MVE550 Stockastiska processer och Bayesiansk inferens

Föreläsning 5/11

- Science - making models for part of reality
  - make predictions
- The model is a model of knowledge
- Deterministic and { stochastic models.
  ↙ { = probabilitic model

It predicts a probability distribution

- a mathematical formula for computing the probability of an event.
- a computer program generating different outputs each time.

Monte Carlo simulations

Frequency of an output ≈ probability of that output.

- We can calculate with uncertainty
- make probability predictions based on uncertain information

Stochastic processes

A collection of random variables $\{X_t, t \in I\}$

The random variables are defined on a common state space S.

Review a random variable X with a state space S.

Is a Real-valued function on S, togheter with a probability P on S.

- $0 \leq P(A) \leq 1$ for any measurable subset $A \subseteq S$
- $P(S) = 1$, $P(\emptyset) = 0$
- $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, $A_1, A_2, \dots$ disjoint subset of S

A random variable X is a function $X: S \to \mathbb{R}$ from a probability space S to $\mathbb{R}$

A discrete random variable:

- The function takes on a finite or contable set of numbers in $\mathbb{R}$.

  står ofta f i boken

  $\downarrow$ $f_X(x)$
- The probability of $x \in \mathbb{R}$, written $\underline{\tilde{\pi}(x)}$, is $P(X^{-1}(x))$
- $\tilde{\pi}(x)$. The probability mass function
  - $\sum_{\text{All } x} \tilde{\pi}(x) = 1$

A continuous random variable:

- The function $X \to \mathbb{R}$ has values in (most often) an interval of real numbers.
- The probability density function $\tilde{\pi}(x)$ is defined by requireinc:

  $$\int_{-\infty}^{y} \tilde{\pi}(x)\,dx = P(X^{-1}((-\infty, y))) \quad (\text{when } \tilde{\pi} \text{ is continuous})$$
- In dobrow: $f_X(x)$ for $\tilde{\pi}_X(x)$
- $\int_{-\infty}^{\infty} \tilde{\pi}(x)\,dx = 1$ $\left\{ \text{we define } \int \tilde{\pi}(x)\,dx \text{ to mean } \sum \tilde{\pi}(x) \text{ when } x \text{ is discrete} \right\}$

Conditional distributions:

Reviews $A, B \subseteq S$ (Event)

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(A \cap B) = P(B)P(A|B)$$

For discrete random variables:

$$\tilde{\pi}(x|y) = \frac{\tilde{\pi}(x,y)}{\tilde{\pi}(y)} \quad \Big| \quad \tilde{\pi}(x,y) = \tilde{\pi}(y)\,\tilde{\pi}(x|y)$$

$$\tilde{\pi}(y) = \int \tilde{\pi}(x,y)\,dx \quad \left(\sum_x \tilde{\pi}(x,y)\right)$$

For continous random variables:

$$\pi(x|y) = \frac{\pi(x,y)}{\pi(y)} = \frac{\pi(x,y)}{\int \pi(x,y)\,dx}$$

$$\pi(x,y,z,w,u) = \pi(x,y|z,w,u)\,\pi(z,w,u)$$

$$\pi(x,y|z,w,u) = \frac{\pi(x,y,z,w,u)}{\pi(z,w,u)}$$

$$\pi(x,y,z,w|u,v) = \pi(x,y|z,w,u,v) = \pi(z,w|u,v)$$

Expectation:

$$E(x) = \int_{-\infty}^{\infty} x\,\pi(x)\,dx \qquad E_x(x)$$

$$E_x(g(x)) = \int g(x)\,\pi(x)\,dx$$

$$E_{x,y}(g(x,y)) = \iint g(x,y)\,\pi(x,y)\,dx\,dy$$

$$E_{x|y}(g(x,y)) = \int g(x,y)\,\pi(x|y)\,dx$$

$$E_y[E_{x|y}[g(x,y)]] = \int\left[\int g(x,y)\,\pi(x|y)\,dx\right]\pi(y)\,dy =$$

$$= \iint g(x,y)\,\pi(x|y)\,\pi(y)\,dx\,dy = \iint g(x,y)\,\pi(x,y)\,dx\,dy =$$

$$= E_{x,y}[g(x,y)]$$

- $Var_x(x) = E_x[(x - E_x(x))^2] = E_x[x^2 - 2xE_x(x) + E_x(x)^2] = E_x(x^2) - E_x(x)^2$

- $Var_{xy}(g(x,y)) = E_{x,y}[g(x,y) - E_{x,y}[g(x,y)^2] = E_{xy}[g(x,y)^2] - E_{xy}[g(x,y)]^2$

$$Var_{x,y}[g(x,y)] = Var_y[E_{x|y}(g(x,y))] + E_y(Var_{x|y}(g(x,y)))$$

$$\left.\begin{array}{l} Var_y(E_{x|y}(g(x,y))) = E_y(E_{x|y}(g(x,y))^2] - E_y[E_{x|y}(g(x,y))]^2 \\ E_y(Var_{x|y}(g(x,y))) = E_y(E_{x|y}(g(x,y)^2)) - E_y[E_{x|y}(g(x,y))^2] \end{array}\right\} \quad \heartsuit$$

$$\heartsuit \quad E_y[E_{x|y}(g(x,y)^2)] - E_y[E_{x|y}(g(x,y))]^2$$

räknestuga 5/10

(1.10) En tärning rullas tills vi får en 3:a. Genom att betingen på första utfallet. Vad är sannolikheten att det kommer krävas ett jämt antal kast?

Lösning: Introducera:

A : Det krävs ett jämt antal kast

B: Vi får en 3:a på direkten

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

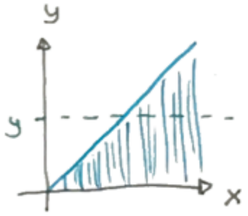
$P(A|B) = 0$, $P(A|B^c) = 1 - P(A)$ sannolikheten att det blir udda antal kast \ första kastet

$$P(A) = 0 \cdot \frac{1}{6} + (1 - P(A)) \cdot \frac{5}{6} = \frac{5}{11}$$

(1.14) Vi har X, Y , $f(x,y) = 4e^{-2x}$, $0 < y < x < \infty$

a) Hitta $f_{X|Y}(X|Y)$

b) Hitta $f_{Y|X}(Y|X)$



$$f_Y(y) = \int_y^{\infty} f(x,y)\,dx$$

Lösning: $f_{X|Y}(X|Y) = \dfrac{f(x,y)}{f_Y(y)}$

a) $f_Y(y) = \int_y^{\infty} f(x,y)\,dx = \int_y^{\infty} 4e^{-2x}\,dx = \left[-2e^{-2x}\right]_y^{\infty} = 2e^{-2y}$

$f_{X|Y}(X|Y) = \dfrac{4e^{-2x}}{2e^{-2y}} = 2e^{2y-2x} \Rightarrow x > y$ se bild

b) $f_Y(x) = \int_0^{Y} 4e^{-2x}\,dY = 4xe^{-2x}$

$f_{Y|X}(Y|X) = \dfrac{4e^{-2x}}{4xe^{-2x}} = \dfrac{1}{x} \quad 0 < Y < x$

(1.16) En pokerhand har 5 kort från 52 st
Vad är det förväntade antal äss givet att det första kortet var ett äss?

Lösning: Y: Antal äss

A: Första kortet är ett äss

sökt: $E(Y/A)$

$$E(Y/A) = \sum k \cdot P(Y=k|A) \quad , \quad P(Y=k|A) = \frac{\binom{3}{k-1}\binom{48}{5-k}}{\binom{51}{4}}$$

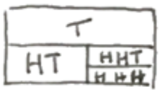
(1.20) Ett rättvist mynt singlas om och om igen

a) Förväntat antal singlingar tills vi får 3 heads på rad

b) Förväntat antal singlingar tills vi får k heads på rad

Lösning: Y: Antal singlingar tills 3 heads

a) $$\mathbb{E}(Y) = \mathbb{E}(Y|T)P(T) + \mathbb{E}(Y|HT)P(HT) + \mathbb{E}[Y|HHT)P(HHT) + $$
$$+ \mathbb{E}(Y|HHH)P(HHH)$$ viktigt med reglerna till detta



$$\mathbb{E}(Y|T) = 1 + \mathbb{E}(Y)$$
$$\mathbb{E}(Y|HT) = 2 + \mathbb{E}(Y)$$
$$\mathbb{E}(Y|HHT) = 3 + \mathbb{E}(Y)$$
$$\mathbb{E}(Y|HHH) = 3$$

$$\mathbb{E}(Y) = (1+\mathbb{E}(Y)) \cdot \frac{1}{2} + (2+\mathbb{E}(Y)) \cdot \frac{1}{4} + (3+\mathbb{E}(Y)) \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} = 14$$

b) Y: Antal singlingar

$$\mathbb{E}(Y) = \sum_{i=1}^{k} (i + \mathbb{E}(Y)) \cdot \left(\frac{1}{2}\right)^i + k \cdot \left(\frac{1}{2}\right)^k = \ldots = 2 - 2^{1-k} + \mathbb{E}(Y)(1-2^k)$$

$$\left\{ \sum_{i=0}^{k} r^i = \frac{1-r^{k+1}}{1-r} \quad , \quad \frac{d}{dr}\left(\sum r^i\right) = \sum i r^{i+1} \right\}$$

(1.24) Visa $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$, X, Y kontinuerliga

Lösning: $\mathbb{E}(\mathbb{E}(Y|X)) = \int_{\mathbb{R}} \mathbb{E}(Y|X=x) f_X(x)\, dx = \{\mathbb{E}(g(x)) = \int g(x) f_x(x) dx\}$

$= \int \left[\int y \cdot f_{Y|X}(y|x) dy \right] f_X(x)\, dx = \int y \int f_{Y|X}(y|x) f_X(x)\, dx dy =$

$= \int Y \underbrace{\int f(x,y)\, dx}_{f_Y(Y)} dy = \int y f_Y(y)\, dy = \mathbb{E}[Y]$

(1.30) $Y = g(X)$, förenkla

a) $\mathbb{E}[Y|X]$ b) $Var(Y|X)$

Lösning: a) $\mathbb{E}[g(x) | X = x] = \mathbb{E}[g(x) | X = x] = g(x)$

$\Rightarrow \mathbb{E}[Y|X] = g(X)$

b) $Var(g(x) | X) = Var(g(x) | X = x) = Var(g(x) | X = x)$

Då de håller för alla x $\Rightarrow Var(Y|X) = 0$

Föreläsning 7/11

Ways of building stochastic models from data:

Frequentist (Classical)

- Define stochastic model with unknown parameter θ.
- Use data to estimate θ. $\hat{\theta}$
- Plug in $\hat{\theta}$ in the model make predictions.

Bayesian:

- Define stochastic model including variable. Representing data y and what you want to predict y_{new}.
- Compute the conditional distribution $y_{new} | y$
- Use this for predictions.

Example: (Dice rolls)

$\boxed{1, 1, 2, 1, 6, 1, 1}$, ②, 2, 1, 3, 5, 5, 1, 4, 2

Simpler example:

Probability for heads: 0.7 or 0.3

Both possibilites are equally likely

Probability of observing k heads in n throws

$$\pi(k) = 0.5 \text{ binomal}(k; n, 0.7) + 0.5 \text{ binomal}(k, n, 0.3)$$

$$\pi(\text{heads} \mid \text{sequence with } y_H \text{ heads and } y_T \text{ tails}) =$$

$$= \frac{\pi(\text{sequence and then heads})}{\pi(\text{sequence})} = \frac{0.5 \cdot 0.7^{y_H+1} \cdot 0.3^{y_T} + 0.5 \cdot 0.3^{y_H+1} \cdot 0.7^{y_T}}{0.5 \cdot 0.7^{y_H} \cdot 0.3^{y_T} + 0.5 \cdot 0.3^{y_H} \cdot 0.7^{y_T}}$$

prior

$$\theta \sim \text{Bernoulli}(0.5, 0.7, 0.3) \quad , \quad k | \theta \sim \text{Binomial}(n, \theta)$$

$y \cdot y = 1$ means that n+1'st throw is heads

$$\pi(y|k) = \sum_{\theta = 0.3 \cdot 0.7} \pi(y \cdot \theta | k) = \sum_{\theta = 0.3 \cdot 0.7} \pi(y | \theta, k) \pi(\theta | k) = \sum_{\theta = 0.3 \cdot 0.7} \pi(y | \theta) \pi(\theta, k)$$

$\theta | k$: posterior

$\pi(y = 1 | \theta) = \theta$

Bayes formula:

$$\pi(\theta | k) = \frac{\pi(k | \theta) \pi(\theta)}{\pi(k)} = \frac{\pi(k | \theta) \pi(\theta)}{\sum_{G} \pi(k | \theta) \pi(\theta)} \longrightarrow$$

In our case:

$$\pi(\theta|k) = \dots = \frac{\text{Binomial}(k, n, \theta)\, 0.5}{\text{Binomial}(k, n, 0.7)\, 0.5 + \text{Binomial}(k, n, 0.3)\, 0.5} =$$

$$= \frac{\binom{n}{k} \theta^k (1-\theta)^{n-k}}{\binom{n}{k} 0.7^k \cdot 0.3^{n-k} + \binom{n}{k} \cdot 0{,}3 \cdot 0{,}7^{n-k}}$$

Next model:

$\theta \sim \text{uniform}(0,1)$

$k|\theta \sim \text{binomial}(n, \theta)$

$$\pi(\theta|k) = \frac{\pi(k|\theta)\pi(\theta)}{\pi(k)} = \frac{\pi(k|\theta)\pi(\theta)}{\int_\theta \pi(k|\theta)\pi(\theta)\, d\theta} = \frac{\text{Binomial}(k,n,\theta)}{\int \text{Binomial}(k,n,\theta)\, d\theta} =$$

$$= \frac{\binom{n}{k} \theta^k (1-\theta)^{n-k}}{\int \binom{n}{k} \theta^k (1-\theta)^{n-k} d\theta} = \boxed{\frac{\theta^k (1-\theta)^{n-k}}{B(k+1, n-k+1)}}$$

This is a beta density!

$\text{Beta}(\theta, k+1, n-k+1)$

$$\pi(y|k) = \int \pi(y|\theta)\pi(\theta|k)\, d\theta = \int \theta \,\text{Beta}(\theta, k+1, n-k+1)\, d\theta =$$

$$= \frac{k+1}{k+1+n-k+1} = \frac{k+1}{n+2}$$

$\propto_\theta$ proportional to $\quad \theta \propto_\theta 3\theta \quad \frac{1}{\theta+1} \propto_\theta \frac{\alpha}{\theta+1}$

$$\pi(\theta|k) \propto_\theta \pi(k|\theta)\pi(\theta) = \text{Binomial}(k,n,\theta) \propto_\theta \theta^k (1-\theta)^{n-k}$$

This must be Beta(k+1, n-k+1) density!

Next example:

$\theta \sim \text{Beta}(\alpha, \beta)$, $k|\theta \sim \text{Binomial}(k, n, \theta)$

$$\pi(\theta|k) \propto_\theta \pi(k|\theta)\pi(\theta) \propto_\theta \text{Binomial}(k,n,\theta)\, \text{Beta}(\theta, \alpha, \beta)$$

$$\propto_\theta \theta^k (1-\theta)^{n-k} \theta^{\alpha-1}(1-\theta)^{\beta-1} = \theta^{\alpha+k-1}(1-\theta)^{\beta+n-k-1}$$

prediction is the expectation of $\longrightarrow$ Beta$(\alpha+k, \beta+n-k)$

$$\frac{\alpha+k}{\alpha+k+\beta+n-k} = \frac{\alpha+k}{\alpha+\beta+n}$$

Conjugacy

Given density $\pi(y|\theta)$ so that for all priors for θ.
In a family of densitites, the posterior is also in the same family => conjugate family **Beta** distribution is conjugate to the **binomial** **likelyhood**

k: number of requests per time unit

θ: expected number of requests for this time unit

Poisson likelyhood: $\pi(k|\theta) = \ell^{-\theta} \frac{\theta^k}{k!}$ $\quad k \in \{0,1,2,\dots\}$
$\theta \in [0, \infty)$

Prior on θ with "No information"

Improper densities: Do not integrate to 1

$\pi(\theta) \propto_\theta 1$

$\pi(\theta) \propto_\theta \frac{1}{\theta}$ Improper

$\pi(\theta|k) \propto_\theta \pi(k|\theta)\pi(\theta) \propto_\theta \ell^{-\theta} \cdot \theta^k \cdot \frac{1}{\theta} = \ell^{-\theta}\theta^{k-1}$ this is a gamma(k, 1
density

Try with prior $\theta \sim \text{Gamma}(\theta, \alpha, \beta)$

$\pi(\theta|k) \propto_\theta \pi(k|\theta)\pi(\theta) \propto_\theta \ell^{-\theta} \cdot \theta^k \cdot \theta^{\alpha-1} \cdot \exp(-\beta\theta) = \theta^{\alpha+k-1} \exp(-(\beta+1)\theta)$

Gamma$(\alpha+k, 1+B)$

The gamma densities are conjugate to the poisson likelyhood.

To compute "The prior predictive"

$$\pi(k) = \left\{\int \pi(k|\theta)\pi(\theta)\,d\theta\right\} = \frac{\pi(k|\theta)\pi(\theta)}{\pi(\theta|k)}$$

$$\pi(k) = \frac{\pi(k|\theta)\pi(\theta)}{\pi(\theta|k)} = \frac{\text{poisson}(k;\theta)\,\text{gamma}(\theta;\alpha,\beta)}{\text{gamma}(\theta,\alpha-k,1+\beta)} = \frac{\ell^{-\theta}\frac{\theta^k}{k!}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}\exp(-\beta\theta}{\frac{(\beta+1)^{\alpha+k}}{\Gamma(\alpha-k)}\theta^{\alpha+k-1}\exp(-(\beta+1)\theta}$$

$$= \frac{\beta^\alpha\Gamma(\alpha+k)}{(\beta+1)^{\alpha+k}\Gamma(\alpha)\cdot k!}$$ negative binomial$\left(\alpha, \frac{1}{1+\beta}\right)$

posterior predictive: $\pi(k_{new}|k) = \int \pi(k_{new}|\theta)\pi(\theta|k)\,d\theta$

räknestuga 9/11

(1.3) Vikter till producerade objektet är normalfördelade med $\mu = 5.4$: $\sigma = 4$.

Den observerade vikten har fordelningen $N(w_i, 1)$ (w sanna vikten)

Vi observerar $y = 5.9$, vad är sannolikheten att $w > \sigma_0$?

Lösning: Introducerar:

W_i : Den sanna vikten
Y: Den observerade vikten

$W \sim N(59.4^2) \quad Y|W \sim N(W, 1^2)$

Vi vill ha fordelningen till $W|Y$

$W \sim N(\mu_0, t_0^{-1}) \quad t_0 = \frac{1}{\sigma^2}$

$Y|W \sim N(W, t^{-1})$

$$p(w|y) \propto_w p(y|w)p(w) = \exp\left(-\frac{t}{2}(y-w)^2\right)\exp\left(-\frac{t_0}{2}(w-\mu_0)^2\right) =$$

$$= \exp\left(-\frac{t}{2}(y^2 - 2yw + w^2) - \frac{t_0}{2}(w^2 - 2w\mu_0 + \mu_0^2)\right)$$

$$\propto_w \exp\left(-\frac{t}{2}(-2yw + w^2) - \frac{t_0}{2}(w^2 - 2w\mu_0)\right) =$$

$$= \exp\left(-w^2\left(\frac{t}{2} + \frac{t_0}{2}\right) + (yt + t_0\mu_0)w =\right.$$

$$= \exp\left(\frac{-t + t_0}{2}\left(w^2 - \frac{2yt + 2t_0\mu_0}{t_0 + t}w\right)\right) \propto_w \exp\left(-\frac{t + t_0}{2}\left(w - \frac{yt + t\mu_0}{t_0 + t}\right)^2\right) =$$

$$= w|y \sim N\left(\frac{yt + t_0\mu_0}{t_0 + t}, \frac{1}{t_0 + t}\right)$$

In our case $y = 59$, $t = 1$, $t_0 = \frac{1}{4^2}$, $\mu_0 = 54$

$\Rightarrow \quad w|y = 51 \sim N\left(\frac{998}{17}, \frac{16}{17}\right)$

$P(w > 60 | y = 59) \approx 0.091)$

So we have a coin, which is not fair.

We don't know if pobability of heads in 0.7 or 0.3

We belive these two cases to be equally likely.

a) What is the probability of y_H heads in a total of $y_H + y_T$ throws?

Løsning: Introduce Y: The number of heads

p_H: The probability of heads

We want the $P(Y = y_H) = \left\{ \begin{matrix} \text{use Law of Total} \\ \text{probability} \end{matrix} \right\} =$

$$= P(Y = y_H \mid p_H = 0.3) P(p_H = 0.3) + P(Y = y_H \mid p_H = 0.7) P(p_H = 0.7)$$

$$P(Y = y_H \mid p_H = 0.3) = \binom{y_H + y_T}{y_H} 0.3^{y_H} \cdot 0.7^{y_T}$$

$$P(Y = y_H \mid p_H = 0.7) = \binom{y_H + y_T}{y_H} 0.7^{y_H} \cdot 0.3^{y_T}$$

$$P(p_H = 0.3) = P(p_H = 0.7) = \frac{1}{2}$$

$$P(Y = y_H) = \binom{y_H + y_T}{y_H} \left(\frac{0.3^{y_H} \cdot 0.7^{y_T} + 0.7^{y_H} \cdot 0.3^{y_T}}{2} \right)$$

b) what is the probability of getting y_H heads, y_T tails and the last flipp should be heads.

Independence $P(A \cap B) = P(A)P(B)$

Løsning: $P(Y = y_H \cap H)$

H: The last throw is a head

Use the law of probability to condition on the probability of getting heads.

$$(♡): P(Y = y_H \cap H) = P(Y = y_H \cap H \mid p_H = 0.3) P(p_H = 0.3) + P(Y = y_H \cap H \mid p_H = 0.7) \cdot P(p_H = 0.7)$$

We have: $P(Y = y_H \cap H \mid p_H = 0.3) =$

$$= P(Y = y_H \mid p_H = 0.3) P(H \mid p_H = 0.3) \text{ by independence condition}$$

With ans in a

$$\Longrightarrow (♡) = \binom{y_H + y_T}{y_H} \frac{0.3^{y_H} \cdot 0.7^{y_T}}{2} (0.3) + \binom{y_H + y_T}{y_H} \frac{0.7^{y_H} \cdot 0.3^{y_T}}{2} \cdot 0.7$$

c) What is the probability of getting a head after y_H heads and y_T tails

We want $P(H \mid Y = y_H)$

Lösning: $P(H \mid Y = y_H) = \dfrac{P(H \cap Y = y_H)}{P(Y = y_H)}$

By a & b $\Rightarrow$ $P(H \mid Y = y_H) = \ldots = \dfrac{0.3^{y_H+1} \cdot 0.7^{y_T} + 0.7^{y_H+1} \cdot 0.3^{y_T}}{0.3^{y_H}\, 0.7^{y_T} + 0.7^{y_H}\, 0.3^{y_T}}$

Föreläsning 12/11

If v is a vector describing the distribution of states i, then vP^{n} is the vector describing the distribution of states at stage $i+n$

Thus the probability to go from state i to state j in n steps is given by $(P^n)_{ij}$ (fortsättning, se slides)

$$[v_1 \ \ldots \ v_n] \cdot \begin{bmatrix} P_{11} & \\ P_{21} & \\ \vdots & \\ P_{n1} & \end{bmatrix} = [0 \ \ldots \ 0]$$

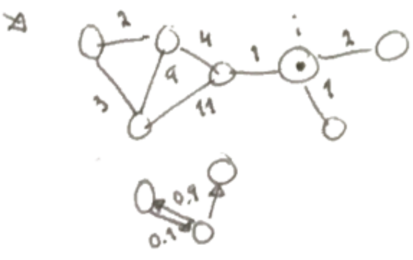
(Limiting distribution):

$$E\left(\frac{1}{n}\sum_{l=0}^{n-1} I(X_l = j) \mid X_0 = i\right)$$

$$\lim_{n\to\infty} \Rightarrow \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} E(I(X_k = j) \mid X_0 = i) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} (P^k)_{ij} =$$

$$= \lim_{n\to\infty} (P^n)_{ij}$$

transisgraf



$$\sum_{j\sim i} \frac{\deg(j)}{\sum_n \deg(n)} \cdot \frac{1}{\deg(j)} = \sum_{j\sim i} \frac{1}{\sum_n \deg(n)} = \frac{\deg(i)}{\sum_n \deg(n)}$$

$\deg(j)$ = number of edges into j

övning 12/11

(2.2) $X_0, X_1, \dots$ is a Markov chain $\alpha = (1/2, 0, 1/2)$

$$P = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 0 & 1/2 & 1/2 \\ 2 & 1 & 0 & 0 \\ 3 & 1/3 & 1/3 & 1/3 \end{array}$$

a) $P(X_2 = 1 | X_1 = 3) = \underline{1/3}$

b) $P(X_1 = 3, X_2 = 1) = \underbrace{P(X_2 = 1 | X_1 = 3)}_{\text{from a)}} P(X_1 = 3)$

$P(X_1 = 3) = \sum P(X_1 = 3 | X_0 = k) P(X_0 = k) = \sum_k P_{k3} \alpha_k = (\alpha \cdot P)_3$

$\Rightarrow P(X_1 = 3, X_2 = 1) = P(X_2 = 1 | X_1 = 3)(\alpha \cdot P)_3 = \underline{5/36}$

c) $P(X_1 = 3 | X_2 = 1) = \dfrac{P(X_1 = 3, X_2 = 1)}{P(X_2 = 1)}$

$P(X_2 = 1) = \sum \underbrace{P(X_2 = 1 | X_0 = k)}_{\text{we need } P^2} P(X_0 = k) = (\alpha P^2)_1 = \underline{\underline{5/9}}$

$$P^2 = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 2/3 & 1/6 & 1/6 \\ 2 & 0 & 1/2 & 1/2 \\ 3 & 4/9 & 5/18 & 5/18 \end{array}$$

d) $P(X_9 = 1 | \underbrace{X_1 = 3, X_4 = 1}_{\text{kan skippa då } X_7 \text{ har infon vi behöver}}, X_7 = 2) = P(X_9 = 1 | X_7 = 2) = 0$

pga X_{21} i matris
är noll

(2.7) $X_0, X_1, \dots$ is a Markov chain with transition matrix P

Let $Y_n = X_{3n}$ for $n = 0, 1, 2, \dots$

Show Y is a Markov chain and find the transition matrix

Lösning: $P(Y_n = j | Y_n = i, Y_{n-2} = i_2, \dots, Y_0 = i_0) = P(X_{3n} = j | X_{3n-3} = i,$
$X_{3n-6} = i_2, \dots, X_0 = i_0) = P(X_{3n} = j | X_{3n-3} = i) = (P^3)_{ij}$

$\Rightarrow$ The transition matrix for Y_n is given P^3

2.16 Assume P is a stochastic matrix with equal rows. We want to show $P^n = P$ $\forall n \geq 1$

$P_{ij} = p_j$

Induction:

Base.case $n = 1$

Assume it is true for n => it is time for $n+1$

$$(P^{n+1})_{ij} = \sum_k P^n_{ik} P_{kj} = \dots = p_j \sum_k P^n_{ik} = P_j$$

=> All rows are equal for P^{n+1}

2.14 There are k songs on a music player. The songs are chosen randomly with replacement.

Introduce: X_0: the number of unique songs that we have heard after n songs

Is X_n a Markov chain?

Give the transition matrix

Løsning: $P(X_n = 1 \mid X_{n-1} = j, X_{n-2} = i_0, \dots X_0 = 0) = P(X_n = i \mid X_{n-1} = j) =$

k songs
j already heard

$$= \begin{cases} j/k & i = j \leftarrow \text{the next we hear have already been played} \\ \underset{\uparrow \text{ new song}}{\frac{k-j}{k}} & i = j+1 \end{cases}$$

=> so this gives the transition matrix

$P_{ij} = P(X_n = j \mid X_{n-1} = i)$

b) If $k = 4$, What is the probability that we have heard all songs after 6 plays

We want $P(X_6 = 4) = P(X_6 = 4 \mid X_0 = 0) P(X_0 = 0) =$

disapears

$(\sum P(X_6 = 4 \mid X_0 = k) P(X_0 = k))$

$= P(X_6 = 4 \mid X_0 = 0) \cdot 1$

unique songs

$P(X_0 = 0) = 1$

we want P^6

$$P = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1/4 & 3/4 & 0 & 0 \\ 2 & 0 & 0 & 2/4 & 2/4 & 0 \\ 3 & 0 & 0 & 0 & 3/4 & 1/4 \\ 4 & 0 & 0 & 0 & 0 & 1 \end{array} \leftarrow \text{songs heard}$$

Calculate P^6 on a computer

$$P^6 = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & & & & & \boxed{\frac{195}{512}} \\ 1 & & & & & \\ 2 & & & & & \\ 3 & & & & & \\ 4 & & & & & \end{array}$$

$P(X_6 = 4 \mid X_0 = 0) = 195/512$

2.20 $X_0, X_1, \ldots$ is a Markov chain with transition matrix.

$$P = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \\ 3 & p & 1-p & 0 \end{array} \qquad 0<p<1 \qquad g(x) = \begin{cases} 0 & x=1 \\ 1 & x=2,3 \end{cases}$$

$Y_n = g(X_n)$, show Y_n not a Markov chain

We want to show that the future is of all past history

Investigate: $P(Y_2 = 0 \mid Y_0 = 0, Y_1 = 1) = P(g(X_2) = 0 \mid g(X_0) = 0, g(Y_1) = 1) =$

$= P(X_2 = 1 \mid X_0 = 1, X_1 \in \{2, 3\}) = P(X_2 = 1 \mid X_0 = 1, X_1 = 2) = 0$

Investigate: $P(Y_2 = 0 \mid Y_0 = 1, Y_1 = 1) = P(X_2 = 1 \mid X_0 \in \{2,3\}, X_1 \in \{2,3\}) =$
$P(X_2 = 1 \mid X_0 = 2, X_1 = 3) = p$

$\Rightarrow$ Since Y_n depends not only on the most recent past it is not a marker chain.

Förelåsning 14/11

A markov chain is defined as this: you can't predict the future and dosen't depend on the past.

ex: (not a markov chain)

kan komma tillbaka ↓

Recurrence and transience

$$\mathbb{E}\left(\sum_{n=0}^{\infty} I(X_n=j) \mid X_0=i\right) = \sum_{n=0}^{\infty} \mathbb{E}\left(I(X_n=j) \mid X_0=i\right) = \sum_{n=0}^{\infty} (P^n)_{nj}$$

Communication classes

j is recurrent, i is communicating with j

$$\sum_{n=0}^{\infty} P_{jj}^{r+n+s} \geq \sum_{n=0} P_{ji}^{s} P_{ii}^{n} P_{ij}^{r} = P_{ji}^{s}\left(\sum_{n=0}^{\infty} P_{ii}^{n}\right) P_{ij}^{r}$$

$$P_{ij}^{r}, P_{ji}^{s} > 0 \qquad \sum_{n=0}^{\infty} P_{jj}^{n} > \sum_{n=0}^{\infty} P_{jj}^{r+n+s} = \infty$$

Time reversibility

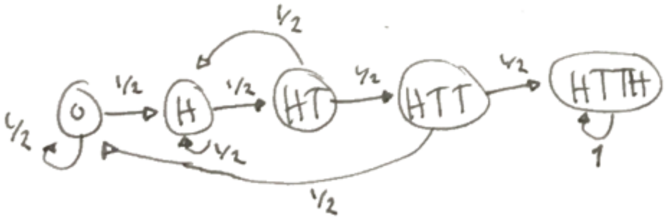
($=Px$??) $(xP)_j = \sum_i x_i P_{ij} = \sum_i x_j P_{ji} = x_j$ => gives $x = Px$ is true.

absorbing state => stuck at one place.

absorbing chains:

$$P^{n+1} = P^n P = \begin{bmatrix} Q^n & (I+Q+\dots+Q^{n-1})R \\ 0 & I \end{bmatrix} \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix} = \begin{bmatrix} Q^{n+1} & (I+Q+\dots+Q^n)R \\ 0 & I \end{bmatrix}$$

$$(I-A)I + A + A^2 + \dots + A^n = \qquad (I + A^{n+1})$$



$$\left.\begin{array}{cccc|c} 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right.$$

övning 16/11

(3.2) A stochastic matrix is called doubly stochastic if the columns sum to 1. as will if a MC has doubly stochastic transition matrix.

Then $\pi = \left(\frac{1}{k}, \frac{1}{k}, \frac{1}{k}, \cdots \frac{1}{k}\right)$

↑
stationary
dice

Solution: We want to find the vector π. $\pi P = \pi$

$\left(\frac{1}{k}, \frac{1}{k}, \frac{1}{k}\right) P = \left(\sum_1 \frac{1}{k} P_{i1}, \ldots \right) = \left(\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k}\right.$

Since P is doubly stochastic: $\sum_1 P_{ij} = 1$

(3.5) A MC has a transition matrix

$$P = \begin{bmatrix} 0 & 1/4 & 0 & 0 & 3/4 \\ 3/4 & 0 & 0 & 0 & 1/4 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 \end{bmatrix}$$

a) Describe the set of stationary distribution for the MC

Solution: Pick a vector $\pi = (a, b, c, d, e)$, $\pi P = \left(\frac{3b}{4} + \frac{e}{4}, \frac{a}{4} + \frac{3e}{4}, c, d, \frac{3a}{4} + \frac{b}{4}\right)$

For stationary: $\pi P \pi = (a, b, c, d, e)$

$$\Rightarrow \left.\begin{array}{l} \frac{3b}{4} + \frac{e}{4} = a \\ \frac{a}{4} + \frac{3e}{4} = b \\ \frac{3a}{4} + \frac{b}{4} = e \end{array}\right\} \text{use linea Algebra} \Rightarrow \begin{array}{l} a = b \\ e = b \end{array}$$

So in total: $\pi = (a, a, c, d, a)$

where they should som to 1

Man kan välja vilka tal som helst så länge summan blir 1

ex: $a = \frac{1}{4}$, $c = \frac{1}{8}$, $d = \frac{1}{8}$ osv.

b) Find longterm behaviour of MC

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 1/3 & 1/3 & 0 & 0 & 1/3 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \end{bmatrix}$$

If we start in 1, 2, 5 => Equally likely that we will end up in these states again

If we start in 3, 4 then we end up in these states again.

3.9 P a stochastic matrix.

If P is regular, is P^2 also regular?

Solution: P is regular if for some $n \geq 1$ $P > 0$

we know $P^n > 0$ for some n

{All elements are > 0 }

Since all elements in $P^n > 0$ => $P^n \cdot P^n > 0$

=> $(P^2)^n = P^{2n} > 0$ => P^2 is also regular

b) If P is transition matrix of an irreducible MC:

Is P^2 also the same?

The MC is irreducible if we from any state can travel to any other state.

Investigate $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then $P^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

if we start in 1 we are stuck there, same if we start in 2.

=> P^2 is not irreducible since we cant travel between state 1 and 2.

(3.14) We model episode (days with high ______) days and non-episode days with a MC ?

$$P = \begin{array}{c|cc} & \text{None} & e \\ \hline \text{Non episode} & 0.77 & 0.23 \\ \text{episode} & 0.24 & 0.76 \end{array}$$

a) What is the longterm probability that a given day is an episode day?

Solution: Our chain is ergodic since transition matrix is regular so we find the stationary distribution to get the limiting distribution

So we want $\pi P = \pi$, we want $\tilde{\pi}$

$$\pi = (x,y) \Rightarrow \begin{cases} 0.77x + 0.23y = x \\ 0.24x + 0.76y = y \end{cases} \Rightarrow (x,y) = (\underbrace{0.51063}, \underbrace{0489}$$

no matter which day we will have these pro

The integer probability that we will end up in an episode is 0.4894

b) Find the amount of episode days in a year.

we get $0.4894 \cdot 365$

$$I_i = \begin{cases} 1 & \text{Episode} \\ 0 & \text{non Episode} \end{cases} \qquad \mathbb{E}(\sum I_i) = 0.4894 \cdot 365$$

c) What is the average amount of days that will transpire between days we get, since the prob. of episode day is 0.4894.

From Geometric (or from other ways) we get that the expected amount of days between episode days is given by $\frac{1}{0.4894} \approx 2.04$

(3.18) Use first-step Analysis to return to state b)
For the MC With

$$P = \begin{array}{c|ccc} & a & b & c \\ \hline a & 1/2 & 1/2 & 0 \\ b & 1/4 & 0 & 1/4 \\ c & 1/2 & 1/2 & 0 \end{array}$$

Introduce: T_b = The time until we get $t = b$.

We want $\mathbb{E}(T_b \mid X_0 = b)$, Introduce $f_x = \mathbb{E}[T_b \mid X_1 = x]$

$$f_b = \mathbb{E}[T_b \mid X_0 = b] \overset{\text{use law of total expectation}}{=} \mathbb{E}[T_b \mid X_1 = a, X_0 = b]\,P(X_1 = a \mid X_0 = b) +$$

$$\mathbb{E}[T_b \mid X_1 = b, X_0 = b]\,P(X_1 = b \mid X_0 = b) + \underbrace{\text{the lost term}}_{\text{Kan vara: } \mathbb{E}(T_b \mid X_1 = c, X_0 = b) P(X_1 = b, X_0 = b)}$$

$$\mathbb{E}[T_b \mid X_1 = a, X_0 = b] = \underset{\uparrow}{1} + \mathbb{E}[T_b \mid X_0 = a]$$

We know we have taken one step from b to a.

Repeat the process for f_c, f_b, f_a

$$\left.\begin{aligned} f_b &= 1/4\,(1 + f_a) + 3/4\,(1 + f_c) \\ f_a &= 1/2\,(1 + f_a) + 1/2 \\ f_c &= 1/2\,(1 + f_a) + 1/2 \end{aligned}\right\} \Rightarrow \begin{aligned} f_a &= f_c = 2 \\ f_b &= 3 \end{aligned}$$

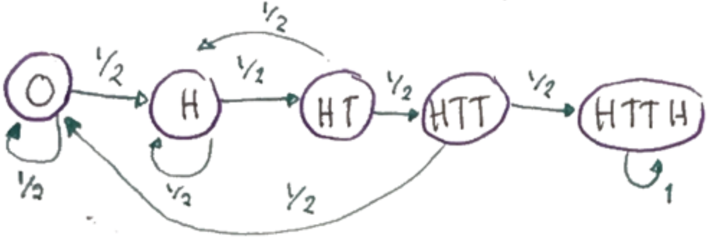
This gives $\mathbb{E}(T_b \mid X_0 = b) = 3$

Föreläsning 19/11

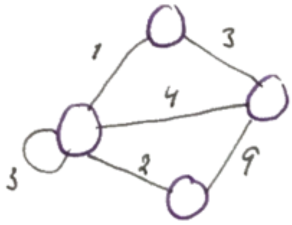
Expected number of visits in state j in chain that starts in state i.

$$E\left[\sum_{n=0}^{\infty} I(X_n=j \mid X_0=i)\right] = \sum_{n=0}^{\infty} E(I(X_n=j \mid X_0=i)) = \sum_{n=0}^{\infty} P_{ij}^{\,n} =$$

$$= \sum_{n=0}^{\infty} Q_{ij}^{n} = \left(\sum_{n=0}^{\infty} Q^{n}\right)_{ij} = F_{ij} \qquad F=(I\cdot Q)^{-1}$$



Time reversibility



Define a weighted random walk

$w(i,j) = w(j,i)$ is weighted between i and j

$$P_{ij} = \frac{w(ij)}{\sum_j w(ij)}.$$

The stationary distribution v:

$$v_n = \frac{\sum_j w(i,j)}{\sum_i \sum_j w(i,j)}$$

$$v_i P_{ij} = \frac{w(i,j)}{\sum_i \sum_j w(i,j)} = v_j P_{ji}$$

Assume a chain is time reversible. Define a graph with an edge between i and j <u>if</u> $\pi_i P_{ij} = \pi_j P_{ji}$ is non zero

weight: $w(i,j) = \pi_i P_{ij}$

The probability to move from i to j:

$$\frac{w(i,j)}{\sum_j w(i,j)} = \frac{\pi_i P_{ij}}{\sum_j \pi_i P_{ij}} = \frac{\pi_i P_{ij}}{\pi_i} = P_{ij}$$

The multinomial Dirchlet conjugacy

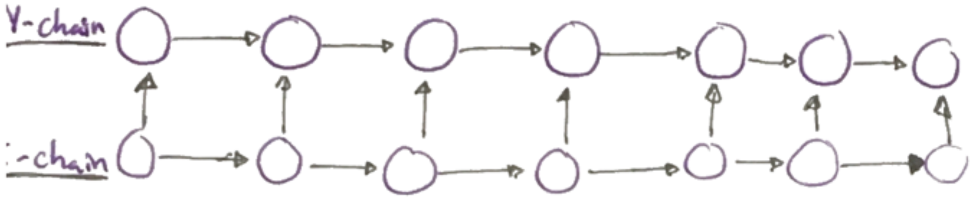
$$\pi(\theta|x) \propto_\theta \tilde{\pi}(x|\theta)\tilde{\pi}(\theta)$$

$$\propto_\theta \theta_1^{x_1}\theta_2^{x_2}\cdots\theta_k^{x_k}\,\theta_1^{\alpha_1-1}\cdots\theta_k^{\alpha_k-1} = \underline{\theta_1^{\boxed{\alpha_1+x_1}-1}\,\theta_2^{\boxed{\alpha_2+x_2}-1}\cdots\theta_k^{\boxed{\alpha_k+x_k}-1}}$$

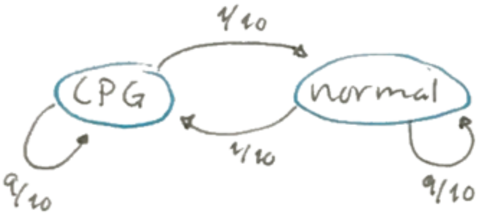
$$\theta|x \sim \text{Dirichlet}(\alpha+x) = \text{Dirichlet}(\alpha_1+x_1, \alpha_2+x_2, \dots, \alpha_k+x_k)$$

$$\pi(\theta|x) = \text{Dirichlet}(\theta; \alpha+x)$$

$$\Rightarrow \tilde{\pi}(x) = \frac{\tilde{\pi}(x|\theta)\tilde{\pi}(\theta)}{\pi(\theta|x)}$$



Hidden markov model
(HMM)



- How to find the transition matrix for y-chain
- Finding the parameters of the hidden x-chain.
- Given data $y_0, \dots, y_T$ for $Y_0 \dots Y_T$

Forward - Backward → - What is the posterior $\tilde{\pi}(x_n|y_0 \dots y_T$

- What is the most likely sequence $x_0, \dots x_T$ maximizing $\tilde{\pi}(x_0 \dots x_T | y_0 \dots y_T)$

Viterbi (← skip)

Learning about the transition matrix from data

Want to estimate (learn about) transition matrix P given data $x_0\ x_1, \ldots x_T$

Likelihood

$$\pi(x_0 .. x_k | P) = \pi(x_0) \prod_{i=1}^{k} \pi(x_i | x_{i-1}, P) = \pi(x_0) \prod_{i=1}^{k} P_{x_1 \ldots x_i}$$

Where c_{ij} is the count of transitions from i to j =

$$= \pi(x_0) \prod_{i=1}^{S} \prod_{j=1}^{S} (p_{ij})^{c_{ij}} \qquad c_{ij} = \sum_{R=1}^{T} I(x_{k-1} = i, x_i = j)$$

Prior $\Theta \sim \text{Dirichlet}(1, 1, 1)$ For first row

Posterior: $\Theta | x \sim \text{Dirichlet}(1+3, 1+4, 1+1)$

Expected value: $\left(\frac{1+3}{1+3+1+4+1+1}, \frac{1+4}{1+3+1+4+1+1}, \frac{1+1}{1+3+1+4+1+1} \right)$

$$\pi(x_{i+1} | y_0 \ldots y_{i+1}) \propto_{x_{i+1}} \pi(y_{i+1} | x_{i+1}, y_0 .. y_i) \pi(x_{i+1} | y_0 .. y_i)$$

$$= \pi(y_{i+1} | x_{i+1}) \int \pi(x_{i+1}, x_i | y_0 .. y_i) \, dx_i$$

$$= \pi(y_{i+1} | x_{i+1}) \int \pi(x_{i+1} | x_i, y_0, .. y_i) \pi(x_i | y_0 .. y_i) \, dx_i$$

$$= \pi(y_{i+1} | x_{i+1}) \int \pi(x_{i+1} | x_i) \pi(x_i | y_0 \ldots y_i) \, dx_i$$

övning 20/11

3.22 Consider the general two step chain

$$P = \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 1-p & p \\ 2 & q & 1-q \end{array}$$ where both p, q are not equal to 0

T: The first return time to 1 if we start in 1

a) Show $P(T \geq n) = p(1-q)^{n-2}$ for $n \geq 2$

Solution: So we get $P(T \geq n) = p(1-q)^{n-2}$ is true because you can't return to state one anymore after you have passed it

b) Find $E(T)$ and verify that $E(T) = 1/\pi$ where π is the stationary distribution

Solution: Since we calculated $P(T \geq n)$ in a)

$$\Rightarrow E(T) = \sum_{n=1}^{\infty} P(T \geq n)$$

since we can't return to T in less than 1 step

So we get: $E(T) = P(T \geq 1) + \sum_{n=2}^{\infty} P(T \geq n) = 1 + \sum_{n=2}^{\infty} P(T \geq n) =$

$$= 1 + \sum_{n=2}^{\infty} p(1-q)^{n-2} \underset{\uparrow}{=} 1 + p \frac{1}{1-(1-q)} = 1 + \frac{p}{q} = \frac{p+q}{q} = 1/\pi$$

By geometric sum

where the last formula comes from the book

Argument for $E(T) = \sum_{n=1}^{\infty} P(T \geq n)$

$$E(T) = \sum_{k=0}^{\infty} k \cdot P(T=k) = \sum_{k=0}^{\infty} \sum_{t=1}^{k} P(T=k)$$

switch directions of summation

$$\Rightarrow \sum_{t=1}^{\infty} \underbrace{\sum_{k=t}^{\infty} P(T=k)}_{= P(T \geq t)}$$

(3.38) Throw 5 dices, set aside those which are 6.
Set aside those which were 6, and repeat until
all dices are 6.

a) X_n: The number of dices that are 6 after n throws
Find the transition matrix for this MC.

For P_{ij} we want the probability that if we have i 6 in round n that we get j 6 in the next round

This is given by: $\binom{5-i}{j-i} \left(\frac{1}{6}\right)^{j-i} \left(\frac{5}{6}\right)^{5-j} \quad 0 \leq i, j \leq 5$

5-i since if we have i 6 in round n, then we only throw 5-i dices

This gives: $P = \begin{array}{c|cccccc} & 0 & 1 & 2 & 3 & 4 & 5 \\ \hline 0 & 0.402 & 0.402 & 0.161 & \cdots & & \\ 1 & 0 & 0.482 & \ddots & & & \\ 2 & 0 & 0 & \ddots & \ddots & & \\ 3 & 0 & 0 & 0 & \ddots & & \\ 4 & 0 & 0 & 0 & 0 & & \\ 5 & 0 & 0 & 0 & 0 & 0 & \end{array}$

b) How many turns does it take in average before you get all 6:es?

We want expected amount of steps before being absorbed
This is related to fundamental matrix $(I-Q)^{-1}$, where Q is the P matrix except the last row and column.

The expected amount of steps is then given by the sum of the elements in the first row of the $(I-Q)^{-1}$ matrix
This gives the expected amount of steps as $\underline{\underline{13.02}}$

(L, 2.4:1) Assume that an experiment can have one of three outcomes 1, 2, 3 with corresponding probabilities p_1, p_2, p_3

13 independent experiments are performed where we observe:

3 outcomes of 1

9 outcomes of 2

1 outcome of 3

The general formula for Multinormal distribution is

$$\frac{n!}{n_1!\, n_2! \,..\, n_k!} \, p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$ for k possible outcomes

a) Use a Dirichlet (α) for $\mathbb{p} = (p_1, p_2, p_3)$ where $\alpha = (1, 1, 1)$

Find posterior $\pi(\mathbb{p} / y)$

Solution: From the experiment we have $\pi(y/\mathbb{p}) \propto p_1^3 p_2^2 p_3^1$

$$\pi(\mathbb{p}) = \frac{\Gamma(1+1+1)}{\Gamma(1)\Gamma(1)\Gamma(1)} \, p_1^{1-1} p_2^{1-1} p_3^{1-1}$$

This gives $\pi(\mathbb{p} | y) \propto \pi(y | \mathbb{p}) \, \pi(\mathbb{p}) \propto p_1^3 p_2^2 p_3^1$

The posterior has the form of Dirichlet distribution

$\Rightarrow \pi(\mathbb{p} | y) = \text{Dirichlet}(4, 10, 2)$

b) We want $E(\mathbb{p}) = (E(p_1), E(p_2), E(p_3))$ (Use formula in appendix)

In appendix we have for Dirichlet(α)

$$\Rightarrow E(\mathbb{p}) = \frac{\alpha}{\Sigma \alpha_i}$$

In our case we get: $E(\mathbb{p}) = \frac{(4, 10, 2)}{16} \left(\frac{4}{16}, \frac{10}{16}, \frac{2}{16}\right) = \left(\frac{2}{8}, \frac{5}{8}, \frac{1}{8}\right)$

c) We want the probability of observing

1 outcome of 1

2 outcomes of 2

1 outcome of 3

with the same prior as before

→

we want posterior predictrive distribution

$\pi(\tilde{y}|y)$ $\begin{cases} \tilde{y} : \text{New observation} \\ y : \text{Old observation} \end{cases}$

$$\pi(\tilde{y}|y) = \int \pi(\tilde{y}, \mathbb{p}|y)\, d\mathbb{p} = \int \underbrace{\pi(\tilde{y}|\mathbb{p})}_{\text{likelihood}} \underbrace{\pi(\mathbb{p}|y)}_{\text{Posterior}}\, d\mathbb{p} =$$

sum

$$= \int \frac{4!}{1!\,2!\,1!} \cdot p_1^{1} \cdot p_2^{2} \cdot p_3^{1} \cdot \frac{\Gamma(16)}{\Gamma(4)\Gamma(10)\Gamma(2)} \cdot p_1^{3}\, p_2^{9}\, p_3^{1}\, d\mathbb{p} =$$

Values from b)

$$= \frac{4!}{2!} \cdot \frac{\Gamma(16)}{\Gamma(4)\Gamma(10)\Gamma(2)} \int p_1^{4} p_2^{11} p_3^{2}\, d\mathbb{p}$$

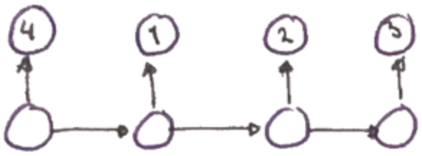
The expression in the integral corresponds to Dirichlet $(5, 12, 3)$

$$\Rightarrow \int p_1^{4} p_2^{11} p_3^{2}\, d\mathbb{p} = \frac{\Gamma(5)\Gamma(12)\Gamma(3)}{\Gamma(20)}$$

$$\Rightarrow \pi(\hat{y}|y) = \frac{4!}{2!} \cdot \frac{\Gamma(16)}{\Gamma(4)\Gamma(10)\Gamma(2)} \cdot \frac{\Gamma(5)\Gamma(12)\Gamma(3)}{\Gamma(20)} \approx 0.113$$

Föreläsning 21/11

The foward algorithm



$x_i = \begin{cases} 0 & \text{no disease} \\ 1 & \text{disease} \end{cases}$

$Y_i | X_i \sim \text{Poisson}(\lambda x_i) \begin{cases} \lambda_0 = 1.3 \\ \lambda_1 = 3.7 \end{cases}$

$\tilde{\pi}(X_0 | Y_0)$ compute this for $X_0 = 1$ and $X_0 = 0$

$i > 0$ compute, for $x_{i+1} = 0$ or 1

$$\tilde{\pi}(x_{i+1} | Y_0 \dots Y_{i+1}) \propto \tilde{\pi}(Y_{i+1} | X_{i+1}) \cdot \int \pi(x_{i+1} | X_i) \tilde{\pi}(X_i | Y_0 \dots Y_i) \, dX_i$$

$$= \tilde{\pi}(Y_{i+1} | X_{i+1}) \sum_{j=0}^{1} \tilde{\pi}(X_{i+1} | X_i = j) \tilde{\pi}(X_i = j | Y_0 \dots Y_i)$$

Backward Algorithm

$$\tilde{\pi}(Y_{i+1} \dots Y_T | X_i) = \sum_{j=0}^{1} \tilde{\pi}(Y_{i+1} | X_{i+1} = j) \tilde{\pi}(Y_{i+1} \dots Y_T | X_{i+1} = j) \tilde{\pi}(X_{i+1} = j | X_i)$$

Branching processes

at each node offsprings are created independently, according to the offsprings distribution, with probability vector a

a_i is the pobability of i offspring

$\boxed{a_0 > 0}$ $\boxed{a_0 + a_1 < 1}$ Let Z_n be the size of the "population" at "Generation" n

$\boxed{Z_0 = 1}$ Z_0, Z_1, Z_2 is a markov chain

- state zero is absorbing
- all other states are transient

A branching process either goes extinct or grows without bounds.

Mean generation size:

$Z_n = \sum_{i=1}^{Z_{n-1}} X_i$ where X_i are independent and follow the offspring distribution. $\boxed{\mu = E(X_i)}$

$$E(Z_n) = E(E(Z_n | Z_{n-1})) = E(E(\sum_{i=1}^{Z_{n-1}} X_i | Z_{n-1})) = E(\sum_{i=1}^{Z_{n-1}} E(X_i) | Z_{n-1}) =$$

$$= E(Z_{n-1}\mu | Z_{n-1}) = \mu E(Z_{n-1})$$

$$E(Z_0) = E(1) = 1$$

$$E(Z_n) = \mu^n$$

$$\lim_{n\to\infty} E(Z_n) = \lim_{n\to\infty} \mu^n = \begin{cases} 0 & \mu < 1 \quad \text{Subcritical} \\ 1 & \mu = 0 \quad \text{critical} \\ \infty & \mu > 1 \quad \text{supercritical case} \end{cases}$$

$$Var(Z_n) = Var(E(Z_n | Z_{n-1})) + E(Var(Z_n | Z_{n-1})) = Var(\mu Z_{n-1}) +$$

$$+ E(Var(\sum_{i=1}^{Z_{n-1}} X_i | Z_{n-1})) = \mu^2 Var(Z_{n-1}) + E(\sum_{i=1}^{Z_{n-1}} Var(X_i) | Z_{n-1}) =$$

$$= \mu^2 Var(Z_{n-1}) + E(Z_{n-1}\sigma^2 | Z_{n-1}) = \mu^2 Var(Z_{n-1}) + \sigma^2 \mu^{n-1}$$

$$\boxed{\sigma^2 = Var(X_i)}$$

$$\boxed{Var(Z_n) = \mu^2 Var(Z_{n-1}) + \sigma^2 \mu^{n-1}} \qquad \boxed{Var(Z_0) = 0}$$

$$Var(Z_n) = \sigma^2 \mu^{n-1} \sum_{k=0}^{n-1} \mu^k = \begin{cases} \text{if } \mu = 1 : Var(Z_n) = n\sigma^2 \\ \text{if } \mu \neq 1 \quad \sigma^2 \mu^{n-1} \dfrac{\mu^{n-1} - 1}{\mu - 1} \end{cases}$$

critical

subcritical $\xrightarrow[n\to\infty]{} 0$

supercritical $\xrightarrow[n\to\infty]{} \infty$

Probability Generating function

Related concepts:
- moment generating function
- characteristic functions
- laplace transforms

Assume X is a random variable on $\{0,1,2,3,\dots\}$

Define $G_X(s) = E(s^X)$ for $|s| \leq 1$

$E(S^X) = \sum_{k=0}^{\infty} s^x P(X=k)$

Example X has distribution given by

x	0	1	2	3
	1/2	1/4	1/8	1/8

$$G_X(s) = \frac{1}{2}s^0 + \frac{1}{4}s^1 + \frac{1}{8}s^2 + \frac{1}{8}s^3$$

Properties: if X and Y are independent, then

- $G_{X+Y}(s) = E(s^{X+Y}) = E(s^X s^Y) = E(s^X)E(s^Y) = G_X(s)G_Y(s)$
- if $G_X(s)$ and $G_Y(s)$ are equal then X and Y have the same distribution.
- $P(X=j) = \dfrac{G^{(j)}(0)}{j!}$ for all j :

$$G_X(s) = \sum_{k=0}^{\infty} s^k p(X=k)$$

$$G_X'(s) = \sum_{k=1}^{\infty} k s^{k-1} p(X=k)$$

$$G_X''(s) = \sum_{k=2}^{\infty} k^2 s^{k-2} p(x=k)$$

Moments: $G(s) = E(S^X)$

$G'(s) = E(X s^{X-1})$, $\boxed{G'(1) = E(X)}$

$G''(s) = E(X(X-1)s^{X-2})$ $G''(1) = E(X(X-1)) = E(X^2-X) =$

$= E(X^2) - E(X)$

$\boxed{Var(X)} = E(X^2) - E(X)^2$ $var(X) : G''(1) + G'(1) - G'(1)^2$

Apply probability generating functions to branching processes

$$\underline{G_{Z_n}(s)} = E(s^{Z_n}) = E(E(s^{Z_n}|Z_{n-1})) = E(E(s^{\sum_{i=1}^{Z_{n-1}} X_i}|Z_{n-1})) =$$

$$= E(E(\prod_{i=1}^{Z_{n-1}} s^{X_i}|Z_{n-1})) = E(\prod_{i=1}^{Z_{n-1}} E(s^{X_i})|Z_{n-1}) = E(G_X(s^{Z_{n-1}}|Z_{n-1})) =$$

$$= \underline{G_{Z_{n-1}}(G_X(s))} \qquad G_{Z_n}(s) = G_{Z_0}(\underbrace{G_X(G_X(\dots G_X(s)\dots)}_{n \text{ times}})$$

$$G_{Z_0} = E(s^{Z_n}) = S$$

Theorem:

The probability of extinction, ℓ, for a branching process is the smallest positive root of $G_X(s) = s$, where G_X is the probability generating function for the offspring process.

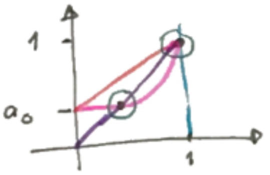
- In the critical cas $\ell = 1$
- In the supercritical case $\ell < 1$

Outline of proof:

- ℓ is a root of $s = G_X(s)$

Def: $\ell_n = P(Z_n = 0)$ Then $\ell_n = G_{Z_n}(0) = G_X(G_{Z_{n-1}}(0))$

$$\ell = \lim_{n\to\infty} = \lim_{n\to\infty} G_X(\ell_{n-1}) = G_X(\lim_{n\to\infty} \ell_{n-1}) = G_X(\ell)$$



$G_X(0) = a_0$ $\qquad G_X''(s) = E(X(X-1)s^{X-2}) > 0$

$G_X(1) = 1$ $\qquad G_X'(1) = E(X) = \mu$

smallest rootc is one.
→ prob = 1

μ bigger than 1
smallest root < 1

övning 23/11

⑤ The k:th factorial moment of a random variable X is given by.

$E(X(X-1)) \dots (X-k+1)) = E\left(\frac{X!}{(X-K)!}\right)$

a) Given G, how do we find ?

Solution: $G_X(s) = \sum_{k=0}^{\infty} s^k P(X=k)$, we have $\frac{d}{ds} G_X(s) =$

$$= \sum_{k=1}^{\infty} k \cdot s^{k-1} P(X=k) \overset{S=1}{=} E(X)$$

By itteration we get $\frac{d^t}{ds^t} G_X(s) = \sum_{k=1}^{\infty} k(k-1)\dots(k-t+1) s^{k-t} P(X=k$

Insert $s=1 \Rightarrow G^{(k)}(1) = E\left(\frac{X!}{(X-h)!}\right)$

b) Find k:th factorial moment for $X \sim Bin(n,p)$

we find the generating function for X.

Solution: X binomial $\Rightarrow X = I_1 + I_2 + \dots + I_n$ where $I_n \sim Bin(p)$

So $G_{I_1}(s) = E(s^x) \underset{\uparrow}{=} s^0(1-p) + s^1 p$

def of expectation

$$G_X(s) = G_{I_1}(s)\, G_{I_2}(s) \dots G_{I_n}(s) = (1-p+sp)^n$$

Use a) to get k:th factorial moment $\Rightarrow$

$$\Rightarrow \left.\frac{d^k G(s)}{ds^k}\right|_{s=1} = \frac{k!}{(n-k)!} p^k$$

4.6 X_1, X_2 i.i.d Bern(p)

$N \sim Po(\lambda)$ $\quad N \underset{\text{independent}}{\perp\!\!\!\perp} X_i \ \forall i$

a) Find $G(s)$ for $Z = \sum_{k=1}^{N} x_i$

Solution: We have $G_Z(S) = E(S^{\sum_{k=1}^{N} x_i})$ = use law of total expectation:

$$= \sum_{n=0}^{\infty} E(S^{\sum_{k=1}^{N} x_i} \mid N=n)\, P(N=n) = \sum_{n=0}^{\infty} (1-p+sp)^n \cdot \frac{\lambda^n e^{-\lambda}}{n!} =$$

$$= e^{-\lambda} \sum_{n=0}^{\infty} \frac{((1-p+sp)\lambda)^n}{n!} = e^{-\lambda} \cdot e^{(1-p+sp)\lambda} = e^{p\lambda(s+1)}$$

b) Use a) to find the distribution of Z.

Solution: Since $e^{-p\lambda(1-s)}$ is the generating function for a $Po(\lambda p)$ random variables, this gives $Z \sim Po(\lambda p)$

4.8 X is discrete random variable with generating function $G(s)$. Show: $P(X \text{ is even}) = \frac{1-G(-1)}{2}$

Solution: $G(-1) = \sum_{i=0}^{\infty} (-1)^k P(X=k) = P(x=0) - P(x=1) + P(x=2) \ldots$

Collect possitive and negative terms =

$= P(X=0) + P(X=2) + \ldots - P(X=1) - P(X=3) - \ldots =$

$= P(x \text{ is even}) - P(x \text{ is odd})$ $\quad (P(X \text{ is even}) + P(x \text{ is odd}) = 1)$

combine the expressions

$P(x \text{ is even}) = \frac{1+G(-1)}{2}$

(4.12) A branching process has offspring distribution $a = (1/4, 1/4, 1/2)$

Find the following:

a) μ c) The extinct probability

b) $G(s)$ d) $G_z(s)$ e) $P(z_z = 0)$

Solution: a) $\mu = 0 \cdot 1/4 + 1 \cdot 1/4 + 2 \cdot 1/2 = 5/4$

b) $G(s) = E(s^x) = \left\{ \text{where } x \text{ have the distribution given by } a \right\}$

$G(s) = s^0 \cdot 1/4 + s^1 \cdot 1/4 + s^2 \cdot 1/2 = 1/4 + s/4 + \frac{s}{2}$ maybe

c) The extintion probability is given by the (smallest possible) solution $s = G(s)$

$s = 1/4 + s/4 + \frac{s^2}{2} \Rightarrow s = 1/2$

d) $G_z(s) = G(G(s))$

$G(G(s) = 1/4 + \frac{G(s)}{4} + \frac{G(s)^2}{2}$ = Insert $G(s)$ =

$= \frac{1}{4} + \frac{s}{4} + \frac{s^2}{2} = \frac{1}{32}(11 + 4s + 9s^2 + 4s^3 + 4s^4)$

e) $P(z_2 = 0)$

Use $G(s) = \sum_{k=0}^{\infty} s^k P(X=k) = G_x(0) = P(X=0)$

For us $P(z_0 = 0) = G_z(0) = 11/32$

④.20 Consider the offspring distribution. $a_k = (1/2)^{k+1}$ $k \geq 0$

a) Find the extinction probability

Solution: Calculate mean and investigate

mean = $\mu = \sum_{k=0}^{\infty} k \cdot (1/2)^{k+1} = \frac{1}{2} \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^k$ = use mean of geometric distribution $\Rightarrow \sum_{k=1}^{\infty} k \left(\frac{1}{2}\right)^k = \frac{1}{1/2} = 2$

$\Rightarrow \mu = \frac{1}{2} \cdot 2 = 1$

since $\mu = 1 \Rightarrow$ Extinction probability $= 1$

b) Show by induction that $G_n(s) = \frac{n-(n-1)s}{n+1-ns}$

Solution: $G(s) = E(s^x) = \sum_{k=0}^{\infty} s^k \cdot \left(\frac{1}{2}\right)^{k+1} = \frac{1}{2} \sum_{k=0}^{\infty} \left(s \cdot \frac{1}{2}\right)^k =$

$= \frac{1}{2} \cdot \frac{1}{1-\frac{s}{2}} = \frac{1}{2-s} = \frac{1-(1-1)s}{1+1-1 \cdot s} \Rightarrow G(s)$ is our base case

Use induction:

$G_n(s) = G_{n-1}(G(s)) = \frac{n-1-(n-2)G(s)}{n-1+1-(n-1)G(s)} =$

= insert $G(s)$ and simpliefy $= \frac{n-(n-1)s}{n+1-ns}$

Föreläsning 26/11

Extinction theorem for branching processes

The probability e of eventual extinction is the smallest positive root of $s = G_x(s)$, where G_x is the probability generating function of the offspring distribution.

$G_x(s) = E(s^x)$ $\quad$ $G_{Z_n}(s) = \underbrace{G_x(G_x(\ldots G_x(s)\ldots)}_{n \text{ iterations}}$

e is a root of $s = G_x(s)$

define $e_n = P(Z_n = 0)$

$e_n = G_{Z_n}(0) = G_x(G_{Z_{n-1}}(0)) = G_x(e_{n-1})$

$e = \lim_{n\to\infty} e_n = G_x\left(\lim_{n\to\infty} e_{n-1}\right) = G_x(e)$

e is the smallest root: Let x be a root

$e_1 = G_{Z_1}(0) = G_x(0) \leqslant G_x(x) = x$

By induction: $e_n = G_{Z_n}(0) = G_x(G_{Z_{n-1}}(0)) = G_x(e_{n-1}) \leqslant G_x(x) = x$

Taking limits: $e \leqslant x$

Example: offspring distribution

0	1	3	4
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

Expected number of offsprings:

$E(x) = 0\cdot\frac{1}{2} + 1\cdot\frac{1}{4} + 2\cdot\frac{1}{8} + 3\cdot\frac{1}{8} = \frac{7}{8} < 1$

Subcritical: probability for extinction is 1

Example:

0	1	2	3
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

$E(x) = 0\cdot\frac{1}{4} + 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 3\cdot\frac{1}{4} = \frac{6}{4} > 1$

$G_x(s) = E(s^x) = s^0\cdot\frac{1}{4} + s^1\cdot\frac{1}{4} + s^2\cdot\frac{1}{4} + s^3\cdot\frac{1}{4}$

find the smallest positive s such that:

$\Rightarrow s = \frac{1}{4} + \frac{1}{4}s + \frac{1}{4}s^2 + \frac{1}{4}s^3$

optimize numerically: $s = 0.4142$

Law of large numbers

$Y_1, \ldots, Y_m$ a random sample, with same distribution Y finite expectation μ.

$$\lim_{m \to \infty} \frac{Y_1 + \ldots + Y_m}{m} = \mu \quad \text{with probability } 1.$$

In the markov chain limit it dosen't matter if you start on X_0 or X_1 aso because when going to m (which is a large number) it matters less and less.

Example: A binary sequence (010110001...) of length m is called "good" if it has no adjacent 1's

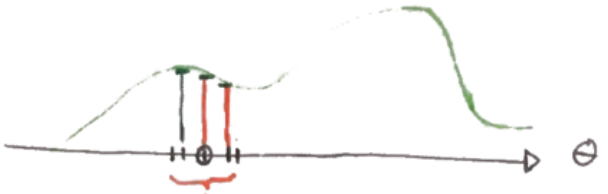
(01010001001) ~~($0\underline{11}001010$)~~

What is the average number of 1's in good sequence of length m?

Solution: Define a uniform distribution on the set of good sequences of length m, estimate the expected number of 1's under this distribution

1) Construct a markov chain on set of good sequences of length n with the uniform distribution as it's stationary distribution
2) Use the chain to simulate $X_1, \ldots, X_m$
3) Approximate the expectation over all good sequences by the average number of 1's in $X_1, \ldots, X_m$

The metropolis Hastings algorithm



* probabilitys to take steps.

Example 5.2

$\pi(\theta)$ is a discrete distribution on 1, 2, 3 ...

$$\pi(\theta) = \frac{\theta^{-3/2}}{\sum_{\theta=1}^{\infty} \theta^{-3/2}}$$

Use metropolis hastings to create an approximate sample

use as proposal function a random walk on:

if you are at i and propose j. What is acceptance probability?

$i, j > 1$: probability is : $\frac{\pi(j)}{\pi(i)}$

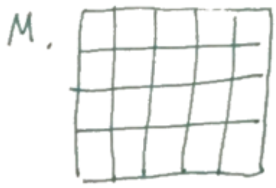
if $i = 1$, $j = 2$

$$\frac{\theta(i|j)}{\theta(j|i)} = \frac{1/2}{1} = 2$$

A coding function is a permutation of all 26 letters (27 characters including space).

A prior distribution on the set of permutations uniform.

- Likelihood: construct a markov chain modelling "resonable" language



$$\text{score}(j) = \prod_{i=1}^{319} M_{j(c_i)\, j(c_{i+1})}$$

$c_1 \dots c_{320}$ is coded message

posterior dist. on the set of permutations is poportional to score(j)

övning 27/11

(4.27) Consider a Branching process whose offspring distribution is a Bernoulli distribution with parameter p.

a) Find the PGF for the n:th generation Z_n
probability generating function

Solution: Use $G_n(s) = G_{n-1}(G(s))$

$$G(s) = E(s^X) = s^0 \cdot (1-p) + s^1 \cdot p = 1-p+sp$$

$$G(s) = G(G(s)) = 1-p+G(s)p = 1-p+(1-p+sp)p = 1-\cancel{p}+\cancel{p}-p^2+sp^2$$
$$= 1-p^2+sp^2$$

Hypothesis: $G_n(s) = 1-p^n+sp^n$

Use induction: $G_n(s) = G_{n-1}(G(s)) = 1-p^{n-1}+G(s)p^{n-1} = 1-p^n+sp^n$ ☐

So PGF for Z_n is $G_n(s) = 1-p^n+sp^n$

This is a PGF for Bernoulli distribution with parameter p^n

b) For $p=0.9$, find extinction probability and the expectation of the total progeny

Solution:

Total progeny: $T = \lim_{n\to\infty} T_n$, where $T_n = Z_0+Z_1+\ldots+Z_n$
(amount of generations?)

Find m of offspring distribution, $\mu = 0 \cdot (1-0.9) + 1 \cdot 0.9 = 0.9$

Since $\mu < 1 \Rightarrow$ Extinction probability is 1.

$$E(T) = E(\lim T_n) = E\left(\lim_{n\to\infty} \sum_{k=0}^{n} Z_k\right) = \lim_{n\to\infty} \sum_{k=0}^{n} E(Z_k) = \lim_{n\to\infty} \sum_{k=0}^{n} \mu^k =$$
$$= \frac{1}{1-\mu} = \frac{1}{1-0.9} = 10$$

If $\sum E(|Z_k|) < \infty$ then you can interchange lim and the expectation.

(4.28) In a Branching process with immigration a random number of immigrants W_n is independently added to the population at the n:th generation.

a) Let H_n be the PGF for W_n, G_n is PGF for Z_n

Show $G_n(s) = G_{n-1}(G(s))\,H_n(s)$

Solution: we get $Z_n = \sum_{i=1}^{Z_{n-1}} x_i + W_n$

$$G_n(s) = E(s^x) = E\left(s^{\sum_{i=1}^{Z_{n-1}} x_i + W_n}\right) = E\left(s^{\sum_{i=1}^{Z_{n-1}} x_i} \cdot s^{W_n}\right) =$$

{ If X, Y are independent then g(x), h(y) are also independent (some restrictions on g and h, however this dosen't effect us) }

$$= \underbrace{E\left(s^{\sum_{i=1}^{Z_{n-1}} x_i}\right)}_{} \underbrace{E\left(s^{W_n}\right)}_{} = \underbrace{G_{n-1}(G(s))}_{} \cdot \underbrace{H_n(s)}_{}$$

(5.1) 4 out of 5 trucks are followed by a car

1 out of 4 cars are followed by a truck

Cars pay 1.5 \$, trucks pay 5 \$

If 1000 vehicles pass, how much money is collected ?

Solution: We get:

$$P = \begin{array}{c|cc} & \text{truck} & \text{car} \\ \hline \text{truck} & 1/5 & 4/5 \\ \text{car} & 1/4 & 3/4 \end{array}$$

Introduce r:

$r(\text{car}) = 1.5$ \$

$r(\text{truck}) = 5$ \$

We have L(N)

$$\lim_{n\to\infty} \frac{r(x_1) + r(x_2) + \ldots + r(x_n)}{n} = E(r(x))$$

where X have the stationary distribution π.

we need π : $\pi P = \pi$ $\Rightarrow$

$$\frac{1}{5}\pi_1 + \frac{1}{4}\pi_2 = \pi_1$$
$$\frac{4}{5}\pi_1 + \frac{3}{4}\pi_2 = \pi_2$$

$$\Rightarrow \pi = \left(\overset{\text{truck}}{\frac{5}{21}}, \overset{\text{car}}{\frac{16}{21}}\right) \qquad E(r(x)) = 5 \cdot \frac{5}{21} + 1{,}5\,\frac{16}{21}$$

To get the amount for 1000 vehicles:

$$1000 \cdot \left(\frac{5}{21} \cdot 5 + 1.5 \cdot \frac{16}{21}\right) \approx 2333 \text{ \$}$$

(5.6) Show how to generate a Poisson RV with parameter λ using Metropolis-hastings. Use symmetric walk as proposal distribution

Solution: In metropolis - Hastings we have:

$\boxed{\pi :}$ Our $Po(\lambda)$ distribution

$\boxed{T :}$ Transition matrix for the symmetric walk

$\boxed{0 :}$ we are in state i

$\boxed{1 :}$ Choose a new state j with the help of T

$\boxed{2 :}$ we decide whether to move to j or not

with $a(i,j) = \frac{\pi_j T_{ji}}{\pi_i T_{ij}}$

$a(i,j) \geqslant 1$ accept

$a(i,j) \leqslant 1$ accept with probability $a(i,j)$

If $i = 0, j = 1$: $a(i,j) = \dfrac{\frac{\lambda^1 e^{-\lambda}}{1!} \cdot \frac{1}{2}}{\frac{\lambda^0 e^{-\lambda}}{0!} \cdot 1} = \frac{\lambda}{2}$

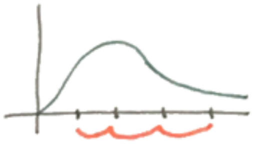
Flip a coin with probability $\frac{\lambda}{2}$, heads => go to 1

Tails => stay at 0

In general: $i = k$, $j = k+1$

$$a(i,j) = \frac{\frac{\lambda^{k+1} e^{-\lambda}}{(k+1)!} \cdot \frac{1}{2}}{\frac{\lambda^k e^{-\lambda}}{k!} \cdot \frac{1}{2}} = \frac{\lambda}{k+1}$$

$i = k$, $j = k-1$ $\qquad a(i,j) = \dfrac{\frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} \cdot \frac{1}{2}}{\frac{\lambda^k e^{-\lambda}}{(k)!} \cdot \frac{1}{2}} = \frac{k}{\lambda}$



Föreläsning 28/11

Darwin's finches (5.4)

Question: Among all incidence matrices with given margins, what is the distribution of the number of "checkerboards"? Is the count 333 "Extreme"

⊗ generate an (approximate) sample mean from the set of incid matrices with given margins, count "checkerbards" compare.

⊗ Not so easy to generate sample from the uniform distribution

⊗ Instead use metropolis hastings:
- target distribution: uniform
- proposal distribution: random walk on such matric

Algorithm:

* start with some incidence matrix θ
* count how many incidence matrices you can generate from θ with swaps: n
* choose one of these at random θ^{new}
* count the number of inc. matrices you can generate from θ^{new} m
* acceptance probability $\min\left(1, \frac{\pi(\theta^{new})Q(\theta|\theta^{new})}{\pi(\theta)Q(\theta^{new}/\theta)}\right) =$

$= \min\left(1, \frac{1/m}{1/n}\right) = \min\left(1, \frac{n}{m}\right.$

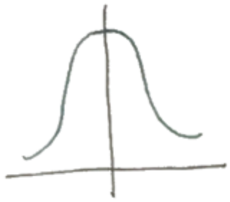
3.3 Generating an approx. sample from posterior.

Two problems:

1) Convergence (mixing)
2) $\boxed{\text{autocorrelation}}$

Random walk metropolis hastings

Symmetric proposal function: $Q(\theta^{new} | \theta) = Q(\theta | \theta^{new})$ for all θ, θ^{new}



Simulate from standard normal normal(0, 1)

proposal function $\theta^{new} | \theta \sim \text{normal}(\theta, \sigma^2)$

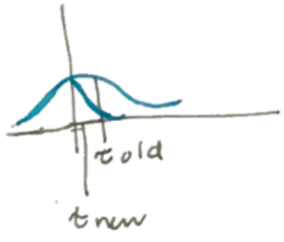
$\sigma = 0.05$?

Independent proposal:

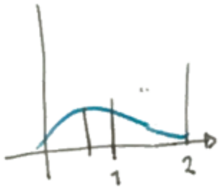
$\theta^{new} | \theta \sim \text{normal}(0, \sigma^2)$

larger values on sigma gives better results.

Proposal for τ:



Proposal for a:



Want to get an approximate sample mean from the posterior $\pi(\theta | \text{Data}) \propto_\theta \pi(\text{Data} | \theta)\, \pi(\theta)$

Acceptance probability: (symmetric prop), $\min\left(1, \frac{\pi(\theta^{new} | \text{Data})}{\pi(\theta | \text{Data})}\right)$

Gibbs sampling

Idea: · Propose change to only one component of θ at a time. keeping all other components fixed.

- The proposed value for the component uses its conditional distribution given all other components.
- The acceptance probability is always 1.

(creates a path!)

Assume $\theta = (\theta_1, \theta_2)$ acceptance probability:

move θ, keeping θ_2 fixed

$$a = \min\left(1, \frac{\pi(\theta_1^{new}, \theta_2)\,\pi(\theta_1 | \theta_2)}{\pi(\theta_1, \theta_2)\,\tilde{\pi}(\theta_1^{new} | \theta_2)}\right) = \min\left(1, \frac{\tilde{\pi}(\theta_1^{new}, \theta_2)\tilde{\pi}(\theta_1, \theta_2)/\tilde{\pi}}{\tilde{\pi}(\theta_1, \theta_2)\tilde{\pi}(\theta_1^{new}, \theta_2)/\tilde{\pi}(\theta}\right.$$

$$= 1.$$

Evning 30/11

(5.2) Consider a random walk on $\{0, 1, \ldots, k\}$ with reflecting boundaries. A walker earns k \$ on $0, k$, lose -1 \$ on $1, 2, \ldots, k-1$

How much will they gain on average if they walk for a long time.

Solution: Introduce $r(k) = \begin{cases} k & x = 0, k \\ -1 & x = 1, \ldots, k-1 \end{cases}$

we want $\lim_{n \to \infty} \dfrac{r(x_1) + r(x_2) + \ldots + r(x_{k-1})}{n} = E(r(x))$ where x has the

Stationary distribution. we need the transition matrix:

$\pi P = \pi$

$$P = \begin{array}{c|cccccc} & 0 & 1 & 2 & \cdots & & k \\ \hline 0 & 0 & 1 & 0 & \cdots & & 0 \\ 1 & 1/2 & 0 & 1/2 & \cdots & & 0 \\ 2 & 0 & 1/2 & 0 & 1/2 & \cdots & 0 \\ \vdots & & & \ddots & \ddots & \ddots & \\ & 0 & \cdots & & 1/2 & 0 & 1/2 \\ k & 0 & \cdots & & 0 & 1 & 0 \end{array}$$

Now we can solve $\pi P = \pi$

(might be easiest to start with small k, find a pattern; and then extend it)

This gives: $\pi \left(\frac{1}{2k}, \frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k}, \frac{1}{2k}\right)$

Calculate $E(r(x))$: $E(r(x)) = r(0) \cdot \frac{1}{2k} + r(1) \frac{1}{k} + \ldots + r(k-1) \frac{1}{k} + r(k) \frac{1}{2k}$

$$= k \cdot \frac{1}{2k} \cdot 2 - \frac{1}{k}(k-1) = \frac{1}{k}$$

The longterm average is $\frac{1}{k}$

$\boxed{5.4}$ The fibonacci sequence 1, 1, 2, 3, 5, 8, 13

which is described a recurence relation

$f_m = f_{m-1} + f_{m-2}$ $\quad f_1 = 1,\ f_2 = 1$

a) study the binary sequence with adjancet 1s

Show that the number of binary sequences of length m with no adjancent 1s is $g_m = f_{m+2}$

Solution: Example: $m=4$

$(1,1,0,1)$ Bad sequence

$(1,0,0,1)$ good sequence

Since we want a recurrence relation, try and express the current g_m in terms of previous ones

Fix first element

copy

$0: \underbrace{(0, \underbrace{\quad\quad}_{m-1 \text{ elements}})}_{m \text{ elements}}$ $\left\{ \text{For the entire sequence to be good we need just } m-1 \text{ elements to be good} \right\}$

This is g_{m-1}

similary to before

$1: \underbrace{(1, 0, \underbrace{\ldots}_{m-2})}_{m}$ $\quad$ copy $\{\}$ $\Rightarrow$ This is g_{m-2}

This then gives: $g_m = g_{m-1} + g_{m-2}$

we need initial conditions g_1: $(0), (1)$

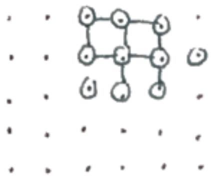
$g_1 = 2$, $\quad g_2$: $(0,0), (1,0), (0,1)$ $\quad$ possible with $(1,1)$ but is not good

$g_2 = 3$

Then we get $g_m = f_{m+2}$

Föreläsning 3/12

The Ising modell



m×m points

at node N the variable σ_v has value +1 or -1

The model consists of $\sigma = \{\sigma_v\}$

Energy: $\mathbb{E}(\sigma) = - \sum_{v \sim m} \sigma_v \sigma_m$ v~m means that v and m are neighbours

Gibbs distribution on all possible σ

$\pi(\sigma) \propto_\sigma \exp(-\beta E(\sigma))$ $\begin{cases} \beta \text{ is some constant} \\ \beta > 0: \text{ high energy. low prob.} \\ \beta = 0: \text{ uniform distribution} \\ \beta < 0: \text{ high energy, most probably} \end{cases}$

$\frac{1}{\beta}$ temperature.

Possible to simulate from this. model using Gibbs sampling

How does Gibbs sampling work here ?

- compute $\pi(\sigma_v = 1 \mid \sigma_{-v})$ σ_{-v} = all nodes expect v

  $\pi(\sigma_v = -1 \mid \sigma_{-v})$

$$\frac{\pi(\sigma_v = 1 \mid \sigma_{-v})}{\pi(\sigma_v = -1 \mid \sigma_{-v})} = \frac{\pi(\sigma_v = 1, \sigma_{-v})}{\pi(\sigma_v = -1, \sigma_{-v})} = \frac{\exp(-\beta E(\sigma_v = 1, \sigma_{-v}))}{\exp(-\beta E(\sigma_v = -1, \sigma_{-v}))} =$$

$$= \exp(-\beta |E(\sigma_v = 1, \sigma_{-v}) - E(\sigma_v = -1, \sigma_{-v})) = \exp(-\beta [\sum_{\substack{m \sim v \\ \sigma_v = 1}} \sigma_v \sigma_m + \sum_{\substack{m \sim v \\ \sigma_n = -1}} \sigma_v \sigma_m]) =$$

$$= \exp(+\beta 2 \sum_{m \sim v} \sigma_m)$$

Monotonicity

$X_{-v}^{(1)} \leq X_{-v}^{(j)} \leq X_{-v}^{(k)} = M$

$X_0^{(1)} \leq X_0^{(j)} \leq X_0^{(k)}$ $X_0^{(1)} = X_0^{(k)}$

Apply to the ising model

$\sigma \leq \tau$ iff $\sigma_v \leq \tau_v$ for all v

minimal element m: all -1

maximal element M: all $+1$

Show that g is monotone:

is possible as long as $\beta > 0$:

Proof: Assume that $\sigma \leq \tau$

update at v. Want to show that results respect same inequality

$$P(\sigma_v = 1 | \sigma_{-v}) = \frac{1}{1 + \exp(-\beta 2 \sum_{m\sim v} \sigma_m)} \leq \frac{1}{1+\exp(-\beta 2 \sum_{m \sim v} \tau_m)} = P(\tau_v = 1 | \tau_{-v})$$

· övning 3/12

5.10 a markov chain has transition matrix

$$P = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 1/2 & 1/2 & 0 \\ 2 & 1/2 & 0 & 1/2 \\ 3 & 0 & 1/2 & 1/2 \end{array}$$

Show this mc is monotone

i.e. $g(x,u) \leq g(y,u)$ if $x \leq y$

Solution: $g(1,u) = \begin{cases} 1 & u < 0.5 \\ 2 & 0.5 \leq u < 1 \end{cases}$ $\quad u \sim u[0,1]$

$$g(2,u) = \begin{cases} 1 & u < 0.5 \\ 3 & 0.5 \leq u < 1 \end{cases}$$

$$g(3,u) = \begin{cases} 2 & u < 0.5 \\ 3 & 0.5 \leq u < 1 \end{cases}$$

we want to show monotonicity

$$\left.\begin{array}{l} 1 \leq 2 \Rightarrow g(1,u) \leq g(2,u) \\ 2 \leq 3 \Rightarrow g(2,u) \leq g(3,u) \end{array}\right\} \Rightarrow \text{This chain is monotone.}$$

5.11 We have a RW on $\{1, \dots, n\}$

If we are on 1, then we go to 1 or 2 with prob. 1/2

if we are on n, then we go to n or n-1 with prob. 1/2

if we are in $i \neq 1, n \Rightarrow$ Then we go to $i+1$, $i-1$ with prob. 1/2

Show this chain is monotone

Solution: Find the transition matrix

$$P = \begin{array}{c|cccccc} & 1 & 2 & 3 & \dots & & n \\ \hline 1 & 1/2 & 1/2 & 0 & \dots & & 0 \\ 2 & 1/2 & 0 & 1/2 & 0 & \dots & 0 \\ 3 & 0 & 1/2 & 0 & 1/2 & 0 \dots & 0 \\ \vdots & \vdots & & \ddots & & \ddots & \vdots \\ \vdots & & & & 1/2 & 0 & 1/2 \\ n & 0 & \dots & & & 1/2 & 1/2 \end{array}$$

Write down update function:

$$g(1,u) = \begin{cases} 1 & u < 0.5 \\ 2 & 0.5 \leq u < 1 \end{cases}$$

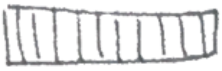
$$g(n,u) = \begin{cases} n-1 & u < 0.5 \\ n & 0.5 \leq u < 1 \end{cases}$$

$$g(x,u) = \begin{cases} x-1 & u < 0.5 \\ x+1 & 0.5 \leq u < 1 \end{cases}$$

we check the requirement: $x \leq y \Rightarrow g(x,u) \leq g(y,u)$

The condition holds

(5.13) Consider the lazy librarian from chapter 2


Books rented with probability P_j

When books are returned they are placed at the front. It can be shown that the time when all books have been selected is a strong stationary time.

Show $V(n) \leq \sum_{j=1}^{k} e^{-P_j n}$ we have k books.

Solution: If T is the strong stationary time.

$V(n) \leq P(T>n)$

T = the time when all books have been selected

$\Rightarrow$ Some book has not yet been taken at time n, $T>n$

The probability that book j has not yet been selected is $(1-P_j)^n$, $P(T>n) = P(\bigcup_j$ book j has not yet been rented$) \leq$

$\leq \sum_{j=1}^{k} (1-P_j)^n$

Since $1-x \leq e^{-x}\ \forall x \Rightarrow \sum_{j=1}^{k}(1-P_j)^n \leq \sum_{j=1}^{k} e^{-P_j n}$

(5.14) Show that the two definitions of total variation distance are equivelent

(1) $V(n) = \max_{i\in S} \max_{A \subseteq S} |P(X_i \in A \mid X_0 = i) - \pi_A|$

(2) $V(n) = \max_{i \in S} \frac{1}{2} \sum_{j\in S} |P_{ij}{}^{n} - \pi_j|$

Solution:
- We ignore the $\max_{i\in S}$ for now.
- we want to find the set A which maximies (1)
- Difficult to find max right away, easier to find a upper bound

Simplifying relation $W(A) = P(X_n \in A \mid X_0 = i)$, $V(A) = \pi_A$

Introduce: $B = \{x : W(x) \geq V(x)\}$, Take $A \subseteq S$

$W(A) - V(A) \overset{\text{Law of total prob.}}{=} W(A\cap B) + W(A \cap B^c) - V(A\cap B) - V(A \cap B^c) =$

$= W(A\cap B) - V(A\cap B) + W(A\cap B^c) - V(A\cap B^c) \longrightarrow$

In B^c then $V \geq w \Rightarrow \underbrace{w(A \cap B^c) - V(A \cap B^c)}_{\leq 0}$

$\Rightarrow w(A \cap B) - V(A \cap B) + w(A \cap B^c) - V(A \cap B^c) \leq w(A \cap B) - V(A \cap B)$

we have $w(A \cap B) - V(A \cap B) \leq w(A \cap B) - V(A \cap B) + \underbrace{w(A^c \cap B) - V(A^c \cap B)}_{\geq 0} =$

$= w(B) - V(B)$

So now we have $A < S$

$w(A) - V(A) \leq w(B) - V(B)$

In the exact same way:

$V(A) - w(A) \leq V(B^c) - w(B^c)$

We have two upper bounds: But they are equal since $w(B) - V(B) =$

$= V(B^c) - w(B^c) \iff w(B) + w(B^c) = V(B) + V(B^c) \iff 1 = 1$

we want a $A \subseteq S$ for which the inequality:

$w(A) - V(A) \leq w(B) - V(B)$ for which we have equality.

If $A = B$ (or B^c) then the inequality is an equality

$\Rightarrow B$ and B^c are the maximal sets.

$$\max_{A \subseteq S} |w(A) - V(A)| = w(B) - V(B) = \frac{1}{2}\left(w(B) - V(B) + V(B^c) - w(B^c)\right) =$$

$$= \frac{1}{2}\left(\sum_{x \in B} \underbrace{w(x) - V(x)}_{\geq 0} + \sum_{x \in B^c} \underbrace{V(x) - w(x)}_{\geq 0}\right) = \frac{1}{2} \sum_{x \in B} |w(x) - V(x) + \sum_{x \in B^c} |V(x) - w($$

$$= \frac{1}{2} \sum_{x \in S} |w(x) - V(x)| = \frac{1}{2} \sum_{x \in S} |P(X_n = x \mid X_0 = i) - \pi_x| =$$

$$= \frac{1}{2} \sum_{x \in S} |P_{ix}^n - \pi_x|$$

Föreläsning 4/12

ex, book 6.1

$P(N_{11} - N_9 > 65) = P(N_2 > 65)$ $N_2 \sim$ Poisson $(30 \cdot 2)$

compute in R: 1 - ppois (65, 30·2)

6.2 $P(N_2 = 18, N_7 = 70) = P(N_2 = 18, N_7 - N_2 = 52) =$

$= P(N_2 = 18) P(N_7 - N_2) = 52) = P(N_2 = 18) P(N_5 = 52) =$

$N_2 \sim$ Poisson $(10 \cdot 2)$ $N_5 \sim$ Poisson $(5 \cdot 10)$

$$= e^{-10 \cdot 2} \frac{(10 \cdot 2)^{18}}{18!} e^{-5 \cdot 10} \frac{(5 \cdot 10)^{52}}{52!} = 0.0045$$

Let x be the time before the first "arrival"

$P(x > t) = P(N_t = 0) = e^{-t\lambda}$

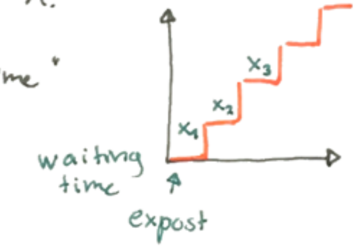
$P(x \leq t) = 1 - e^{-t\lambda}$ x ~ exponential (λ)

$\tilde{u}(x) = \lambda e^{-tx}$

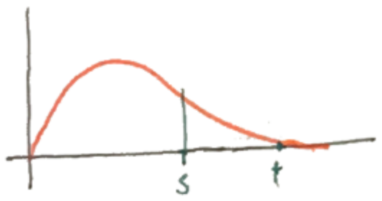
Poissonprocess - the different waiting times are exponential distributed with parameter λ.

$S_u = X_1 + \dots + X_u$ "arrival time"

$S_k - S_{k-1}$ interarival times



Random variable X is called memoryless if $P(X > s+t \mid X > s) = P(X > t)$
$t, s > 0$



probabillty that you get to t from s is the same as you get to t.

$X \sim \exp(\lambda)$

$$P(x > s+t / x > s) = \frac{P(x > s+t, x > t)}{P(x, s)} = \frac{P(x > s+t)}{P(x > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} =$$

$$= e^{-\lambda t} = P(x > t)$$

If exponential variables: the minimum of this

$M = \min(X_1, \dots, X_n)$, $M \sim \mathrm{Exp}(\lambda_1 + \dots + \lambda_n)$

$$P(M = X_k) = \frac{\lambda_k}{\lambda_1 + \dots + \lambda_n}$$

$$P(M > t) = P(X_1 > t, X_2 > t, \dots, X_n > t) = P(X_1 > t)P(X_2 > t) \dots P(X_n > t) =$$

$$= e^{-t\lambda_1} e^{-t\lambda_2} \dots\dots e^{-t\lambda_n} = e^{-t(\lambda_1 + \lambda_2 + \dots + \lambda_n)}$$

$S_n = X_1 + \dots + X_n$ where ind. $X_i \sim \exp(\lambda)$, $S_n \sim \mathrm{gamma}(n, \lambda)$

$E(S_n) = \frac{n}{\lambda}$ $\quad$ $E(\text{first arrival}) = \frac{1}{\lambda}$

$\mathrm{Var}(S_n) = \frac{n}{\lambda^2}$ $\quad$ $\mathrm{Std}(S_n) = \sqrt{\frac{n}{\lambda^2}}$

Brauching process:

c) Know distribution from a and have the answear if you have λ

example 3 in LN.

compute average - combine numerically.

not superexact, program youself or R function from book

In 60-minute game, find prob. that fourth goal happen in last 5 min

$S_n \sim \mathrm{gamma}(5, \lambda)$ $\lambda = 15$

$P(55 < S_i < 60) = \frac{1}{6} \int_{55}^{60} \left(\frac{1}{15}\right)^4 r^3 e^{-115} dt = 0.068$ see 283 in the book

pgamma(60, 4, 1/15) - pgamma(55, 4, 1/15) → 0.06766...

We write $f(h) = O(h)$ if $\lim_{n \to 0} \frac{f(h)}{h} = 0$

We write $f(h) = O(g(h))$ if $\lim_{n \to 0} \frac{f(h)}{g(h)} = 0$

Poisson Process is cocting process fullfilling

- $N_0 = 0$
- The process has stationary and independent increments
- We have
  - $P(N_n = 0) = 1 - \lambda n + O(h)$
  - $P(N_n = 1) = \lambda h + O(h)$
  - $P(N_n > \underline{\Phi}) = O(h)$

Def 1 implies def 3

$P(N_n = 0) = e^{-\lambda n}$ $\quad \frac{e^{-\lambda n} - (1 - \lambda_n)}{h} \xrightarrow[h \to 0]{} 0$

$P(N_n = 1) = e^{-\lambda h} \lambda h = \lambda h + O(h)$

$(= (1 - \lambda h + O(h)) \lambda h = \lambda h - (\lambda h)^2 + O(h) \lambda h = \lambda h + O(h))$

$P(N_n > 1) = 1 - P(N_n = 1) - P(N_n = 0) = 1 - (\lambda h + O(h)) - (1 - \lambda h + O(h)) = O(h)$ ish

Föreläsning 10/12

The Chapman-Kolmogorov Equations

$$P(s+t)_{ij} = P(X_{s+t} = i \mid X_0 = i) = \sum_k P(X_{s+t} = j \mid X_s = k) P(X_s = k \mid X_0 = i) =$$

$$= \sum_k P(X_t = j \mid X_0 = k) P(X_s = k \mid X_0 = i) =$$

$$= \sum_k P(s)_{ik} P_{kj}(t) = [P(s)P(t)]_{ij}$$

Poisson process example

$$P_{ij}(t) = P(N_{s+t} = j \mid N_s = i) = \frac{P(N_{s+t} = j, N_s = i)}{P(N_s = i)} = \frac{P(N_{s+t} - N_s = j-i, N_s = i)}{P(N_s = i)} =$$

$$= P(N_{s+t} - N_s = j-i) = P(N_t = j-i) =$$

$$\{N_t \sim \text{Poisson}(\lambda t)\} \qquad = e^{-\lambda t} \frac{\lambda t^{(j-i)}}{(j-i)!}$$

Memorylessness of T_i (Holding time)

$$P(T_i > s+t \mid X_0 = i) = P(T_i > s+t, T_i > s \mid X_0 = i) =$$

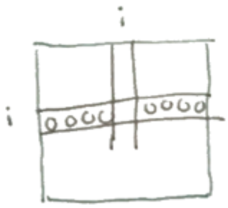
$$= P(T_i > s+t \mid X_s = i) P(T_i > s \mid X_0 = i) =$$

$$= P(T_i > t \mid X_0 = i) P(T_i > s \mid X_0 = i) \qquad \{T_i \sim \text{Exponential}(q_i)\}$$

$X \sim \text{Exponential}(\lambda)$

$\boxed{P(X > t) = e^{-\lambda t}}$ $\qquad$ $\boxed{\hat{\pi}(x) = \lambda e^{-\lambda t}}$

$P(X \leq t) = 1 - e^{-\lambda t}$



Alarm clocks

exponential(Q_{ij}), $i \neq j$

The derivate at zero

Assume $i \neq j$ (assume differentiability)

$$P_{ij}'(0) = \lim_{h\to 0^+} \frac{P_{ij}(h) - P_{ij}(0)}{h} = \lim_{h\to 0^+} \frac{P_{ij}(h)}{h} =$$

$$= \lim_{h\to 0^+} \frac{\text{expected \# of changes from i to j in } [0,h]}{h} = q_{ij}$$

$$P_{ii}'(0) = \lim_{h\to 0^+} \frac{P_{ii}(h) - \overbrace{P_{ii}(0)}^{=1}}{h} = \lim_{h\to 0^+} \frac{1 - \sum_{j\neq i} P_{ij}(h) - 1}{h} =$$

$$= -\lim_{h\to 0^+} \frac{1}{h} \sum_{j\neq i} P_{ij}(h) = -\sum_{j\neq i} q_{ij} = -q_i$$

Kolmogorov forward backward

$$P'(t) = \lim_{h\to 0^+} \frac{P(t+h) - P(t)}{h} = \lim_{h\to 0^+} \frac{P(t)P(h) - P(t)}{h} =$$

$$= \lim_{h\to 0^+} \frac{1}{h} P(t)(P(h) - I) = P(t)\left[\lim_{h\to 0^+} \frac{1}{h}(P(h) - P(0))\right] =$$

$$= P(t)\, P'(0) - P(t)\, Q$$

The matrix exponential

$$e^{A+B} = \sum_{n=0}^{\infty} \frac{1}{n!}(A+B)^n = \left\{ (A+B)^n = \sum_{n=0}^{n} \frac{n!}{k!(n-k)!} A^k B^{n-k} \right.$$

$$\boxed{\text{only } \overset{\text{easily}}{\text{proven}} \text{ if } AB = BA!}$$

$$\left. (A\times B)(A\times B) = AA + AB + BA + BB \qquad A^2 + 2AB + B^2 \right\}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} A^k B^{n-k} =$$

$$= \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{1}{k!} A^k \frac{1}{(n-k)!} B^{n-k} = \sum_{n=0}^{\infty} \sum_{s=0}^{\infty} \frac{1}{k!} A^k \frac{1}{s!} B^s =$$

$$= \sum_{n=0}^{\infty} \frac{1}{k} A^k \sum_{s=0}^{\infty} \frac{1}{s!} B^s = e^A e^B$$

$$\frac{d}{dt} e^{tA} = \frac{d}{dt} \sum_{n=0}^{\infty} \frac{1}{n!} (tA)^n = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{d}{dt} t^n A^n =$$

$$= \sum_{n=1}^{\infty} \frac{1}{n!} n t^{n-1} A^n = \left[\sum_{n=1}^{\infty} \frac{1}{(n-1)!} (tA)^{n-1}\right] A = \left[\sum_{n=0}^{\infty} \frac{1}{n!} (tA)^n\right] A = e^{tA} A$$

Computing the matrix exponential

$$e^{tQ} = \sum_{n=0}^{\infty} \frac{1}{n!} (t \cdot Q)^n = \sum_{n=0}^{\infty} \frac{1}{n!} (t SDS^{-1})^n = \sum_{n=0}^{\infty} \frac{1}{n!} t^n (SDS^{-1})^n =$$

$$= \left\{ \underbrace{SD\overbrace{S^{-1} \cdot S}^{I}DS^{-1} \dots SDS^{-1}}_{n \text{ times}} = SD^n S^{-1} \right\} = S\left[\sum_{n=0}^{\infty} \frac{1}{n!} t^n D^n\right] S^{-1} = S e^{tD} S^{-1}$$

D diagonal.

$$e^{tD} = \sum_{n=0}^{\infty} \frac{1}{n!} (t \cdot D)^n = \sum_{n=0}^{\infty} \frac{1}{n!} t^n \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_k \end{bmatrix}^n = \sum_{n=0}^{\infty} \frac{1}{n!} t^n \begin{pmatrix} \lambda_1^n & & \\ & \ddots & \\ & & \lambda_k^n \end{pmatrix} =$$

$$= \begin{bmatrix} \sum_{n=0}^{\infty} \frac{1}{n!} t^n \lambda_1^n & 0 \\ 0 & \sum_{n=k}^{\infty} \frac{1}{n!} t^n \lambda_k^n \end{bmatrix} = \begin{bmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{t\lambda_k} \end{bmatrix}$$

Övning 10/12

7.4 Customers arrive at a restaurant at a rate of 120/hour
The restaurant has one line, with three service workers taking orders.
Each worker take on average 1 min to serve a customer.
Denote X_t: The numbers of customers in the restaurant
Exhibit the generator matrix

Solution: We have two alarm clocks, one for customers entering the restaurant, one for customers beeing served.
Put the alarm clocks on equal scale

$\Rightarrow$ 120 customer/hour $\Rightarrow$ 2 customers/minute
(1 customer/30 sec)

$\Rightarrow$ service $\sim \text{Exp}(1)$
customers $\sim$ poi (2)

when there are 0 customers: $q_{01} = 2$

when there are 1 customer: $q_{10} = 1$, $q_{12} = 2$

when there are 2 customers:

The min (service 1, service 2) describes the rate from $2 \rightarrow 1$

min $(\text{Exp}(1), \text{Exp}(1)) \sim \text{Exp}(1+1)$

$q_{21} = 2 \quad q_{23} = 2$

$$Q = \begin{array}{c|cccccccc|}
 & 0 & 1 & 2 & 3 & 4 & 5 & \cdots & \\
0 & -2 & 2 & 0 & 0 & 0 & 0 & \cdots & \\
1 & 1 & -3 & 2 & 0 & 0 & \cdots & & \\
2 & 0 & 2 & -4 & 2 & 0 & \cdots & & \\
3 & 0 & 0 & 3 & -5 & 2 & \cdots & & \\
4 & 0 & 0 & 0 & 0 & 3 & -5 & 2 & \cdots \\
5 & \cdots & & \cdots & & & 3 & -5 & 2 \\
\vdots & & & & & & & & \ddots
\end{array}$$

7.5 If a customer enter and all service stations are full, they leave

Denite Y_t Number of service stations busy at time t

Exhibit the generator matrix

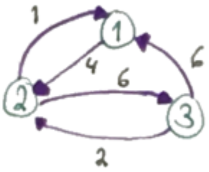
Solution: Proceed identically to 7.4

Since there is a 1-1 correspondence

i.e. 1 customer in restaurant | service station busy

$$Q = \begin{array}{c|cccc} & 0 & 1 & 2 & 3 \\ \hline 0 & -2 & 2 & 0 & 0 \\ 1 & 1 & -3 & 2 & 0 \\ 2 & 0 & 2 & -4 & 2 \\ 3 & 0 & 0 & 3 & -3 \end{array}$$

7.14 For the MC with transition graph



a) Find the generator matrix

b) Find the stationary distribution of the continuous time MC

c) Transition matrix of the embedded chain

Solution:

a) $$Q = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & -4 & 4 & 0 \\ 2 & 1 & -7 & 6 \\ 3 & 6 & 2 & 8 \end{array}$$

b) One can show that stationary π fullfills $\pi Q = 0$

Solve $Q^T \pi^T = 0 \Rightarrow \begin{cases} \pi_1 = 11/6 \cdot t \\ \pi_2 = 4/3 \cdot t \\ \pi_3 = t \end{cases} = \pi_1 + \pi_2 + \pi_3 = 1$

$(\pi_1, \pi_2, \pi_3) = (11/25, 8/25, 6/25$

c) We have $p_{ij} = \frac{q_{ij}}{q_i}$ for the transition probabilities

$$\tilde{P} = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 0 & 1 & 0 \\ 2 & 4/7 & 0 & 6/7 \\ 3 & 6/8 & 2/8 & 0 \end{array}$$

6-32 Investors purchase 1000 $ bonds at random according to a Poisson-process with parameter λ Interest rate is σ. The present value of the bonds at time t is $1000\, e^{-rs_i}$.

Show the total expected present value at t is $\dfrac{1000\lambda(1-e^{-rs_i})}{r}$

Solution: Introduce:

N_t : Total amount of bonds purchased at time t

$E\left(\sum_{i=1}^{N_t} e^{-rs_i}\right)$, s_i : arrival time

Use law of total expectation:

$$E\left(\sum_{i=1}^{N_t} e^{-rs_i}\right) = \sum_{n=0}^{\infty} E\left(\sum_{i=1}^{N_t} e^{-rs} \,\middle|\, N_t = n\right) \cdot P(N_t = n)$$

we have $E\left(\sum_{i=1}^{N_t} e^{-rs} \,\middle|\, N_t = n\right) = E\left(\sum_{i=1}^{n} e^{-rU_{(i)}}\right)$

where $U_i \sim unif[0,t]$, $U_{(i)}$ is the order statistic

Order stastic:

Normally: $U_1, U_2, \ldots, U_n$ ← order doesn't matter

order statistic: $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$

↑ smallest ↖ largest

Since $\sum e^{-rU_{(i)}}$ is independent of the ordering

$$\Rightarrow E\left(\sum_{i=1}^{n} e^{-rU_{(i)}}\right) = E\left(\sum_{i=1}^{n} e^{-rU_{(i)}}\right) = n\,E\left(e^{-rU}\right) = n \int_0^t \frac{e^{-rs}}{t}\, ds = \frac{n\cdot(1-e^{-rt})}{rt}$$

$$\sum_{n=0}^{\infty} E\left(\sum_{i=0}^{N_t} e^{-rs} \,\middle|\, N_t = n\right) \cdot P(N_t = n) = \sum_{n=0}^{\infty} \frac{n(1-e^{-rt})}{rt} \cdot \frac{(\lambda t)^n e^{-\lambda t}}{n!} =$$

$$= \frac{(1-e^{-rt})}{rt} \sum_{n=0}^{\infty} n\, \frac{(\lambda t)^n e^{-\lambda t}}{n!} = \frac{(1-e^{-rt})}{rt}\, \underbrace{\lambda t}_{E(N_t)} = \frac{\lambda(1-e^{-rt})}{r}$$

Forelasning 12/12

Countnous-time discrete state-space Markov chains

$P(t)$: $P(t)P(s) = P(t+s)$, $P(0) = I$

Holding times exponentially distributed: rates q_i

Alarm clocks: q_{ij} $(i \neq j)$

$$Q = \begin{bmatrix} -q_1 & q_{12} & \dots & q_{1n} \\ q_{21} & -q_2 & \dots & q_{2n} \\ \vdots & & & \\ q_{k1} & \dots & & q_k \end{bmatrix}$$

Rows sum to zero!

Embedded chain:

$$\tilde{P} = \begin{bmatrix} 0 & \frac{q_{12}}{q_1} & \dots & \frac{q_{1k}}{q_1} \\ \frac{q_{21}}{q_2} & 0 & \dots & \frac{q_{2k}}{q_2} \\ \vdots & & & \\ \frac{q_{k1}}{q_k} & \dots & \dots & 0 \end{bmatrix}$$

$P'(0) = Q$, $P'(t) = QP(t) = P(t)Q$

$P(t) = e^{tQ}$ $\boxed{R: \text{expm}}$

$$P(t) \xrightarrow[t\to\infty]{} \begin{bmatrix} \pi \\ \hline \\ \hline \\ \hline \\ \hline \end{bmatrix}$$

The fundamental limit theorem

Assume that v is a stationary distribution

$vP(t) = v$ for all t.

Differentiating w.r.a. t: $vP'(t) = 0$ for all t

$vP'(0) = 0$ $vQ = 0$

Assume that $0 = vQ$

$0 = vQP(t) = vP'(t) = \frac{d}{dt}(vP(t))$ so $vP(t)$ is constant

$vP(t) = vP(0) = vI = v$

Stationary distribution of the embedded chain

Assume π is stationary distribution for the continous chain

$$\underline{\pi Q = 0} \qquad \pi_j Q_j = \sum_{\substack{i=1 \\ i \neq j}}^{k} \pi_i Q_{ij}$$

Want to prove that Ψ is a stationary distr. for the embedded chain.

$$[\Psi P]_j = \Psi_j \qquad \sum_{i=1}^{k} c\pi_i Q_i \cdot \tilde{P}_{ij} = \sum_{\substack{i=1 \\ i \neq j}}^{k} c\pi_i Q_i \frac{q_{ij}}{Q_i} = c \sum_{\substack{i=1 \\ i \neq j}}^{k} \pi_i Q_{ij} = c\pi_j Q_j$$

M/M/1

Finding the stationary distribution π for the M/M/1 model:

Birth-and-death with birth rate λ and death rate μ

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} = \pi_0 \prod_{i=1}^{k} \frac{\lambda}{\mu} = \pi_0 \left(\frac{\lambda}{\mu}\right)^k$$

$$\left(\pi_0, \left(\frac{\lambda}{\mu}\right)\pi_0, \left(\frac{\lambda}{\mu}\right)^2 \pi_0, \ldots\right)$$

Birth-and-death process

The geometric distribution with parameter p: The number of trials you need to get one sucess, when the prob of succ. is (p) (parameter) so $1-p = \frac{\lambda}{\mu}$ $\boxed{p = 1 - \frac{\lambda}{\mu}}$

expected value is $\boxed{\frac{1}{p} = \frac{1}{1-\frac{\lambda}{\mu}}}$

övning 14/12

(7.18) For a general Birth-and-Death process with birthrate λ_i and deathrate μ.

Denote T_i : Time from state i to state $i+1$

a) $E(T_i) = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i} E(T_{i-1})$, $i = 1, 2, 3, \ldots$

Solution: Introduce: M: The birth process $\sim Exp(\lambda_i)$

D: The death process $\sim Exp(\mu_i)$

We condition on $M < D$

we have $P(M < D) = \frac{\lambda_i}{\lambda_i + \mu_i}$ since $P(M < D) = \int_0^\infty P(M < t) f_D(t)\, dt =$

$$= \int_0^\infty (1 - e^{-\lambda t}) \mu_i e^{-\mu_i t}\, dt = 1 - \frac{\mu_i}{\lambda_i + \mu_i} = \frac{\lambda_i}{\lambda_i + \mu_i}$$

We have $E(T_i) = E(T_i | M < D) P(M < D) + E(T_i | D < M) P(D < M)$

$$\Rightarrow E(T_i) = \frac{\lambda_i}{\lambda_i + \mu_i} E(T_i | M < D) + \frac{\mu_i}{\lambda_i + \mu_i} E(T_i | D < M)$$

We have: $E(T_i | M < D) = \frac{1}{\lambda_i}$

$$E(T_i | D < M) = E(T_{i-1}) + E(T_i) \quad (??)$$

Insert into (*) and solve for $E(T_i)$

$$\Rightarrow E(T_i) = \frac{\lambda_i}{\lambda_i + \mu_i} \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i + \mu_i} \left(E(T_i) + E(T_{i-1}) \right)$$

$$\Rightarrow E(T_i) \left(1 - \frac{\mu_i}{\mu_i + \lambda_i} \right) = \frac{1}{\lambda_i + \mu_i} + \frac{\mu_i}{\lambda_i + \mu_i} E(T_{i-1})$$

$$E(T_i) \frac{\lambda_i}{\mu_i + \lambda_i} = \frac{1}{\mu_i + \lambda_i}$$

makear inte mycket sense nu alltså....

b) Solve $E(T_i)$ when $\lambda_i = \lambda$, $\mu_i = \mu$

Solution:

We have a recursive equation: (no one can die only born)

$E(T_0) = \frac{1}{\lambda}$ initial condition

Use generating functions (check solutions)

Assume $E(T_i) = \dfrac{1 - \left(\frac{\mu}{\lambda}\right)^{i+1}}{\lambda - \mu}$ (or see geometric sum)

$$\text{We get } E(T_{i+1}) = \frac{1}{\lambda} + \frac{\mu}{\lambda} E(T_i) = \frac{1}{\lambda} + \frac{\mu}{\lambda}\left(\frac{1 - \left(\frac{\mu}{\lambda}\right)^{i+1}}{\lambda - \mu}\right)$$

$$\frac{1}{\lambda} + \frac{\frac{\mu}{\lambda} - \left(\frac{\mu}{\lambda}\right)^{i+2}}{\lambda - \mu} = \frac{\lambda - \mu + \mu - \frac{\mu^{i+2}}{\lambda^{i+1}}}{(\lambda - \mu)\lambda} = \frac{1 - \frac{\mu^{i+2}}{\lambda^{i+2}}}{\lambda - \mu}$$

7.23 We have 4 machines, two repair workers

Individual machines fail on average every 10 hours

It takes a repair on average 4 hours

Everything is independent.

a) Find the generator matrix.

Solution:

$$Q = \begin{array}{c c} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} -1/2 & 1/2 & 0 & 0 & 0 \\ 1/10 & -3/5 & 1/2 & 0 & 0 \\ 0 & 2/10 & -7/10 & 1/2 & 0 \\ 0 & 0 & 3/10 & -11/20 & 1/4 \\ 0 & 0 & 0 & 4/10 & -4/10 \end{bmatrix} \end{array}$$

Failure times $\sim \text{Exp}(1/10)$

Repair times $\sim \text{Exp}(1/4)$

b) What is the long term average?

Finite amount of scates and irreducible

$\Rightarrow$ limiting = stationary

To get the stationary dist. solve $\pi Q = 0$

$\Rightarrow \pi = (0.0191 ; 0.0955 ; 0.7888 ; 0.3979 ; 0.2487)$

$\Rightarrow$ Long term average

$$0 \cdot \pi_0 + 1 \cdot \pi_1 + 2\pi_2 + 3\pi_3 + 4\pi_4 = 2.76$$

c) If all 4 mathines are initially Working, find the probability that only two machines are working after 5 hours.

Solution: $P(X_5 = 2 \mid X_0 = 4) = (e^{5Q})_{42} \overset{\text{solve numerically}}{=} 0.188$

X_t : Number of machines working at time t.

7.26 Three maschines, three mechanics

Machines break down at rate of 1 per 24 hours.

Mechanics fix machines on average in 6 hours

Let X_t be the number of maschines working at time t.

Find the long term probability that all maschines are working.

Solution:

we want

$$Q = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} -3/6 & 3/6 & 0 & 0 \\ 1/24 & -3/8 & 2/6 & 0 \\ 0 & 2/24 & -6/24 & 1/6 \\ 0 & 0 & 3/24 & -3/24 \end{bmatrix} \end{array}$$

↑ fungerande maskiner

Mechanics fixing $\sim \text{Exp}(1/6)$

Machings breaking time $\sim \text{Exp}(1/24)$

We find stationary distribution:

Solve $\pi Q = 0$

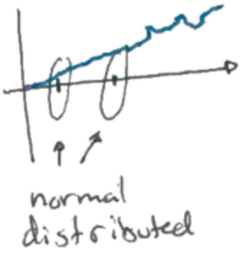
$$\Rightarrow \pi = \left(\frac{1}{125}, \frac{12}{125}, \frac{48}{125}, \underbrace{\frac{64}{125}}_{\text{de sökta}}\right)$$

Long term probability : $\pi_3 = \frac{64}{125}$

Föreläsning 17/12

Standard Brownian motion

infinity number of random variables



Corr?

{proper way:
use measure theory
(not in this course)}

We want it to be a continious line.

Compute:

$$\beta_1 + \beta_3 + 2\beta_7 = \beta_1 + \beta_3 + 2(B_7 - \beta_3 + \beta_3) = \beta_1 + 3\beta_3 \pm 2(\underbrace{\beta_7 - \beta_3}_{\text{independent}}) =$$

$$= \beta_1 + 3(B_3 - B_1 + \beta_1) + 2(B_7 - B_3) = 4\beta_1 + 3(B_3 - \beta_1) + 2(B_7 - \beta_3)$$

$B_1 \sim \text{Normal}(0,1)$

σ $4B_1 \sim \text{Normal}(0,16)$ σ^2

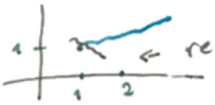
$B_3 - \beta_1 \sim \text{Normal}(0,2)$

$3(\beta_3 - \beta_1) \sim \text{Normal}(0, 2\cdot 9)$

$2(\beta_7 - \beta_3) \sim \text{Normal}(0, 4\cdot 4)$

This has distribution
normal$(0, 16+18+16)$
$=$ Normal$(0, 50)$

$P(\beta_2 > 0 \mid B_1 = 1) = P(B_1 > -1)$ →



⇐ resa tilk $0 = -1$ igrafen

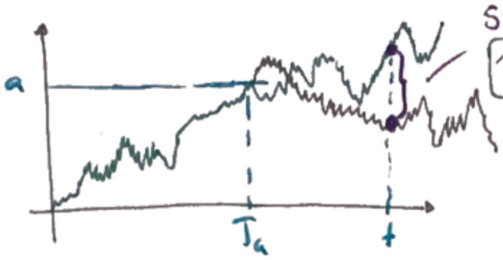
$\beta_1 \sim \text{Normal}(0,1)$

$$\text{COV}(\beta_s, B_t) = E(B_s B_t) - E(B_s)E(B_t) \overset{(t>s)}{=} E(B_s(B_t + B_s - B_s)) =$$

$$= E(B_s(B_t - B_s) + B_s B_s) \Rightarrow \underbrace{E(B_s)}_{\text{exp } 0}\underbrace{E(B_t - B_s)}_{\text{Exp } 0} + E[B_s^2] =$$

$$= \underbrace{E(B_s^2) - E(B_s)^2}_{\text{Var}} + \underbrace{E(B_s^2)}_{=0} = \text{Var}(B_s) = S$$

First hiting Time



(first time "hiting" a)

T_a = stocking time

(spelar ingen roll vad som händer sen)
samma sannolikhet

$$\boxed{P(B_t > a \mid T_a < t) = \frac{1}{2}}$$

$\underline{a > 0}$

$$\frac{1}{2} = P(B_t > a \mid T_a < t) =$$

$$= \frac{P(B_t > a, T_a < t)}{P(T_a < t)} =$$

$$= \frac{P(B_t > a)}{P(T_a < t)} = \{ P(T_a < t) =$$

$$= 2P(B_t > a) = 2\int_a^\infty \frac{1}{\sqrt{2\pi t}} \exp\left(\frac{-1}{2t} x^2\right) dx$$

$$P(T_a < t) = 2\int_{|a|}^\infty \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{1}{2t} x^2\right) dx = 2\int_{\frac{|a|}{\sqrt{t}}}^\infty \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} x^2\right) dx$$

$$\Rightarrow \pi(t) = \frac{|a|}{\sqrt{2\pi t^3}} \exp\left(-\frac{a^2}{2t}\right)$$

Maximum of Brownian motion

$\underline{a > 0}$

$$P(M_t > a) = P(M_t > a, B_t > a) + P(M_t > a, B_t \leq a)$$

$$= P(B_t > a) + P(B_t > a)$$

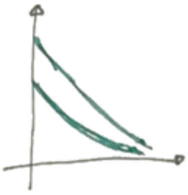
$$= 2P(B_t > a)$$

This is equal to $P(M_t > a, B_t \geq a)$

reflection principal

$$= P(M_t > a, B_t > a) =$$

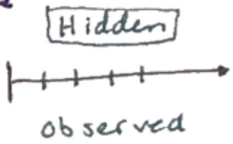
$$= P(B_t > a)$$



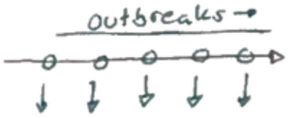
Föreläsning 18/12

Sammanfattning

Hidden markov models

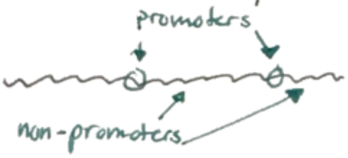


Example 1 Detecting west-nile virus outbreaks in horses in France



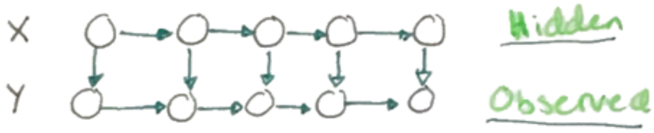
weekly reports on from
veterinaries on health of horses

Example: Detection of promoter sequence in DNA



Observed: DNA - sequence

Mathematical models



Questions:

→ Given "training data" where both X_i and Y_i are observed, Find the parameters of the model. Then use this model for predictions.

→ Assuming given parameters, and observed Y_i, make inference about the x.

- **Simpler Question:** one Markov chain, observed states; infer parameters

- **For the second Question:** Given $Y_0, \dots, Y_T$ observed, find the marginal posterior distribution for each X_i.

Inference for a time-homogeneous discrete-time markov chain, with finite state space (Lecture notes chapter 2)

Example 4-state chain

observed values 1 2 2 1 3 4 4 1 4

count the transitions

	1	2	3	4
1	0	1	1	1
2	1	1	0	0
3	0	0	0	1
4	1	0	0	1

Altenative to using frequencies:

Bayesian inference:

Transition matrix $p = \begin{bmatrix} p_{11} & \cdots & p_{14} \\ \vdots & & \vdots \\ p_{41} & \cdots & p_{44} \end{bmatrix}$

what is the probability of data given parameters multinomial distribution : $x = (c_1, \dots, c_4)$, parameters $p = (p_{11}, \dots, p_{14})$

$$\pi(x \mid p) = \frac{(c_1 + c_2 + c_3 + c_4)!}{c_1!\, c_2!\, c_3!\, c_4!} \, p_1^{c_1} \cdot p_2^{c_2} \cdot p_3^{c_3} \cdot p_4^{c_4}$$

Prior: Dirichlet distribution

$p = (p_{i1}, \dots, p_{i4})$ has a dirichlet distribution with parameters $\alpha = (\alpha_1, \dots, \alpha_4)$ if

$$\pi(p \mid \alpha) = \frac{\Gamma(\alpha_1 + \dots + \alpha_4)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3 \quad \Gamma(\alpha_4)} \, p_{i1}^{\alpha_1 - 1} \cdots p_{i4}^{\alpha_4 - 1}$$

$$\pi(p \mid x) \propto_p \pi(x \mid p)\, \pi(p) \propto_p p_{i1}^{\alpha_1 + c_1 - 1} \, p_{i2}^{\alpha_2 + c_2 - 1} \, p_{i3}^{\alpha_3 + c_3 - 1} \, p_{i4}^{\alpha_4 + c_4 - 1}$$

$p \mid x$: Dirichlet$(\alpha_1 + c_1, \alpha_2 + c_2, \alpha_3 + c_3, \alpha_4 + c_4)$

Assume: $\alpha_1 = \frac{1}{4}$

Result for $p = (p_{11}, p_{12}, p_{13}, p_{14})$

$$p_i \mid \text{Data} \sim \text{Dirichlet}\left(\tfrac{1}{4} + 0, \tfrac{1}{4} + 1, \underbrace{\tfrac{1}{4} + 1}_{}, \tfrac{1}{4} + 1\right)$$

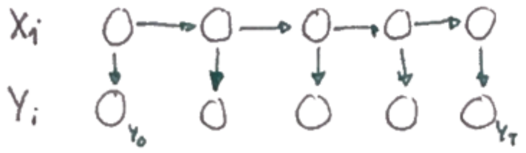
For example: p_{13} has a beta distribution

$$p_{13} \sim \text{Beta}\left(\tfrac{1}{4} + 1, \tfrac{1}{4} + \tfrac{5}{4} + \tfrac{5}{4}\right)$$

Expectation of p_{13} : $\dfrac{5/4}{5/4 + 11/4} = \dfrac{5}{16} = 0.3125$

$$\left\{ (p_{11}, p_{12} + p_{13}, p_{14}) \sim \text{Dirichlet}\left(\tfrac{1}{4} + 0, \tfrac{1}{4} + 1 + \tfrac{1}{4} + 1, \tfrac{1}{4} + 1\right) \right\}$$

The Forward-Backward algorithm



Discrete time, finite state spaces, known parameters

objective given $Y_0, \dots, Y_T$ and the marginal posterior distribution for each X. $\pi(X_i \mid Y_1, Y_2, \dots, Y_T)$

Forward algorithm:

Objective: For $i = 0, 1, \dots, T$, compute $\pi(X_i \mid Y_1, \dots, Y_i)$ using recursion:

For $i = 0$ $\pi(X_0 \mid Y_0) \propto_{Y_0} \pi(Y_0 \mid X_0)\, \pi(X_0)$

For $i > 0$ $\pi(X_i \mid Y_0, Y_1) \propto_{X_i} \pi(Y_i \mid X_1, Y_0 \dots Y_{i-1})\, \pi(X_1 \mid Y_0 \dots Y_{i-1}) =$

$$= \pi(Y_i \mid X_i) \sum_{X_{i+1}} \pi(X_i \mid X_{i-1})\, \pi(X_{i-1} \mid Y_0 \dots Y_{i-1})$$

Backward algorithm:

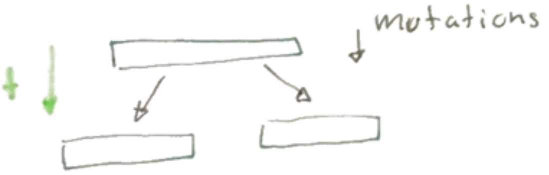
Objective: For $i = T, T-1, \dots, 0$ compute $\pi(Y_{i+1}, Y_{i+2}, \dots Y_T \mid X_i)$

$$\pi(Y_{i+1} \dots Y_T \mid X_i) = \sum_{X_{i+1}} \pi(Y_{i+1} \mid X_{i+1})\, \pi(Y_{i+2} \dots Y_T \mid X_{i+1})\, \pi(X_{i+1} \mid X_i)$$

Forward-Backward:

$$\pi(X_i \mid Y_0 \dots Y_T) \propto_{X_i} \pi(Y_{i+1} \dots Y_T \mid X_i)\, \pi(X_i \mid Y_0 \dots Y_i)$$

Building on Example 7.11 in Dobrow



mutation model: a rate of R per year of any mutation ($A \to T$, $A \to C$, $C \to G$

$$Q = \begin{bmatrix} -3R & R & R & R \\ R & -3R & R & R \\ R & R & -3R & R \\ R & R & R & -3R \end{bmatrix} \qquad P(t) = e^{tQ}$$

Given some rate differences what is the estimate for R?

prob that two nucleotides are the same after time t.

$$Q_t = (P_{AA}(t))^2 + (P_{CA}(t))^2 + (P_{GA}(t))^2 + (P_{TA}(t))^2$$

$$Q = \underset{\overset{\shortparallel}{S}}{\begin{bmatrix} -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}} \begin{bmatrix} -4R & 0 & 0 & 0 \\ 0 & -4R & 0 & 0 \\ 0 & 0 & -4R & 0 \\ 0 & 0 & 0 & -4R \end{bmatrix} S^{-1} \qquad \longrightarrow\!|$$

$$P(t) = e^{tQ} = S \begin{bmatrix} e^{-4Rt} & 0 & 0 & 0 \\ 0 & e^{-4Rt} & 0 & 0 \\ 0 & 0 & e^{-4Rt} & 0 \\ 0 & 0 & 0 & e^{-4Rt} \end{bmatrix} S^{-1} = \frac{1}{4} \begin{bmatrix} 1+e^{-4Rt} & 1-e^{-4Rt} & \dots & \\ & \ddots & \text{samme} & \\ \text{samme} & & \ddots & \\ & & & 1+e^{-4Rt} \\ & & & \text{R sama} \end{bmatrix}$$

$$Q_R = \frac{1}{4} + \frac{3}{4} e^{-8Rt}$$