

## Föreläsning 4: Sannolikhetsmodeller

Om man utför ett experiment  $N$  gånger, under så lika förutsättningar som möjligt, och studerar en händelse  $A$  som kan inträffa, brukar man beskriva sannolikheten att  $A$  händer,  $\mathbb{P}(A)$ , i ett experiment som  $N(A)/N$  där  $N(A)$  är antalet gånger händelse  $A$  inträffade. Detta kommer alltid att vara ett tal mellan 0 och 1.

Rent matematiskt är ett sannolikhetsrum en trippel  $(\Omega, F, \mathbb{P})$  där  $\Omega$  är utfallsrummet,  $F$  en sigma algebra och  $\mathbb{P}$  ett sannolikhetsmått.

En stokastisk process är en sekvens slumpvariabler  $X_1, X_2, \dots$  eller  $\{X_t\}_t$  där  $t$  ofta tolkas som tid. Varje slumpvariabel  $X_i$  kan vara diskret eller kontinuerlig, liksom tiden  $t$  kan vara diskret  $t = 0, 1, 2, \dots$  eller kontinuerlig  $t \in \mathbb{R}_+$ . En sådan process kan beskrivas av en sannolikhetsfördelning

$$\mathbb{P}(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_1 = x_1)$$

Ofta är man mer intresserad av den betingade sannolikheten

$$\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_1 = x_1)$$

**Exempel 1** (Bernoulli processen). *Varje  $X_i$  är oberoende och*

$$X_i = \begin{cases} 1 & \text{med sannolikhet } p \\ 0 & \text{med sannolikhet } 1 - p \end{cases}$$

*En modell för en sekvens slantsinglingar kan då ges av*

$$\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_1 = x_1) = \mathbb{P}(X_n = x_n)$$

*Alltså: utfallet av singling  $n$  är oberoende av utfallet av de  $n - 1$  tidigare utfallen.*

**Exempel 2** (Mynt i en urna). *Betrakta en urna som innehåller 5 enkronor och 5 femkronor. Vi drar nu 3 mynt slumpmässigt och låter  $X_t$  beteckna det totala värdet av de dragna mynten. Exempelvis kan vi dra  $\{5, 5, 1\}$ , och då har vi  $X_3 = 11$ . I detta fall är historiken avgörande:*

$$\mathbb{P}(X_4 = x_4 \mid X_3 = x_3, X_2 = x_2, X_1 = x_1)$$

### Markovkedjor

I många fall är man intresserad av ett mellanting, där nästa tillstånd beror på nuvarande tillstånd, men inte tillstånd längre bak i tiden! Formellt har vi

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_1 = x_1) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

## Diskussion

Ge exempel på en Markov och en icke Markov process!

Om utfallsrummet är diskret kan processen beskrivas med hjälp av övergångssannolikheter:

$$p_{i,j} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

där vi antar att detta är oberoende av tiden  $n$ . Vi kan tolka  $p_{i,j}$  som elementen i en s.k. övergångsmatris  $P$ .

**Exempel 3** (Social rörlighet). *Vi delar in Sveriges befolkning efter inkomst i tre kategorier, låg, medel samt höginkomsttagare. Antag att vi vet sannolikheten att föräldrar i en kategori får barn som är i respektive kategori. Exempelvis kan övergångsmatrisen ges av*

$$P = \begin{pmatrix} 0.65 & 0.28 & 0.07 \\ 0.15 & 0.67 & 0.18 \\ 0.12 & 0.36 & 0.52 \end{pmatrix}$$

*Detta kan också representeras med en graf. Observera att ingen social historia finns från tidigare generationer!*

**Exempel 4** (Moran processen). *Moran processen är en stokastisk process som används inom biologi för att beskriva ändliga populationer av exempelvis förekomsten av gener förändras över tid, under följande antaganden:*

- *Konstant populationsstorlek  $N$*
- *Två alleler:  $A$  och  $B$ , med  $N_A + N_B = N$*
- *Ingen reproduktiv skillnad (ingen fördel/nackdel)*
- *Inga mutationer*

*I varje tidssteg dör 1 individ, och ersätts av någon annan individ (alla har lika stor chans). Vi har då övergångssannolikheter ( $P_{i,j}$  beskriver sannolikheten att gå från att ha  $i$  av typ  $A$ , till att ha  $j$  av typ  $A$ ).*

$$\begin{aligned} P_{i,i-1} &= \frac{N-i}{N} \frac{i}{N} \\ P_{i,i+1} &= \frac{i}{N} \frac{N-i}{N} \\ P_{i,i} &= 1 - P_{i,i+1} - P_{i,i-1} \end{aligned}$$

Vidare har vi  $P_{0,1} = 0 = P_{N,N-1}$  eftersom när en allel är borta kan den inte komma tillbaka (ingen mutation).

Nu får vi resultatet att sannolikheten att  $A$  fixeras är  $\mathbb{P}(A \text{ fixeras} \mid N_A(0) = i) = \frac{i}{N}$ . (Alla har samma sannolikhet att bli ensam stamfader till populationen, och det finns "i" st. av typ  $A$  vid start).

Medeltiden till fixering ges av  $k_i = N \left[ \sum_{j=1}^i \frac{N-i}{N-j} + \sum_{j=i+1}^{N-1} \frac{i}{j} \right]$ , om  $N_A(0) = i$ .

#### Diskussion

Hur kan man modifiera modellen om man vill inkludera exempelvis mutation och fitness?

## Kontinuerlig tid

Vi kan också beskriva system som förändras i kontinuerlig tid. Övergångssannolikheter ersätts då med övergångsintensiteter. Låt  $X(t)$  vara tillståndet av vår process vid tid  $t$ .

$$\mathbb{P}(X(t + \Delta t) = j \mid X(t) = i) = \begin{cases} q_{i,j} \Delta t & \text{om } i \neq j \\ 1 - v_i \Delta t & \text{om } i = j \end{cases}$$

Där vi definierar  $v_i = \sum_j q_{i,j}$ . Detta innebär att  $q_{i,j} \Delta t$  beskriver sannolikheten att gå från tillstånd  $i$  till  $j$  i ett tidsintervall av längd  $\Delta t$ , och  $v_i$  att vi går från  $i$  till  $i$ .

**Exempel 5** (Populationsdynamik). Antag att varje bakterie delar sig med intensitet  $\lambda$ , och  $N(t)$  är antalet bakterier. Då har vi

$$q_{i,j} = \begin{cases} \lambda \cdot i & \text{om } j = i + 1 \\ 0 & \text{annars} \end{cases}$$

Genom att låta  $P_j(t) = \mathbb{P}(X(t) = j)$  och använda ovanstående har vi

$$P_j(t + \Delta t) = P_j(t)(1 - v_j \Delta t) + \sum_{i \neq j} P_i(t) q_{i,j} \Delta t$$

$$\frac{P_j(t + \Delta t) - P_j(t)}{\Delta t} = -P_j(t) v_j + \sum_{i \neq j} P_i(t) q_{i,j}$$

och genom att låta  $\Delta t \rightarrow 0$  erhåller vi

$$\frac{dP_j}{dt} = -P_j(t) v_j + \sum_{i \neq j} P_i(t) q_{i,j}$$

som brukar kallas Kolmogorovs framåt ekvation.

**Exempel 6** (Bakterier igen). Med  $v_j = \sum_{i \neq j} q_{i,j} = \lambda j$  har vi

$$P'_j(t) = -v_j P_j(t) + \sum_{i \neq j} P_i(t) q_{i,j}$$

$$P'_j(t) = -j\lambda P_j(t) + (j-1)\lambda P_{j-1}(t)$$

Detta är en ekvation vi kan lösa och får då

$$P_j(t) = e^{-\lambda t} (1 - e^{-\lambda t})^{j-1}$$

Från detta kan man studera till exempel medel och varians hos fördelningen!