Jos satunnaismuuttujat X ja Y ovat riippumattomia, niin X:n ja Y:n reunajakaumat antavat kaiken tiedon satunnaismuuttujaparin (X, Y) jakautumisesta, esimerkiksi kertymäfunktioille pätee

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$

Kun yleisessä tapauksessa haluamme antaa tyhjentävän kuvan X:n ja Y:n välisestä vuorovaikutuksesta, ”riippuvuudesta”, käytämme X:n ja Y:n *yhteisjakaumaa* eli satunnaismuuttujaparin (X, Y) jakaumaa. Tästä käsin voimme sitten kartoittaa satunnaismuuttujien X ja Y vuorovaikutusta tutkimalla esimerkiksi ehdollisia jakaumia.

Riippumattomuuden ”vastakohtana” – toisena äärimmäisyytenä kahden satunnaismuuttujan vuorovaikutuksesta – on *funktionaalinen riippuvuus*; toisen satunnaismuuttujan arvo määrää täysin myös toisen arvon. Erityisen vahva funktionaalinen riippuvuus on *lineaarinen riippuvuus*, so. tapaus, jossa on olemassa reaaliluvut a, b ja c ($a^2 + b^2 > 0$) siten, että

$$P(aX + bY + c = 0) = 1.$$

Lineaarinen riippuvuus edustaa *singulaarista jakaumaa* (tämä tarkoittaa, että todennäköisyys on keskittynyt johonkin tason nollamittaiseen joukkoon; diskreetti jakauma on erikoistapaus singulaarisesta jakaumasta, tällöinhän kaikki todennäköisyys ”sijaitsee” numeroituvassa joukossa), jossa kaikki todennäköisyys on keskittynyt suoralle

$$A = \{(x, y) \in \mathbb{R}^2 : ax + by + c = 0\}.$$

Yleisessä tapauksessa satunnaismuuttujien välillä ei tietenkään vallitse riippumattomuus eikä funktionaalinen riippuvuus: toisen muuttujan arvon tunteminen vaikuttaa toisen jakautumiseen, mutta ei määrää sen arvoa yksikäsitteisesti. Tällaista riippuvuutta on tapana kutsua *stokastiseksi riippuvuudeksi*.

Esimerkki. Mahdollisimman yksinkertaisena esimerkkinä tarkastelemme symmetrisen nopan heittoa. Jos

$$\begin{aligned}
X &= \text{”1. nopan pisteluku”},\\
Y &= \text{”2. nopan pisteluku”},\\
Z &= X + Y,\\
U &= \text{”parittomien pistelukujen lukumäärä”},\\
V &= \text{”parillisten pistelukujen lukumäärä”},
\end{aligned}$$

niin

X ja Y ovat riippumattomia,

X:n ja Z:n välillä on stokastinen riippuvuus

(eivät riippumattomia, eivätkä funktionaalisesti riippuvia),

Z:n ja U:n välillä vallitsee stokastinen riippuvuus,

U:n ja V:n välillä on funktionaalinen riippuvuus

(vieläpä lineaarinen), sillä $U + V = 2$.

Korrelaatiokertoimen tarkoituksena on saada tiivistetyksi tietoa muuttujien välisestä riippuvuudesta yhteen ainoaan tunnuslukuun. Tarkastelemmekin seuraavaksi yhden lähestymistavan avulla, miten hyvin tämä onnistuu.

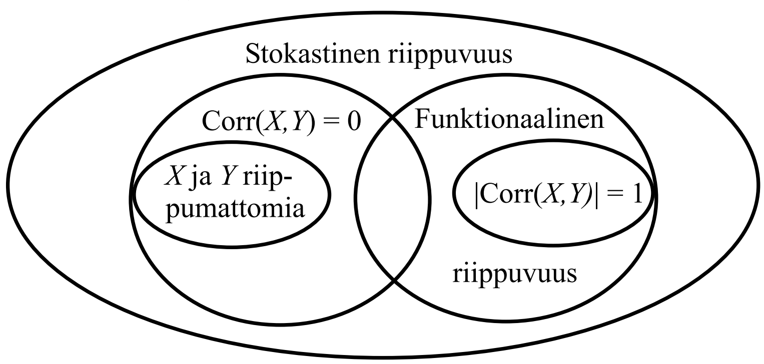
Korrelaatiokertoimen väärinkäytöstä

Korrelaatiokertoimeen liittyy suuri houkutus vetää liian suuria johtopäätöksiä muuttujien välisestä riippuvuudesta. Korostettakoon siksi vielä, että ainoat matemaattisesti oikeutetut johtopäätökset ovat:

(i) $\mathrm{Corr}(X, Y) \neq 0 \;\Rightarrow\; X$ ja Y eivät ole riippumattomia,

(ii) $|\mathrm{Corr}(X, Y)| = 1 \;\Leftrightarrow\; X$ ja Y ovat lineaarisesti riippuvia.

Pyrimme vielä havainnollistamaan satunnaismuuttujien välistä riippuvuutta seuraavalla kartalla (jota ei pidä ottaa liian vakavasti, se on tarkoitettu lähinnä muistin tueksi).



Esimerkki. Korrelaatiokertoimen soveltamisessa käytäntöön liittyy myös muita syitä olla erityisen varovainen. Valaisemme tätä esimerkillä: Tarkastelemme tiettyä aluetta ja merkitsemme

U = ”haikaroiden lukumäärä”,

V = ”syntyvien lasten lukumäärä”,

W = ”naimaikäisten naisten lukumäärä”.

Olkoot

$$X = \frac{U}{W} \quad \text{ja} \quad Y = \frac{V}{W}.$$

Jos käytännössä havaitsemme huomattavan positiivisen korrelaation X:n ja Y:n välillä, voimmeko päätellä, että haikarat tuovat lapsia?