

SANNOLIKHETSLÄRA OCH STATISTIK

FÖRELÄSNING 6

Mattias Villani

**Avdelningen för Statistik och Maskininlärning
Institutionen för datavetenskap
Linköpings universitet**



ÖVERSIKT

- ▶ Stokastiska processer
- ▶ Markovkedjor
- ▶ Binomialprocess
- ▶ Poissonproces

STOKASTISKA PROCESSER

- ▶ **Stokastisk process:** En sekvens av slumpvariabler X_1, X_2, \dots, X_T observerade över tid.
- ▶ Ex. X_t = antalet påträffade buggar under dag t , $t = 1, 2, \dots, T$.
- ▶ Ex. X_t = slutkursen på Ericsson aktie vid dag t .
- ▶ Ex. 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 $\mathcal{T} = \{1, 2, 3, \dots\}$
 - ▶ $\omega \in \Omega$, är utfallet i ett experiment (precis som förut).

STOKASTISKA PROCESSER

- ▶ **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 $\mathcal{T} = \{1, 2, 3, \dots\}$
 - ▶ $\omega \in \Omega$, är utfallet i ett experiment (precis som förut).
- ▶ $X(t, \omega)$ kan betraktas på två sätt:
 - ▶ För givet $t \in \mathcal{T}$ är $X_t(\omega)$ en vanlig slumpvariabel
 - ▶ För givet $\omega \in \Omega$ är $X_\omega(t)$ en deterministisk funktion av tiden t .
- ▶ $X_\omega(t)$ kallas för en **realisation** (eng. **sample path** eller trajectory) av processen $X(t, \omega)$.

STOKASTISKA PROCESSER

- ▶ Värderna på $X(t, \omega)$ kallas **tillstånd** (eng. **states**)
- ▶ 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: en robots position vid tidpunkten t .

MARKOVPROCESSER

- ▶ **Markovprocess:** prognosen för morgondagen beror endast på idag:

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

- ▶ **Markovprocess:** för alla tidpunkter $t_1 < \dots < t_n < t$ och händelser A, A_1, \dots, A_n

$$\mathbf{P} \{ X(t) \in A | X(t_1) \in A_1, \dots, X(t_n) \in A_n \} = \mathbf{P} \{ X(t) \in A | X(t_1) \in A_1 \}$$

- ▶ Många processer är inte Markov. Praktiskt **antagande**.

MARKOVKEDJOR

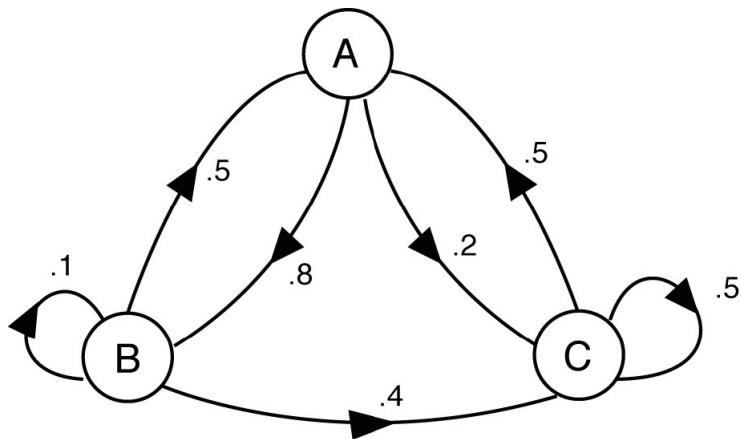
- ▶ **Markovkedja**: diskret tid, diskreta tillstånd.
- ▶ $\mathcal{T} = \{0, 1, 2, \dots\}$ och $\Omega = \{1, 2, \dots, n\}$ (godtyckliga labels, jfr partier).
- ▶ **Transitionssannolikheter (en-steps)**

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

- ▶ **Transitionssannolikheter (h-steps)**

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

MARKOVKEDJOR



MARKOVKEDJOR

- ▶ **Homogen Markovkedja**: transitionssannolikheterna är konstanta över tiden:

$$p_{ij}(t) = p_{ij}$$

- ▶ **Transitionsmatris**

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

- ▶ Ex. två tillstånd, $\Omega = \{\text{sol, regn}\}$

$$\mathbf{P} = \begin{pmatrix} 0.9 & 0.1 \\ 0.7 & 0.3 \end{pmatrix}$$

- ▶ Ex. tre tillstånd, $\Omega = \{\text{RödGröna, Alliansen, SD}\}$

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

MARKOVKEDJOR

► Transitionssannolikheter (h-steps)

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

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

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

- 3-steps transitionssannolikhet $1 \rightarrow 2$:

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

MARKOVKEDJOR

- Transitionsmatris 1-steg

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

- Transitionsmatris h -steg

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

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

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

MARGINALFÖRDELNING

- ▶ Initialfördelning vid $t = 0$ är radvektorn

$$P_0 = (P_0(1), P_0(2), \dots, P_0(n))$$

- ▶ Sannolikhetsfördelning över tillstånden efter h steg (tidsperioder)

$$P_h = (P_h(1), P_h(2), \dots, P_h(n))$$

- ▶ Resultat

$$P_h = P_0 P^h$$

- ▶ Ex. $P_0 = (1/3, 1/3, 1/3)$ och

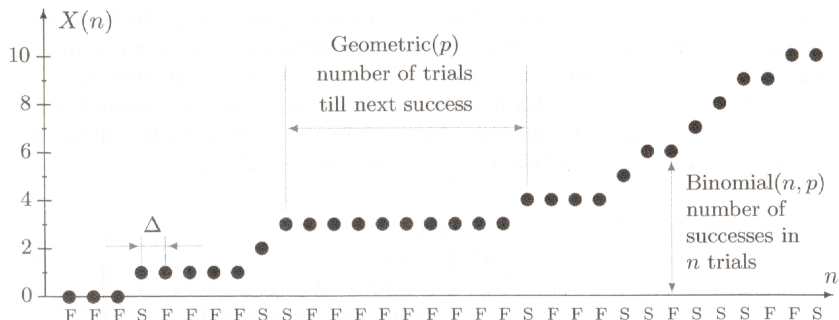
$$P = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.5 & 0.4 \end{pmatrix}$$

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

BERNOULLIPROCESS

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

BERNOULLIPROCESS (BARON S. 149)



BERNOULLIPROCESS

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

$$\mathbb{E}X\left(\frac{t}{\Delta}\right) = \frac{t}{\Delta}p,$$

dvs $\lambda = p/\Delta$ lyckade per sekund.

- ▶ **Ankomstfrekvensen** (arrival rate): $\lambda = p/\Delta$ är **förväntat antal lyckade per tidsenhet** (t ex sekund).

BERNOULLIPROCESS

- ▶ **Interarrival time** T är tiden mellan lyckade försök.
- ▶ Y = antalet försök mellan två lyckade.
- ▶ $Y \sim \text{Geometrisk}(p)$
- ▶ $T = Y\Delta$. Följer en skalad geometrisk fördelning med support $\Delta, 2\Delta, 3\Delta, \dots$
- ▶ Notera att Binomialprocessen kan vara restriktiv om Δ väljs för stort. Endast en Bernoullihändelse i varje time frame Δ .

POISSONPROCESS

- ▶ **Poissonprocessen** fås genom att låta $\Delta \downarrow 0$ samtidigt som λ hålls konstant (dvs även $p \downarrow 0$).
- ▶ Kom ihåg: $X(t) \sim \text{Binomial}(n, p) \rightarrow \text{Poisson}(\lambda)$ när $n \rightarrow \infty$ och $p \rightarrow 0$ och $\lambda = np$ är konstant.
- ▶ Poissonprocessen är en process i **kontinuerlig tid**. Jfr frames i filmer.
- ▶ Interarrival time $T \sim \text{Exp}(\lambda)$. Används vid simulering, see [SimulatePoissonProcess.R](#).
- ▶ Interarrival för k framtida händelser $T_k \sim \text{Gamma}(k, \lambda)$.
- ▶ Se definition 6.14 för alternativ definition av Poissonprocessen.

POISSONPROCESS (BARON s. 155)

