TDAB01 Sannolikhetslära och Statistik

Jose M. Peña
IDA, Linköpings Universitet

Föreläsning 6

Översikt

- **Stokastiska processer**
- **Markovkedjor**
- **Binomialprocess**
- **Poissonprocess**

Stokastiska processer

- **Stokastisk process**: En **sekvens** av slumpvariabler $X_1, X_2, ..., X_T$ **observerade över tid**.
  - Exempel: X_t = antalet påträffade buggar under dag t, $t = 1, 2, \ldots, T$.
  - Exempel: X_t = slutkursen på Ericsson aktie vid dag t.
  - Exempel: X_t = temperaturen vid en viss plats vid tidpunkt t.
- **Stokastisk process**: En slumpvariabel $X(t, \omega)$ som också beror av tiden, där
  - $t \in \mathcal{T}$ och $\mathcal{T}$ är en mängd tidpunkter, t ex $\mathcal{T} = \{1, 2, 3, \ldots\}$.
  - $\omega \in \Omega$ är utfallet i ett experiment (precis som förut).
  - Värden på $X(t, \omega)$ kallas **tillstånd**.
  - Härefter $X(t) = X(t, \omega)$ (precis som förut).
- Uppdelning av processer:
  - Diskreta eller kontinuerliga **tillstånd**.
  - Diskret eller kontinuerlig **tid**.
- Diskret tillstånd, kontinuerlig tid: Väljarsympatier över tid.
- Diskret tillstånd, diskret tid: Väljarsympatier på valdagen.
- Kontinuerligt tillstånd, diskret tid: Dagens högsta temperatur.
- Kontinuerligt tillstånd, kontinuerlig tid: Temperaturen över tid.

Markovprocesser

- **Markovprocess**: Prognosen för morgondagen beror endast på idag:

$$P(\text{framtiden}|\text{nu}, \text{historiken}) = P(\text{framtiden}|\text{nu})$$

- **Markovprocess**: För alla tidpunkter $t_1 < \ldots < t_n < t_{n+1}$ gäller att

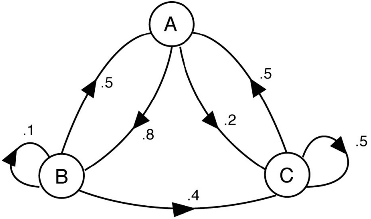
$$P(x(t_{n+1})|x(t_1), \ldots, x(t_n)) = P(x(t_{n+1})|x(t_n))$$

  dvs, $X(t_{n+1})$ är oberoende av $X(t_1), \ldots, X(t_{n-1})$ givet $X(t_n)$.
- Många processer är inte Markov. Praktiskt **antagande**.

Markovkedjor

- **Markovkedja**: Markovprocess med **diskret tid** och **diskreta tillstånd**.
- **Transitionssannolikheter (en-stegs)**

$$p_{ij}(t) = \boldsymbol{P}\left\{X(t+1) = j | X(t) = i\right\}$$



- **Transitionssannolikheter (h-stegs)**

$$p_{ij}^{(h)}(t) = \boldsymbol{P}\left\{X(t+h) = j | X(t) = i\right\}$$

Markovkedjor

- **Homogen Markovkedja**: Transitionssannolikheterna är konstanta över tiden, dvs
$$p_{ij}(t) = p_{ij}$$
- **Transitionsmatris**
$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$
- Example: $\Omega = \{\text{sol}, \ \text{regn}\}$
$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.7 & 0.3 \end{pmatrix}$$
- Example: $\Omega = \{\text{RödGröna}, \text{Alliansen}, \ \text{SD}\}$
$$P = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.5 & 0.4 \end{pmatrix}$$
- Se `SimulateMarkovChain.R`.

Markovkedjor

- **Transitionssannolikheter (h-stegs)**

$$p_{ij}^{(h)}(t) = \boldsymbol{P}\left\{X(t+h) = j | X(t) = i\right\}$$

- Komplext. Det finns många vägar som tar oss $i \to j$ när $h > 1$.
- Example: $\Omega = \{1, 2\}$. Om $h = 2$ kan vi göra resan $1 \to 2$ på två sätt:
  - $1 \to 1 \to 2$
  - $1 \to 2 \to 2$
- 2-stegs transitionssannolikhet $1 \to 2$:

$$p_{12}^{(2)} = p_{11}p_{12} + p_{12}p_{22}$$

- 3-stegs transitionssannolikhet $1 \to 2$:

$$\begin{aligned} p_{12}^{(3)} &= p_{11}p_{11}p_{12} + p_{11}p_{12}p_{22} + p_{12}p_{21}p_{12} + p_{12}p_{22}p_{22} \\ &= p_{11}(p_{11}p_{12} + p_{12}p_{22}) + p_{12}(p_{21}p_{12} + p_{22}p_{22}) \\ &= p_{11}p_{12}^{(2)} + p_{12}p_{22}^{(2)} \end{aligned}$$

Markovkedjor

- Transitionsmatris 1-steg

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}$$

- Transitionsmatris h-steg

$$P^{(h)} = \begin{pmatrix} p_{11}^{(h)} & p_{12}^{(h)} & \cdots & p_{1n}^{(h)} \\ p_{21}^{(h)} & p_{22}^{(h)} & \cdots & p_{2n}^{(h)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1}^{(h)} & p_{n2}^{(h)} & \cdots & p_{nn}^{(h)} \end{pmatrix}$$

- **Resultat**: $P^{(h)}$ är h:te matrispotensen av P

$$P^{(h)} = P \cdots P \cdots P = P^h$$

- Se Example 6.9 i Baron.

Marginalfördelning

- **Initialfördelning** vid $t = 0$ är radvektorn

$$P_0 = (P_0(1), P_0(2), ..., P_0(n))$$

- **Sannolikhetsfördelning över tillstånden** efter h steg

$$P_h = (P_h(1), P_h(2), ..., P_h(n))$$

- **Resultat**:

$$P_{\boldsymbol{h}} = P_0 P^{\boldsymbol{(h)}} = P_0 P^{\boldsymbol{h}}$$

- Example: $P_0 = (1/3, 1/3, 1/3)$ och

$$P = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.5 & 0.4 \end{pmatrix}$$

Då

$$P_3 = (1/3, 1/3, 1/3) \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.5 & 0.4 \end{pmatrix}^3 = (0.333, 0.407, 0.259)$$

- Se Example 6.10 i Baron.

Stationärfördelning

- Vad är sannolikhetsfördelningen över tillstånden efter många steg ?
- **Stationärfördelning** är radvektorn

$$\pi = \lim_{h \to \infty} P_h$$

- Obs. $\pi P = \pi$.
- En Markovkedja är **reguljär** om det finns en h så att

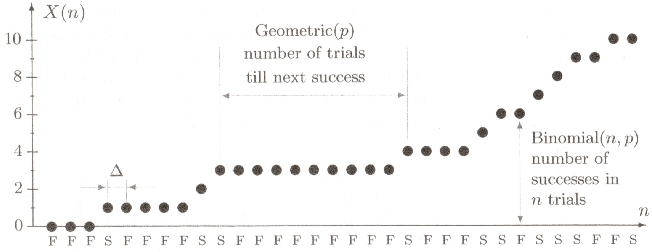
$$p_{ij}^{(h)} > 0$$

  för alla i, j.
- **En reguljär Markov kedja har en stationärfördelning**. Man hittar den genom att lösa ekvationssystemet

$$\begin{cases} \pi P = \pi \\ \sum_x \pi_x = 1 \end{cases}$$

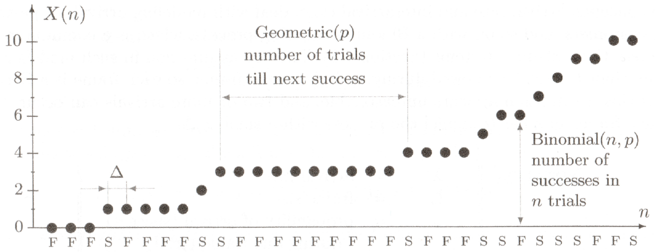
Binomialprocess

- **Räkneprocesser:** $X(t)$ är antalet räknade saker t o m tidpunkt t.
- **Binomialprocess**: $X(n)$ är **antalet lyckade försök i de n första** i en sekvens av oberoende Bernoulliförsök med sannolikhet p.
- $X(n) \sim Binomial(n, p)$
- Y = antalet försök mellan två lyckade.
- $Y \sim Geometrisk(p)$



- Se SimulateBinomialProcess.R.

Binomialprocess



- Ett nytt Bernoulliförsök var Δ sekund. $\Delta =$ **time frame**.
- n försök tar $t = n\Delta$ sekunder att utföra. Så, $n = t/\Delta$.
- Processen kan defineras som funktion av (klock)tid, dvs $X(n) = X(t/\Delta)$.
- Förväntat antal lyckade under hela tidsperioden t är $\mathbb{E}(X(n)) = np$.
- **Förväntat antal lyckade försök under t sekunder**:

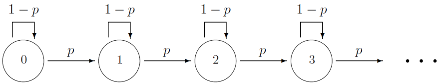
$$\mathbb{E}(X(t/\Delta)) = tp/\Delta,$$

dvs $\lambda = p/\Delta$ förväntat antal lyckade per sekund, också kallad **ankomstfrekvens**.

Binomialprocess

- **Interarrival time** T är **tiden** mellan lyckade försök.
- Y = **antalet** försök mellan två lyckade.
- $Y \sim Geometrisk(p)$
- $T = Y\Delta$. Följer en skalad geometrisk fördelning med support $\Delta, 2\Delta, 3\Delta, \ldots$ Då, $\mathbb{E}(T) = \mathbb{E}(Y\Delta) = \Delta\mathbb{E}(Y) = \Delta/p = 1/\lambda$.
- Binomialprocessen innebär en diskret tid, diskreta tillstånd, Markov- och homogenprocess. Så, den kan representeras som Markovkedja med transitionssannolikheter

$$p_{ij} = \begin{cases} p \text{ om } j = i+1 \\ 1-p \text{ om } j = i \\ 0 \text{ annars} \end{cases}$$

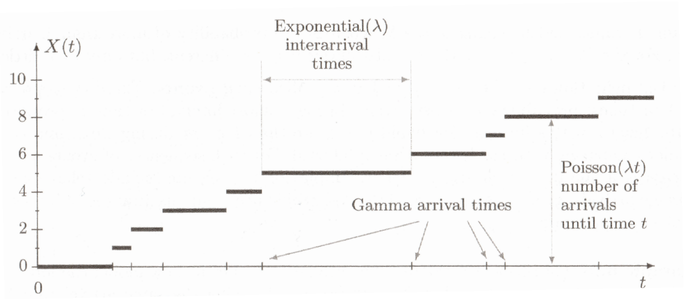


Obs. Markovkedjan är icke-reguljär, dvs ingen stationärfördelning. Obs. flera vägar till 1, 2, ..., vilken innebär binomialfördelningen.

- Notera att Binomialprocessen kan vara restriktiv om Δ väljs för stort, eftersom endast en Bernoullihändelse kan inträffa i varje time frame Δ.
- Se Example 6.18 i Baron.

Poissonprocess

- **Poissonprocessen** fås genom att låta $\Delta \to 0$ samtidigt som λ hålls konstant, dvs $n \to \infty$ och $p \to 0$.
- **Kom ihåg**: $X(t) \sim Binomial(n, p) \to Poisson(\lambda)$ när $n \to \infty$ och $p \to 0$ och $\lambda = np$ är konstant.



- Poissonprocessen är en process i **kontinuerlig tid**.
- Interarrival time $T \sim Exp(\lambda)$. Används vid simulering, se `SimulatePoissonProcess.R`.
- Interarrival för k framtida händelser $T_k \sim Gamma(k, \lambda)$.
- Se Example 6.20 i Baron.

Översikt

- **Stokastiska processer**
- **Markovkedjor**
- **Binomialprocess**
- **Poissonprocess**