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 $X_1, X_2, ..., X_T$ **observerade över tid**.

- Ex. $X_t =$ antalet påträffade buggar under dag t, $t = 1, 2, ..., T$.
- 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 $\mathcal{T} = \{1, 2, 3, ...\}$
  - $\omega \in \Omega$, 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 $\mathcal{T} = \{1, 2, 3, ...\}$
  - $\omega \in \Omega$, 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

- Värden 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:

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

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

$$\mathbf{P}\left\{X(t) \in A \middle| X(t_1) \in A_1, ...X(t_n) \in A_n\right\} = \mathbf{P}\left\{X(t) \in A \middle| X(t_n) \in A_n\right\}$$

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

Markovkedjor

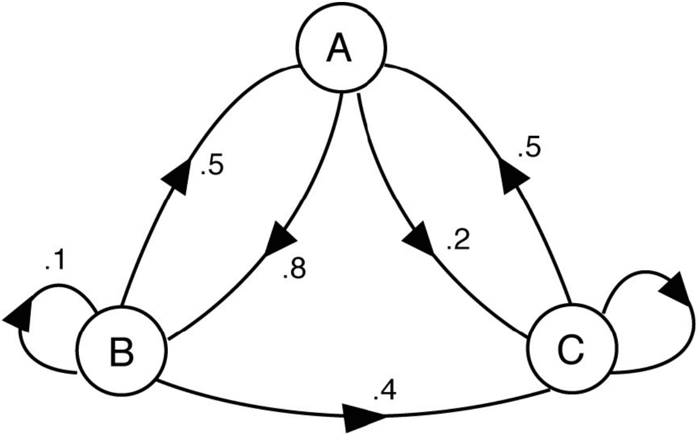
- **Markovkedja**: **diskret tid**, **diskreta tillstånd**.
- $\mathcal{T} = \{0, 1, 2, ...\}$ och $\Omega = \{1, 2, ..., n\}$ (godtyckliga labels, jfr partier).
- **Transitionssannolikheter (en-stegs)**

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

- **Transitionssannolikheter (h-stegs)**

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

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, $\Omega = \{\text{sol}, \text{ regn}\}$
$$\mathbf{P} = \begin{pmatrix} 0.9 & 0.1 \\ 0.7 & 0.3 \end{pmatrix}$$
- Ex. tre tillstånd, $\Omega = \{\text{RödGröna}, \text{Alliansen}, \text{ SD}\}$
$$P = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.5 & 0.4 \end{pmatrix}$$

Markovkedjor

- **Transitionssannolikheter (h-stegs)**

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

- Komplext. Det finns många vägar som tar oss $i \to j$ när $h > 1$.
- Ex: $\Omega = \{1, 2\}$. 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

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

- 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

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

- Sannolikhetsfördelning över tillstånden efter h steg (tidsperioder)

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

- Resultat

$$P_h = P_0 P^h$$

- Ex. $P_0 = (1/3, 1/3, 1/3)$ och

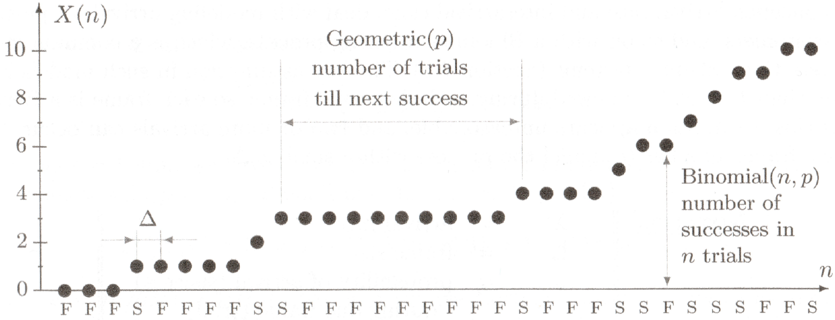
$$P = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.1 & 0.5 & 0.4 \end{pmatrix}$$

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

Bernoulliprocess

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

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. $n = t/\Delta$.
- Processen kan defineras 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 lyckade försök under t sekund**er:

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

  dvs $\lambda = p/\Delta$ förväntat antal 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.
- $Y \sim \mathit{Geometrisk}(p)$
- $T = Y\Delta$. Följer en skalad geometrisk fördelning med support $\Delta, 2\Delta, 3\Delta, ...$
- 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: $X(t) \sim Binomial(n, p) \to Poisson(\lambda)$ när $n \to \infty$ och $p \to 0$ och $\lambda = np$ är konstant.
- Poissonprocessen är en process i **kontinuerlig tid**. Jfr frames i filmer.
- Interarrival time $T \sim Exp(\lambda)$. Används vid simulering, see `SimulatePoissonProcess.R`.
- Interarrvival för k framtida händelser $T_k \sim Gamma(k, \lambda)$.
- Se definition 6.14 för alternativ definition av Poissonprocessen.

Poissonprocess (Baron s. 155)