4. Centrala gränsvärdessatsen

4.1. Centrala gränsvärdes satsen och normal approximation

fi: keskeinen raja-arvo lause

motivation: svaga STL (3.5)

om: $X_1, X_2 \ldots$ följd s.v. $\Omega \to \mathbb{R}$
$\{X_j \mid j \in \mathbb{N}\}$ oberoende
$E(X_j) = \mu, \quad D^2(X_j) = \sigma^2 \quad \forall j$

$$\overset{STL}{\Longrightarrow} P\left(\left\{\omega : \left|\frac{1}{n}\sum_{j=1}^{n} X_j(\omega) - \mu\right| \geq \varepsilon\right\}\right) \xrightarrow[n\to\infty]{} 0$$

$\forall$ fixerade $\varepsilon > 0$, beteckna $S_n(\omega) = \sum_{j=1}^{n} X_j(\omega)$ summa s.v. av $X_1 \ldots X_n$

$$P\left(\{\omega : |S_n(\omega) - n\mu| \geq n\cdot\varepsilon\}\right) \xrightarrow[n\to\infty]{} 0 \quad \text{dvs.} \quad S_n \approx n\mu$$

Problem: vad kan sägas mera noggrant om $S_n - n\mu$?

"svar": $S_n - n\mu \approx \sigma\sqrt{n}$

Svar: CGS efter lämplig "normalisering":

Centrala gränsvärdessatsen (CGS, "central limit theorem") CLT — specialfall

Anta:

- $X_1, X_2, \ldots$ följd av s.v. $\Omega \to \mathbb{R}$, dvs följd $(X_j)_{j\in\mathbb{N}}$ av s.v.
- $\{X_j \mid j \in \mathbb{N}\}$ $\perp\!\!\!\perp$ oberoende (dvs. $\{X_{j_1}, X_{j_2}, \ldots X_{j_n}\}$ $\perp\!\!\!\perp$ s.v. för $\forall j_1 < j_2 < \ldots < j_n$)
- $X_j \sim X_1 \quad \forall j = 1, 2 \ldots$ (X_j har samma fördelning $\forall j \in \mathbb{N}$)
- låt $E(X_j) = \mu, \quad Var(X_j) = \sigma^2 \quad \forall j \in \mathbb{N}$ (antar: X_j har varians $\forall j$)
- $\sigma > 0$ (dvs $X_j \neq$ konstant s.v.)

Beteckna $S_n = \sum_{j=1}^{n} X_j = X_1 + \ldots + X_n$, $n \in \mathbb{N}$,

och

$$Z_n = \frac{S_n - E(S_n)}{D(S_n)} = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

(där $D(S_n) = \sqrt{Var(S_n)}$)

kom ihåg:
$E(S_n) = n\mu$
$Var(S_n) \overset{\perp\!\!\!\perp}{=} \sum_{j=1}^{n} \underbrace{Var(X_j)}_{=\sigma^2} = n\sigma^2$

Då gäller, fördelningsfunktionen

$$P\left(\{\omega : Z_n(\omega) \leq t\}\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{x^2}{2}}\,dx \overset{def}{=} \Phi(t) \quad \forall t \in \mathbb{R}$$

f.f. för $N(0,1)$-fördelad s.v.

Terminologi: Z_n konvergerar till s.v. $X \sim N(0,1)$ normalfördelad i fördelning (= distribution) då $n \to \infty$, (dvs. fördelningsfkt. till Z_n konv. till $N(0,1)$-fördelning).

dvs. f.f. $F_{Z_n}(t) \xrightarrow[n\to\infty]{} F_X(t) \equiv \Phi(t) \quad \forall t \in \mathbb{R}$, där $X \sim N(0,1)$

betecknа ovan: $Z_n \xrightarrow[n\to\infty]{d} X \quad (\sim N(0,1))$

Bevis (se [T], s. 118 använder 3.9 (*) karakteristiska fkt.)

OBS ① normalisering: $Z_n = \dfrac{S_n - E(S_n)}{D(S_n)} \Rightarrow E(Z_n) = \dfrac{E(S_n)-E(S_n)}{D(S_n)} = 0,$

och variansen

$$Var(Z_n) = \frac{1}{D(S_n)^2} Var(S_n - E(S_n)) \overset{\text{kap 3.}}{=} \frac{1}{D(S_n)^2} Var(S_n) = 1 \quad \left(ty\ D(S_n) = \sqrt{Var(S_n)}\right) \quad \forall n \in \mathbb{N}$$

dvs. som $X \sim N(0,1)$ $\quad (\Rightarrow E(X)=0,\ D^2(X)=1)$.

② CGS gäller under allmännare antaganden (se [T], s. 120).
exempelvis: X_j inte samma fördelning

③ jämförelse med STL: $\bar{X}_n = \frac{1}{n} S_n \Rightarrow S_n = n\bar{X}_n$

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{n}{\sigma\sqrt{n}}(\bar{X}_n - \mu) = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow[n\to\infty]{d} X, \quad X \sim N(0,1)$$

Exempel 4.1 normalapproximation

anta $\{X_1, \dots, X_{20}\}$ $\perp\!\!\!\perp$ $\quad X_j \sim \text{Tas}(0,1) \quad \forall j = 1, \dots, 20$

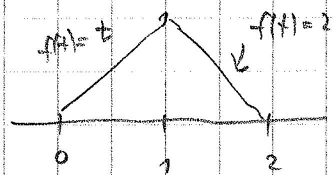
låt $S = \sum_{j=1}^{20} X_j$ (tolkning: S = summa av 20 slumptal från $(0,1)$)

Fråga fördelningsfunktionen av S? $\quad$ exakt f-f till S svår beräkna

2.6

19 konvolutioner $\Rightarrow$ exakt frekvens-fkt. till S!

(2 st: $X_1 + X_2$ frekvens-fkt.)



i stället alternativt CGS (ger approximation)

$$\mu = E(X_j) = \int_0^1 x\,dx = \frac{1}{2}$$

$$\sigma^2 = Var(X_j) = E(X_j^2) - (E(X_j))^2 = \int_0^1 x^2\,dx - \frac{1}{4} = \frac{1}{12}$$

$\Rightarrow$

$$E(S) \overset{\text{lin.}}{=} 20 \cdot \frac{1}{2} = 10, \quad Var(S) \underset{\perp\!\!\!\perp}{=} \sum_{j=1}^{20} Var(X_j) = \frac{20}{12} = \frac{5}{3}$$

CGS: betrakta $\quad Z = \dfrac{S - E(S)}{D(S)} = \dfrac{S-10}{\sqrt{\frac{5}{3}}}$

där $D(S) = \sqrt{Var(S)}$

betrakta $n = 20$ "stort" $\Rightarrow$ f-f $\approx \Phi(?)$ $\quad N(0,1)$ – fördelat

$$P(\{\omega : S(\omega) \le t\}) \overset{\text{normalisera korrekt}}{=} P\left(\left\{\omega : \underbrace{\frac{S(\omega)-10}{\sqrt{\frac{5}{3}}}}_{\sim N(0,1)} \le \frac{t-10}{\sqrt{\frac{5}{3}}}\right\}\right) \overset{\text{CGS}}{\approx} \Phi\left(\frac{t-10}{\sqrt{\frac{5}{3}}}\right), \quad t \in \mathbb{R}$$

hur bra uppskattning?
"Large Deviations" – teori

jämförelse olika värden på t
se [T], s. 119

CGS gäller **inte** alltid

Exempel 4.2 (a) om variansen $D^2(X_j) = \sigma^2 = 0 \overset{3.2}{\Rightarrow} X_j =$ "konstant" (väsentligen)

$$\Rightarrow S_n = \sum_{j=1}^n X_j \text{ väs. "konstant"} \quad (\not\sim N(0,1)).$$

(b) anta $\{X_j \mid j \in \mathbb{N}\}$ $\perp\!\!\!\perp$ s.v.

$$X_j \sim \text{Poisson}\left(\frac{1}{2^j}\right), \quad \forall j = 1, 2 \ldots$$

här har X_j **inte** samma fördelning!

$$2.2 \Rightarrow \mathbb{E}(X_j) = \frac{1}{2^j}, \quad \text{var}(X_j) = \frac{1}{2^j}$$

2.6/3.6 fallet n=2

$$\Rightarrow S_n = \sum_{j=1}^n X_j \sim \text{Poisson}\left(\sum_{j=1}^n \frac{1}{2^j}\right),$$

$$\sum_{j=1}^n \frac{1}{2^j} \underset{\text{serie}}{\overset{\text{geom.}}{=}} 1 - \frac{1}{2^n}$$

eller direkt

$$\Rightarrow \mathbb{E}(S_n) = 1 - \frac{1}{2^n}, \quad \text{var}(S_n) \overset{\text{Poisson}}{=} 1 - \frac{1}{2^n}$$

"Lätt" verifiera: ⊛

$S_n \xrightarrow{d} X$,

$X \sim \text{Poisson}(1)$

diskreta s.v., fr.sl

$$\Rightarrow \frac{S_n - \mathbb{E}(S_n)}{D(S_n)} = \frac{S_n - \left(1 - \frac{1}{2^n}\right)}{\sqrt{1 - \frac{1}{2^n}}} \xrightarrow[n\to\infty]{d} X - 1, \quad \not\sim N(0,1)$$

$\frac{1}{2^n} \to 0$

där $X \sim \text{Poisson}(1)$!

17-11.

---

kursmässigt inte ⊛ 2015

ovan: $X_j : \Omega \to \mathbb{N} \Rightarrow S_n = X_1 + \ldots + X_n$ <u>diskret</u> s.v. $\forall n$

<u>Påst</u> $S_n \xrightarrow[n\to\infty]{d} X \sim \text{Poisson}(1)$ $\quad X_1(\omega) + \ldots + X_n(\omega) \in \mathbb{N} \quad \forall n, \forall \omega$

diskreta s.v. $\Rightarrow$ räcker fr.sl: $\quad X \sim \text{Poisson}(\lambda_n), \quad \lambda_n = 1 - \frac{1}{2^n}$

$$P(\{\omega : S_n(\omega) = k\}) = e^{-\lambda_n} \frac{(\lambda_n)^k}{k!} = e^{-\left(1 - \frac{1}{2^n}\right)} \frac{\left(1 - \frac{1}{2^n}\right)^k}{k!} \xrightarrow[n\to\infty]{} e^{-1} \frac{1}{k!}$$

$$= P(\{\omega : X(\omega) = k\}) \quad \text{då} \quad X \sim \text{Poisson}(1).$$

eftersom $\begin{cases} x \mapsto e^{-x} \text{ kontin.} \\ x \mapsto x^k \text{ kontin.} \end{cases}$

För flera exempel på CGS tittar vi först på klassiska specialfallet $X_n \sim \text{Bin}(n, p)$.

4.2. Normal approximation av diskreta s.v.

exempelvis: $\left\{\begin{array}{l}\text{antalet 6:or i oberoende upprepade tärningskast} \\ \text{antalet kronor i — " — — " — myntsinglingar}\end{array}\right.$ etc

Korollarium 4.3 (de Moivre–Laplace, 1700-t). (Kor. till CGS)

Anta $X_n \sim Bin(n,p)$, $n \in \mathbb{N}$ och $0<p<1$ fixerat. Då gäller att

$$\frac{X_n - np}{\sqrt{npq}} \xrightarrow[n\to\infty]{d} N(0,1)$$

$\boxed{\text{Obs Högra sidan kontin. fördelad s.v.}}$

Idess.
$$P\left(\left\{\omega : \frac{X_n(\omega) - np}{\sqrt{npq}} \le t\right\}\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx, \quad \forall t \in \mathbb{R}$$

Bevis (Obs kan inte tillämpa CGS på $X_n, n \in \mathbb{N}$, ty har inte samma fördelning!)

påst. beror endast på f-fkt (frekvensf.) till $X_n \Rightarrow$ kan anta

$$X_n = \sum_{j=1}^{n} 1_{A_j} \quad \text{där } A_j = \text{händelse } A \text{ inträffar vid } j\text{:s ober. upprepning}, \; j=1,\dots,n$$

$\{A_1, \dots, A_n\}$ $\perp\!\!\!\perp$ händelser, $P(A_j) = P(A) = p \quad (j=1,\dots,n)$ $\forall n \in \mathbb{N}$

(försök med oberoende upprepningar, se kap. 3.3)

Alltså: $\{1_{A_j} \mid j \in \mathbb{N}\}$ $\perp\!\!\!\perp$ oberoende s.v. (eftersom $\{A_1, \dots, A_n\}$ $\perp\!\!\!\perp$ händelser (upprepn.) för alla n)

vet $1_{A_j} \sim 1_{A_1}$

CGS: $\mathbb{E}(1_{A_j}) = p \Rightarrow \mathbb{E}(X_n) \overset{\text{lin}}{=} np \quad \left(= \sum_{j=1}^{n} \mathbb{E}(1_{A_j})\right)$

$Var(1_{A_j}) \overset{3.2}{=} P(A_j) P(A_j^c) = pq \quad \forall j = 1,2,\dots$ där $q = 1-p$

$$\Rightarrow Var(X_n) \overset{\perp\!\!\!\perp}{=} \sum_{j=1}^{n} Var(1_{A_j}) = npq$$

$$\text{CGS} \Rightarrow \quad \frac{X_n - np}{\sqrt{npq}} \xrightarrow[n\to\infty]{d} N(0,1) \qquad \square.$$

Tumregel: $npq > 10 \Rightarrow$ normalapproximation bra

Exempel 4.4 Ett mynt singlas 10.000 ggr (oberoende)

låt X = antalet kronor (i 10.000 kast)

Uppskatta sl. $P(4850 < X < 5150) = ?$

Lösningen ovan: $n = 10.000$, $p = q = \frac{1}{2} \Rightarrow np = 5000$

$$\sqrt{npq} = \sqrt{10.000 \cdot \tfrac{1}{4}} = \sqrt{2500} = 50$$

normalisera för CGS:
$$X_0 = \frac{X - 5000}{50} \Rightarrow$$

$$P(4850 < X < 5150) = P(-150 < X - 5000 < 150) = P\left(-3 < \frac{X-5000}{50} < 3\right)$$

$$= P(-3 < X_0 < 3) \underset{\substack{\text{normal-}\\\text{approximation}\\(\text{Kor } 4.3)}}{\approx} \frac{1}{\sqrt{2\pi}} \int_{-3}^{3} e^{-\frac{x^2}{2}} dx = \Phi(3) - \underbrace{\Phi(-3)}_{= 1-\Phi(3)\ \text{symmetri}}$$

$$= 2\underset{\substack{\uparrow\\ 0.998650}}{\Phi(3)} - 1 \approx 0.9973$$

□

Fråga

vilfrekvens sd

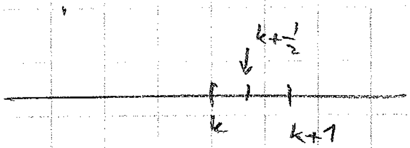
$P(X = 5000) \approx ?$ [erhåller exakt 5000 kronor]

vid normalapproximation:
$$P(X = 5000) = P(X_0 = 3) \approx \frac{1}{\sqrt{2\pi}} \int_3^3 e^{-\frac{x^2}{2}} dx = 0 \quad \text{inte användbart}$$

(*)

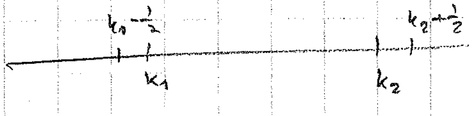
Allmänt: $X : \Omega \to \mathbb{Z}$ diskret s.v ($\mathbb{Z}$ = heltalen) $\Rightarrow$

ersätt
$$\{\omega : X(\omega) \le k\} \underset{\substack{\uparrow\\ X(\Omega) \subset \mathbb{Z}}}{=} \{\omega : X(\omega) \le k + \tfrac{1}{2}\}, \quad k \in \mathbb{Z}$$



likaså: händelsen
$$\{\omega : k_1 \le X(\omega) \le k_2\} = \{\omega : k_1 - \tfrac{1}{2} \le X(\omega) \le k_2 + \tfrac{1}{2}\}$$
$k_1 < k_2$, $k_1, k_2 \in \mathbb{Z}$

ger (vanligen) bättre resultat vid approximation



ovanstående kallas kontinuitetsmodifikation

Exempel 4.4 (fortsätt.) ovan $X_0 = \frac{X-5000}{50}$

$$P(X = 5000) = P(4999.5 \le X \le 5000.5) = P\left(-\tfrac{1}{2} < X - 5000 < \tfrac{1}{2}\right)$$

$$\overset{\cdot\frac{1}{50}}{=} P\left(-\frac{1}{100} < \frac{X-5000}{50} < \frac{1}{100}\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-0.01}^{0.01} e^{-\frac{x^2}{2}} dx \overset{\text{som ovan}}{=} 2\Phi(0.01) - 1$$

$$\approx \underline{0.00798}$$

alternativ: 0.01 litet

$$2\Phi(0.01) - 1 = 2\left(\Phi(0.01) - \underset{\frac{1}{2}}{\Phi(0)}\right) \overset{\text{derivata}}{\approx} 2 \cdot \underbrace{\Phi'(0)}_{=\frac{1}{\sqrt{2\pi}}} \cdot 0.01 = \underline{0.00797}$$

eftersom $\Phi'(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ (frekvensfunktionen).

(*) exakt värde: $X \sim \text{Bin}(10.000, \tfrac{1}{2}) \Rightarrow$
$$P(X = 5000) = \binom{10000}{5000} \frac{1}{2^{10000}} = ??$$

□

25/2/2015
23.11

Exempel 4.5 (kursprov 1981)

Sök ett närmevärde till så att relativa frekvensen av 6:or i 1000 tärningskast ligger i intervallet $[0.16, 0.17]$. Svaret får innehålla värden för f.f. Φ till $X \sim N(0,1)$.

Lösn s.v X = antalet 6:or i 1000 kast $\sim Bin(1000, \frac{1}{6})$

$$P(\text{sexa}) = \frac{1}{6} \approx 0.1666\ldots, \quad n = 1000 \Rightarrow$$

$$np = \frac{1000}{6}, \quad \sqrt{npq} = \sqrt{1000 \cdot \frac{1}{6} \cdot \frac{5}{6}} = \sqrt{\frac{5000}{36}} = \frac{50\sqrt{2}}{6} = \frac{25\sqrt{2}}{3}$$

relativa frekvens 6:or i $[0.16, 0.17]$ betyder att

$$160 \le X \le 170 \qquad (X = \text{antalet 6:or i 1000 kast}).$$

Vi söker

$$P(160 \le X \le 170) = P\left(160 - \frac{1000}{6} \le X - \frac{1000}{6} \le 170 - \frac{1000}{6}\right)$$

$$= P\left(\frac{960-1000}{6} \le X - \frac{1000}{6} \le \frac{1020-1000}{6}\right) = P\left(-\frac{40}{6} \le X - \frac{1000}{6} \le \frac{20}{6}\right)$$

$$= P\left(-\frac{40}{6} \cdot \frac{3}{25\sqrt{2}} \le \frac{X - \frac{1000}{6}}{\sqrt{\frac{5000}{36}}} \le \frac{20}{6} \cdot \frac{3}{25\sqrt{2}}\right) = P\left(-\frac{4}{5\sqrt{2}} \le \frac{X - \frac{1000}{6}}{\sqrt{\frac{5000}{36}}} \le \frac{2}{5\sqrt{2}}\right)$$

(beteckna $a = \frac{4}{5\sqrt{2}}$, $b = \frac{2}{5\sqrt{2}}$)

$$\underset{\text{de Moivre-Laplace}}{\approx} \Phi(b) - \underbrace{\Phi(-a)}_{= 1 - \Phi(a)} = \Phi(b) + \Phi(a) - 1$$

(kan bestämmas från tabell över $\Phi(\cdot)$).

Anmärkning: i ovanstående ex. kunde även använda kontinuitets modifikationen

$$P(160 \le X \le 170) = P(159{,}5 \le X \le 170{,}5) = \ldots\ldots$$

men ger mera komplicerade beräkningar

Exempel 4.6 En tärning kastas 20 ggr. Vilken är så att poängsumman S av ögontalen satisfierar $60 \le S \le 80$?

Inte på föreläsn 2013

Lösn

kapitel 3.6 exakt formel (svår beräkna) för detta problem (av de Moivre) hellre normal approximation ?

Låt X_j = poängtalet vid j:s kast, $j = 1, \ldots, 20 \Rightarrow$

$$S = \sum_{j=1}^{20} X_j$$

där $\{X_1, \ldots, X_{20}\}$ oberoende s.v.

och $X_j \sim X_1$, $j = 1, \ldots, 20$

väntevärdet $E(X_j) = \frac{1}{6}(1+2+\ldots+6) = \frac{7}{2}$ (eftersom $1+2+\ldots+6 = \frac{6 \cdot 7}{2} = 21$).

variansen $Var(X_j) = E(X_j^2) - (E(X_j))^2 = \frac{1}{6}(1^2+2^2+\ldots+6^2) - \left(\frac{7}{2}\right)^2 =$

med induktion: $1^2+2^2+\ldots+n^2 = \frac{n(n+1)(2n+1)}{6}$ för alla n (123)

$$\overset{\downarrow}{=} \frac{1}{6}(6\cdot 7\cdot 13)\cdot\frac{1}{6} - \frac{49}{4} = \frac{7\cdot 13}{6} - \frac{49}{4} = \frac{182-147}{12} = \frac{35}{12}$$

kom ihåg: X diskret s.v. $\Omega \to \mathbb{N}$, $p_k = P(X=k)$, $k=0,1,2,\ldots$, frekvensfkl $\Rightarrow$

$$E(X^2) = \sum_{k=0}^{\infty} k^2 p_k$$

(se kapitel 3.1)

$$\Rightarrow E(S) \overset{\text{lin}}{=} 20\cdot\frac{7}{2} = 70, \quad \mathrm{Var}(S) \overset{\text{II}}{=} \sum_{j=1}^{20} \underbrace{\mathrm{Var}(X_j)}_{\frac{35}{12}\ \forall j} = 20\cdot\frac{35}{12} = \frac{175}{3}$$

$$\Rightarrow P(60 \le S \le 80) \overset{\text{kont. modif.}}{=} P(59{,}5 \le S \le 80{,}5) = P(59{,}5-70 \le X-70 \le 80{,}5-70)$$

$$= P(-10{,}5 \le X-70 \le 10{,}5) = P\left(\frac{-10{,}5}{\sqrt{\frac{175}{3}}} \le \frac{X-70}{\sqrt{\frac{175}{3}}} \le \frac{10{,}5}{\sqrt{\frac{175}{3}}}\right)$$

$$\overset{\text{CGS}}{\approx} \Phi\left(\frac{10{,}5}{\sqrt{\frac{175}{3}}}\right) - \underbrace{\Phi\left(-\frac{10{,}5}{\sqrt{\frac{175}{3}}}\right)}_{1-\Phi\left(\frac{10{,}5}{\sqrt{\frac{175}{3}}}\right)} = 2\,\Phi\left(\frac{10{,}5}{\sqrt{\frac{175}{3}}}\right) - 1 \underset{\text{tabell}}{\approx} 0.831$$

med enklare beteckningar:

låt $a = \frac{10.5}{\sqrt{\frac{175}{3}}}$ $\Rightarrow$

$$P\left(-a \le \frac{X-70}{\sqrt{\frac{175}{3}}} \le a\right) \overset{\text{CGS}}{\approx} \Phi(a) - \underbrace{\Phi(-a)}_{=1-\Phi(a)} = 2\,\Phi(a) - 1 \overset{\text{ovan}}{\approx} 0.831$$

D.

Addendum: 2-dimensionella fördelningar, korrelationskoefficienten (statistik etc)

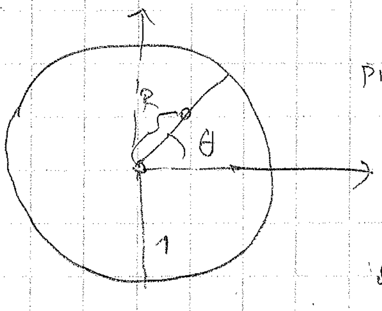
antag $(\Omega, \mathcal{F}, P)$ sl-rum, X och Y s.v på Ω.

Det är ofta viktigt att studera stokastiska vektorer (X, Y) dvs.

$$\omega \mapsto (X(\omega), Y(\omega)) \qquad \Omega \to \mathbb{R}^2$$

exempelvis: (X,Y) beskriver hur X och Y beror på varandra eller korrelerar med varandra

Exempel A slumpmässig pilkastning



pil kastas på måfå mot tavla med radie 1

träffpunkten

$$P = (R\cos\theta, R\sin\theta) \quad \text{i polära koordinater}$$

där R = avståndet av P från origo O och θ = vinkeln med positiva x-axeln

alltså: $0 \le R \le 1$, $0 \le \theta \le 2\pi$

ger stokastisk vektor (R, θ)

på måfå? Olika val av fördelningar, exempelvis $R \sim Jas(0,1)$, $\theta \sim Jas(0, 2\pi)$ likformiga, och $R \perp\!\!\!\perp \theta$

eller $\theta \sim Jas(0,1)$, R så att $P\left(\frac{k-1}{10} \le R < \frac{k}{10}\right) = k\alpha$, där $\alpha > 0$ vald så att $\alpha \cdot \sum_{k=1}^{10} k = 1$.

enklaste specialfallet: diskreta 2-dim. fördelningar

Låt X och Y: $\Omega \to \mathbb{R}$ vara diskreta s.v.

$$X(\Omega) = \{x_i : i = 1, 2 ..\}, \quad Y(\Omega) = \{y_j : j = 1, 2 ..\}$$ ändliga / numrerbart oändliga

Då är stokastiska vektorn (X,Y) diskret, med värdemängden

$$\{(X(\omega), Y(\omega)) : \omega \in \Omega\} \subset \{(x_i, y_j) : i,j\}$$ (obs kan vara $\subsetneq$ strikt olikhet)

Motsvarande frekvensfkn är

$$p_{ij} = P(\{\omega \in \Omega : X(\omega) = x_i, Y(\omega) = y_j\}) \quad \forall i,j \quad (\text{kan vara } = 0)$$

som entydigt bestämmer (X,Y) samt motsvarande 2-dim fördelningsfunktion är

$$F_{(X,Y)}(s,t) \overset{bet}{=} F(s,t) \overset{def}{=} \sum_{\substack{\text{par } (i,j) \\ \text{med } x_i \le s,\ y_j \le t}} p_{ij}, \quad \text{då } s,t \in \mathbb{R}.$$

Fördelningarna till X och Y erhålls som <u>marginal fördelningarna</u>

$$P(\{X=x_i\}) = \sum_j p_{ij} \qquad (i \text{ fixerat})$$

$$P(\{Y=y_j\}) = \sum_i p_{ij} \qquad (j \text{ fixerat})$$

<u>Exempel B</u> Två tärningar kastas

$X=$ poängtalet av 1. tärningen

$Y=$ summan av poängtalen av tärning 1 och tärning 2

Låt $\Omega = \{1,\dots,6\}^2$, $X(n_1,n_2)=n_1$, $Y(n_1,n_2)=n_1+n_2$ $(n_1,n_2)\in\Omega$

Bestäm frekvensfkn till (X,Y):

$$p_{ij} = P(\{X=i\}\cap\{Y=j\}) \quad \text{där } i\in\{1,\dots,6\},\ j\in\{2,\dots,12\}$$

Bassfl $P((n_1,n_2)) = \frac{1}{36} = p$ i Ω för $(n_1,n_2)\in\Omega$

tabellera summan

$12=6+6$, förekommer 1. gång

$11 = 5+6 = 6+5$ 2 gr

Y \ X	1	2	3	4	5	6	
12					o	p	1p
11				o	p	p	2p
10			o	p	p	p	
9		o	p	p	p	p	
8	o	p	p	p	p	p	
7	p	p	p	p	p	p	6p
6	p	p	p	p	p	o	
5	p	p	p	p	o		
4	p	p	p	o			
3	p	p	o				
2	p	o					1p
	6ggr 6p					6p	

p_{ij} ← $P(Y=j)$

↑ $P(X=i)$

<u>Obs</u> $X \not\perp Y$ inte oberoende

$P(X=2,\ Y=3) = \frac{1}{36} = p$, eftersom

$\{X=2\} = \{(2,n_2):\ n_2\in\{1\dots6\}\}$

$\{Y=3\} = \{(1,2),(2,1)\}$

$\Longrightarrow \{X=2\}\cap\{Y=3\} = \{(2,1)\}$.

å andra sidan

$$P(X=2) = \frac{6}{36},\quad P(Y=3) = \frac{2}{36} \Rightarrow$$

$$P(X=2)\cdot P(Y=3) = \frac{6}{36}\cdot\frac{2}{36} = \frac{12}{36} = \frac{1}{3} \neq P(\{X=2\}\cap\{Y=3\}).$$

alltså X och Y <u>inte</u> oberoende s.v. □

allmänt: 2-dim fördelningsfunktion till stokastisk vektor (X,Y) definieras av

$$F(s,t) = P\left(\{\omega : X(\omega) \le s\} \cap \{\omega : Y(\omega) \le t\}\right) \quad \text{för alla } s,t \in \mathbb{R}$$

kontin. fördelning avser 2-dim integraler av frekvensfunktion $f: \mathbb{R}^2 \to \mathbb{R}$

(vilket betydligt mera komplicerat jämfört med stokastiska variabler).

Korrelation: antag $(\Omega, \mathcal{F}, P)$ sl-rum, X och Y s.v. på Ω.

korrelationskoefficienten till X och Y definieras som

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}, \quad \text{om } \mathrm{Var}(X) > 0 \text{ och } \mathrm{Var}(Y) > 0.$$

där kovariansen $\mathrm{Cov}(X,Y) = \mathbb{E}(X\cdot Y) - \mathbb{E}(X)\cdot\mathbb{E}(Y)$ (kap 3)

$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ variansen till X

Om $\mathrm{Var}(X) = 0$ eller $\mathrm{Var}(Y) = 0$, sätter man $\mathrm{Corr}(X,Y) = 0$ ("konvention")

Alltså: $\mathrm{Cov}(X,Y) = 0 \iff \mathbb{E}(X\cdot Y) = \mathbb{E}(X)\cdot\mathbb{E}(Y)$ (i fallet $\mathrm{Var}(X)\mathrm{Var}(Y) > 0$)

speciellt: om $X \perp\!\!\!\perp Y$, så $\mathrm{Cov}(X,Y) = 0$ dvs. $\mathrm{Corr}(X,Y) = 0$.

Terminologi: X och Y korrelerade om $\mathrm{Corr}(X,Y) \ne 0$, okorrelerade om $\mathrm{Corr}(X,Y) = 0$.

Exempel C: X, Y s.v. så att $X \not\perp\!\!\!\perp Y$ inte oberoende, men

$\mathrm{Corr}(X,Y) = 0$ (och $\mathrm{Var}(X)\mathrm{Var}(Y) > 0$).

Nämligen, låt $X \sim N(0,1)$ standard normal fördelad $\Rightarrow \mathbb{E}(X) = 0$

och $Y = X^2$ transformerad s.v. av X.

$$\mathrm{Cov}(X,Y) = \mathbb{E}(X\cdot Y) = \mathbb{E}(\underbrace{X\cdot X^2}_{=X^3}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \underbrace{x^3 \cdot e^{-\frac{x^2}{2}}}_{\text{udda fktn}}\,dx = 0$$

(eller integrera partiellt)

$h(-t) = -h(t) \ \forall t \in \mathbb{R}$

Pss. X och Y är inte oberoende:

$$P\left(\{X(\omega) \le 1\} \cap \{\underbrace{X(\omega)^2 \le 1}_{-1 \le X(\omega) \le 1}\}\right) = P(-1 \le X(\omega) \le 1) = \Phi(1) - \Phi(-1)$$

där $\Phi = N(0,1)$ f.f.

och

$$P(X \le 1)\, P(X(\omega)^2 \le 1) = \underbrace{\Phi(1)}\left(\Phi(1) - \Phi(-1)\right)$$

där $\Phi(1) \ne 1$ ty $\Phi(t) < 1 \ \forall t \in \mathbb{R}$.

påminn: $X \perp\!\!\!\perp Y$ betyder att

$$P(\{X \le a\} \cap \{Y \le b\}) = P(X \le a)\,P(Y \le b) \quad \text{för alla } a,b \in \mathbb{R}.$$

$\square$