P}\{B_i \cap B_j\} \;=\; 0$

Tällöin $A = A \cap \mathcal{S} = A \cap (\cup_i B_i) = \cup_i (A \cap B_i)$ ja



$$\boxed{\mathrm{P}\{A\} = \sum_{i=1}^{n} \mathrm{P}\{A \cap B_i\} = \sum_{i=1}^{n} \mathrm{P}\{A \,|\, B_i\}\mathrm{P}\{B_i\}}$$

A:n todennäköisyyden laskenta ehdollistamalla tapahtumiin B_i. Tyypillisesti tapahtumat B_i edustavat jonkun kokeen kaikkien mahdollisten tulostapahtumien joukkoa.

Bayesin kaava

Olkoon jälleen $\{B_1, \ldots, B_n\}$ otosavaruuden ositus.

Kysytään, mikä on tapahtuman B_i todennäköisyys, kun tiedetään, että A on tapahtunut.

$$\boxed{\mathrm{P}\{B_i \,|\, A\} = \frac{\mathrm{P}\{A \cap B_i\}}{\mathrm{P}\{A\}} = \frac{\mathrm{P}\{A \,|\, B_i\}\mathrm{P}\{B_i\}}{\sum\limits_j \mathrm{P}\{A \,|\, B_j\}\mathrm{P}\{B_j\}}}$$

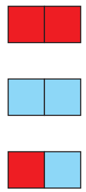
Bayesin kaava mahdollistaa ehdollisen todennäköisyyden laskemisen, kun tunnetaan käänteiset ehdolliset todennäköisyydet.

<u>Esimerkki:</u> kolme korttia, joiden eri puolet ovat erivärisiä.

pp: molemmat puolet punaisia

ss: molemmat puolet sinisiä

ps: toinen puoli punainen, toinen sininen



Satunnaisesti valitun kortin yläpuoli on punainen. Mikä on todennäköisyys sille, että toinen puoli on sininen?

$$\begin{aligned}\mathrm{P}\{\mathrm{ps} \,|\, \mathrm{pun}\} &= \frac{\mathrm{P}\{\mathrm{pun} \,|\, \mathrm{ps}\}\mathrm{P}\{\mathrm{ps}\}}{\mathrm{P}\{\mathrm{pun} \,|\, \mathrm{pp}\}\mathrm{P}\{\mathrm{pp}\} + \mathrm{P}\{\mathrm{pun} \,|\, \mathrm{ss}\}\mathrm{P}\{\mathrm{ss}\} + \mathrm{P}\{\mathrm{pun} \,|\, \mathrm{ps}\}\mathrm{P}\{\mathrm{ps}\}} \\ &= \frac{\frac{1}{2}\times\frac{1}{3}}{1\times\frac{1}{3}+0\times\frac{1}{3}+\frac{1}{2}\times\frac{1}{3}} = \frac{1}{3}\end{aligned}$$

Tilastollinen riippumattomuus

Tapahtumat A ja B ovat riippumattomia silloin ja vain silloin kun

$$\mathrm{P}\{A \cap B\} = \mathrm{P}\{A\} \cdot \mathrm{P}\{B\}$$

Riippumattomille tapahtumille pätee

$$\mathrm{P}\{A \,|\, B\} = \frac{\mathrm{P}\{A \cap B\}}{\mathrm{P}\{B\}} = \frac{\mathrm{P}\{A\}\mathrm{P}\{B\}}{\mathrm{P}\{B\}} = \mathrm{P}\{A\}$$

"B:llä ei ole vaikutusta A:n esiintymiseen".

Esim. 1: Nopanheitto kahdella nopalla, $A = \{n_1 = 6\},\ B = \{n_2 = 1\}$

$A \cap B = \{(6,1)\}$, $\quad \mathrm{P}\{A \cap B\} = \frac{1}{36}$, kaikki yhdistelmät yhtä todennäköisiä

$\mathrm{P}\{A\} = \mathrm{P}\{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\} = \frac{6}{36} = \frac{1}{6}$; samoin $\mathrm{P}\{B\} = \frac{1}{6}$

$\mathrm{P}\{A\}\mathrm{P}\{B\} = \frac{1}{36} = \mathrm{P}\{A \cap B\} \quad \Rightarrow$ ovat riippumatomia

Esim. 2: $A = \{n_1 = 6\},\ B = \{n_1 + n_2 = 9\} = \{(3,6),(4,5),(5,4),(6,3)\}$

$A \cap B = \{(6,2)\}$

$\mathrm{P}\{A\} = \frac{1}{6}$, $\quad \mathrm{P}\{B\} = \frac{4}{36}$, $\quad \mathrm{P}\{A \cap B\} = \frac{1}{36}$

$\mathrm{P}\{A\} \cdot \mathrm{P}\{B\} \neq \mathrm{P}\{A \cap B\} \quad \Rightarrow \quad A$ ja B riippuvia

Todennäköisyyslaskenta: yhteenveto

- Tärkeä reaalimaailman ilmiöiden mallintamisessa
  - esim. tietoliikennejärjestelmät
- Todennäköisyysteorialla on luonnollinen, intuitiivinen tulkinta ja yksinkertaiset matemaattiset aksioomat
- Kokonaistodennäköisyyden kaava mahdollistaa ongelman jaon osiin
  - analyyttinen lähestymistapa
  - keskeinen väline stokastisessa mallinnuksessa
- Riippumattomien tapahtumien yhteistapahtuman todennäköisyys on tulo yksittäisten tapahtumien todennäköisyyksistä

Satunnaismuuttujat ja jakaumat

Satunnaismuuttuja

Usein meitä kiinnostaa enemmän kuin kokeessa esiintyvä tapahtuma itsessään jokin tulokseen liittyvä arvo.

Esim. 1. Klaavojen lukumäärä rahanheittosarjassa pikemmin kuin kruuna/klaava-sekvenssi itsessään

Reaaliarvoinen satunnaismuuttuja X on kuvaus
$X: \quad \mathcal{S} \mapsto \mathcal{R}$
joka liittää reaaliluvun $X(e)$ jokaiseen alkeistapahtumaan $e \in \mathcal{S}$.

Esim. 2. Klaavojen lkm kolmen pituisissa heittosarjoissa (klaava = `h` (head), kruuna=`t` (tail))

e	$X(e)$
`hhh`	3
`hht`	2
`hth`	2
`htt`	1
`thh`	2
`tht`	1
`tth`	1
`ttt`	0

- X:n arvot tulevat “arvotuiksi” sitä kautta, että e “arvotaan”
- e edustaa “arpalippua”, johon on kirjoitettu X:n arvo

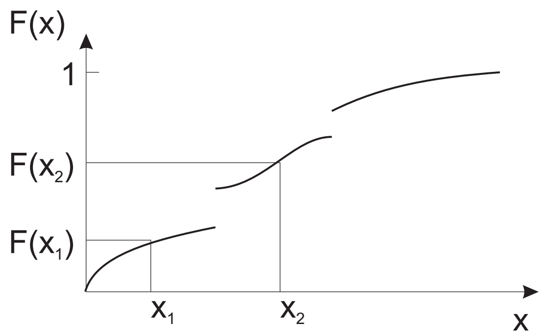
Satunnaismuuttujan X arvojoukko eli X:n otosavaruus

$$\boxed{\mathcal{S}_X = \{x \in \mathcal{R} \mid X(e) = x,\ e \in \mathcal{S}\}}$$

- voi olla äärellinen tai numeroituvasti ääretön: diskreetti jakauma
- ylinumeroituvasti ääretön: jatkuva jakauma

Kertymäfunktio (jakaumafunktio) (cdf, cumulative distribution function)

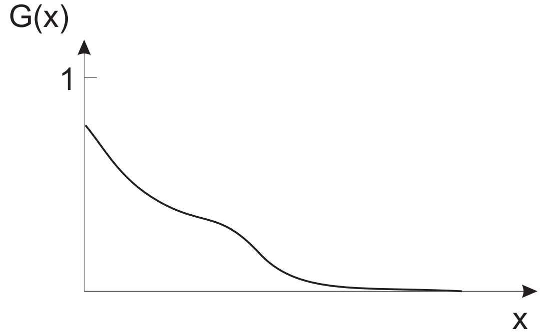
$$\boxed{F(x) = \mathrm{P}\{X \le x\}}$$



Välin todennäköisyys

$$\boxed{\mathrm{P}\{x_1 \le X \le x_2\} = F(x_2) - F(x_1)}$$

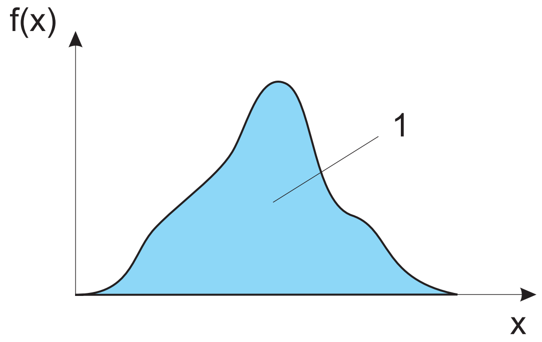
Komplementaarinen jakaumafunktio (häntäjakauma) (tail distribution)



$$\boxed{G(x) = 1 - F(x) = \mathrm{P}\{X > x\}}$$

Jatkuvan sm:n tiheysfunktio (pdf, probability density function)

$$\boxed{f(x) = \frac{dF(x)}{dx} = \lim_{dx \to 0} \frac{\mathrm{P}\{x < X \leq x + dx\}}{dx}}$$



Diskreetti satunnaismuuttuja

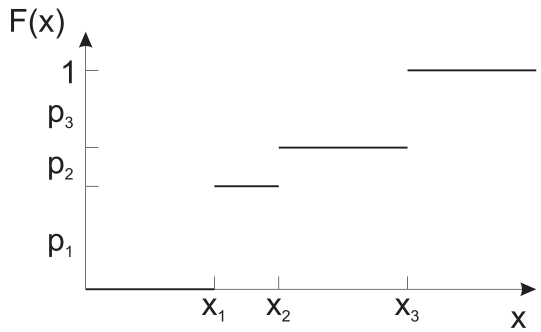
Diskreetin satunnaismuuttujan X saamien arvojen joukko on äärellinen tai numeroituvasti ääretön, $X \in \{x_1, x_2, \ldots\}$.

Näihin arvoihin liittyvät pistetodennäköisyydet

$$p_i = \mathrm{P}\{X = x_i\}$$

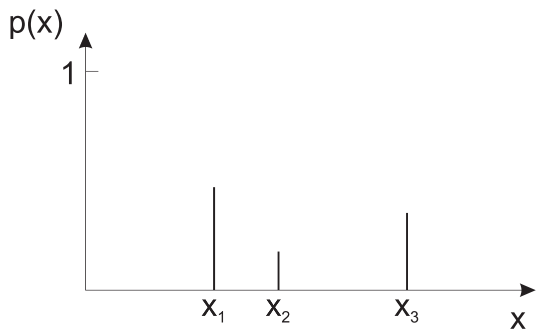
jotka määräävät diskreetin jakauman.

Kertymäfunktio on porrasfunktio, jolla on p_i:n suuruiset hyppäykset kohdissa x_i.



Pistetodennäköisyysfunktio eli tn-massafunktio (pmf, probability mass function)

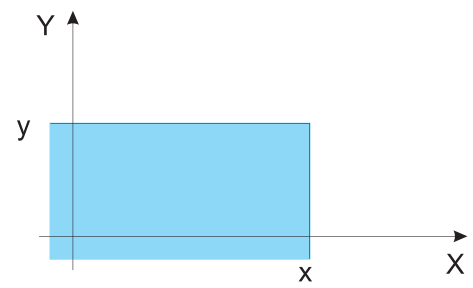
$$p(x) = \mathrm{P}\{X = x\} = \begin{cases} p_i & \text{kun } x = x_i \\ 0, & \text{muuten} \end{cases}$$



Usean muuttujan jakaumat

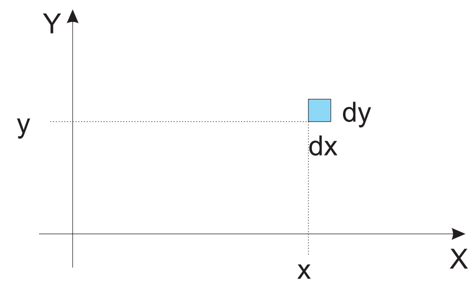
Yhteiskertymäfunktio (joint cdf)

$$\boxed{F_{X,Y}(x,y) = \mathrm{P}\{X \le x, Y \le y\}}$$



Yhteisjakauman tiheysfunktio (joint pdf)

$$\boxed{f_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{X,Y}(x,y)}$$



Ylläolevat yleistyvät luonnollisella tavalla useammalle satunnaismuuttujalle.

Riippumattomuus

Satunnaismuuttujat X ja Y ovat riippumattomia silloin ja vain silloin, kun tapahtumat $\{X \le x\}$ ja $\{Y \le y\}$ ovat riippumattomia, jolloin

$$\boxed{F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)}$$

$$\boxed{f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)}$$

(ehdot ovat ekvivalentteja)

Satunnaismuuttujan funktio

Olkoon X (reaaliarvoinen) satunnaismuuttuja ja $g(\cdot)$ funktio ($g\colon\ \mathcal{R} \mapsto \mathcal{R}$). Soveltamalla funktiota g satunnaismuuttujan X arvoihin, saadaan toinen satunnaismuuttuja $Y = g(X)$.

$\boxed{F_Y(y) = F_X(g^{-1}(y))}$ sillä $Y \le y \;\Leftrightarrow\; g(X) \le y \;\Leftrightarrow\; X \le g^{-1}(y)$

Jos erityisesti valitaan $g(\cdot) = F_X(\cdot)$ (arvoalue [0,1]), niin

$$F_Y(y) = F_X(F_X^{-1}(y)) = y$$

ja Y:n tiheysfunktio $f_Y(y) = \frac{d}{dy}F_Y(y) = 1$ eli Y noudattaa tasaista jakaumaa (välillä (0,1)).

$\boxed{F_X(X) \sim U}$ $\boxed{X \sim F_X^{-1}(U)}$ $\sim$ tarkoittaa “samoin jakautunut”

Tämä mahdollistaa mielivaltaisen satunnaismuuttujan X (jakaumafunktio $F_X(x)$) arvojen generoinnin esim. simuloinneissa, kun käytettävissä on satunnaislukugeneraattori, joka tuottaa välille (0,1) tasanjakautuneita satunnaismuuttujan U arvoja.

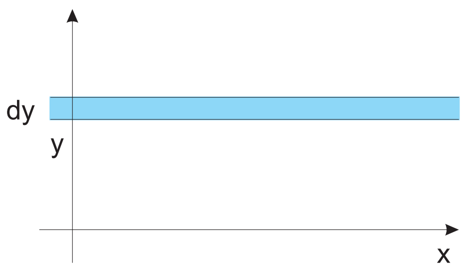
Ehdollisen jakauman tiheysfunktio

Olkoot X ja Y kaksi satunnaismuuttujaa (yleensä toisistaan riippuvia). Voidaan tarkastella satunnaismuuttujaa X ehdollistettuna siihen, että Y on saanut jonkun tietyn arvon y. Merkitään tätä ehdollistettua satunnaismuuttujaa $X_{|Y=y}$:llä.

Tämän tiheysfunktiolle käytetään merkintää $f_{X_{|Y=y}} = f_{X|Y}(x,y)$ ja sille pätee

$$f_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

nimittäjässä Y:n reunajakauma on $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$



jakauma on rajoitettu kaistaleeseen $Y \in (y, y+dy)$
$f_{X,Y}(x,y)dydx$ on tietyn kaistaleessa olevan $dxdy$-elementin todennäköisyys
$f_Y(y)dy$ on koko kaistaleen todennäköisyysmassa

Muuttujien X ja Y ollessa riippumattomia, on $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ ja $f_{X|Y}(x,y) = f_X(x)$, eli ehdollistaminen ei vaikuta jakaumaan, kuten ei pidäkään.

Jakaumien tunnusluvut

Odotusarvo

Merkitään $\mathrm{E}[X] = \bar{X}$

Jatkuva jakauma: $\boxed{\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx}$

Diskreetti jakauma: $\boxed{\mathrm{E}[X] = \sum_i x_i p_i}$

Yleisesti: $\boxed{\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, dF(x)}$ $dF(x)$ on välin dx todennäköisyys

Odotusarvon ominaisuuksia

$\mathrm{E}[cX] = c\mathrm{E}[X]$	c vakio
$\mathrm{E}[X_1 + \cdots X_n] = \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]$	aina
$\mathrm{E}[X \cdot Y] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$	ainoastaan, kun X ja Y riippumattomia

Varianssi

Merkitään $\mathrm{V}[X]$ (myös $\mathrm{Var}[X]$)

$$\boxed{\mathrm{V}[X] = \mathrm{E}[(X - \bar{X})^2]} = \mathrm{E}[X^2] - \mathrm{E}[X]^2$$

Kovarianssi

Merkitään $\mathrm{Cov}[X, Y]$

$$\boxed{\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \bar{X})(Y - \bar{Y})]} = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$$

$\mathrm{Cov}[X, X] = \mathrm{V}[X]$

Jos X ja Y riippumattomia, niin $\mathrm{Cov}[X, Y] = 0$

Varianssin ominaisuuksia

$\mathrm{V}[cX] = c^2\mathrm{V}[X]$	c vakio; huom. toinen potenssi
$\mathrm{V}[X_1 + \cdots X_n] = \sum_{i,j=1}^{n} \mathrm{Cov}[X_i, X_j]$	aina
$\mathrm{V}[X_1 + \cdots X_n] = \mathrm{V}[X_1] + \cdots + \mathrm{V}[X_n]$	<u>ainoastaan, kun X_i:t ovat riippumattomia</u>

Kovarianssin ominaisuuksia

$$\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X]$$

$$\mathrm{Cov}[X + Y, Z] = \mathrm{Cov}[X, Z] + \mathrm{Cov}[Y, Z]$$

Ehdollinen odotusarvo

Satunnaismuuttujan X odotusarvo ehdolla, että toinen sm Y saa arvon $Y = y$ on

$$\boxed{\mathrm{E}\,[X\,|\,Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x, y) dx}$$

eli se saadaan käyttämällä X:n ehdollista jakaumaa.

$\mathrm{E}\,[X\,|\,Y = y]$ on y:n funktio. Laskemalla tämän funktion arvo sm:n Y määräämässä pisteessä saadaan satunnaismuuttuja $\mathrm{E}\,[X\,|\,Y]$ (sm:n Y funktio).

<u>Ehdollisen odotusarvon ominaisuuksia</u>

$\mathrm{E}\,[X\,|\,Y] = \mathrm{E}[X]$	<u>jos X ja Y ovat riippumattomia</u>
$\mathrm{E}\,[c\,X\,|\,Y] = c\,\mathrm{E}\,[X\,|\,Y]$	c on vakio
$\mathrm{E}\,[X + Y\,|\,Z] = \mathrm{E}\,[X\,|\,Z] + \mathrm{E}\,[Y\,|\,Z]$	

Ehdollinen varianssi

$$\boxed{\mathrm{V}[X\,|\,Y] = \mathrm{E}\left[(X - \mathrm{E}\,[X\,|\,Y])^2\,|\,Y\right]}$$

Huom. poikkema ehdollisen odotusarvon suhteen

Ehdollinen kovarianssi

$$\boxed{\mathrm{Cov}[\,X,Y\,|\,Z] = \mathrm{E}\,[(X - \mathrm{E}\,[X\,|\,Z])(Y - \mathrm{E}\,[Y\,|\,Z])\,|\,Z]}$$

Ketjusäännöt (ehdollistamissäännöt)

$$\mathrm{E}[X] = \mathrm{E}[\mathrm{E}\,[X\,|\,Y]] \quad \text{(sisempi, ehdollinen odotusarvo on sm:n } Y \text{ funktio)}$$

$$\mathrm{V}[X] = \mathrm{E}[\mathrm{V}[X\,|\,Y]] + \mathrm{V}[\mathrm{E}\,[X\,|\,Y]]$$

$$\mathrm{Cov}[X,Y] = \mathrm{E}[\mathrm{Cov}[\,X,Y\,|\,Z]] + \mathrm{Cov}[\mathrm{E}\,[X\,|\,Z]\,,\mathrm{E}\,[Y\,|\,Z]]$$

Riippumattomien sm:ien maksimin ja minimin jakaumat

Olkoot $X_1, \ldots, X_n$ riippumattomia satunnaismuuttujia
(kertymäfunktiot $F_i(x)$ ja häntäjakaumat $G_i(x),\ i = 1, \ldots, n$)

Maksimin jakauma

$$\begin{aligned} \mathrm{P}\{\max(X_1, \ldots, X_n) \le x\} &= \mathrm{P}\{X_1 \le x, \ldots, X_n \le x\} \\ &= \mathrm{P}\{X_1 \le x\} \cdots \mathrm{P}\{X_n \le x\} \qquad \text{(riippumattomuus!)} \\ &= F_1(x) \cdots F_n(x) \end{aligned}$$

Minimin jakauma

$$\begin{aligned} \mathrm{P}\{\min(X_1, \ldots, X_n) > x\} &= \mathrm{P}\{X_1 > x, \ldots, X_n > x\} \\ &= \mathrm{P}\{X_1 > x\} \cdots \mathrm{P}\{X_n > x\} \qquad \text{(riippumattomuus!)} \\ &= G_1(x) \cdots G_n(x) \end{aligned}$$