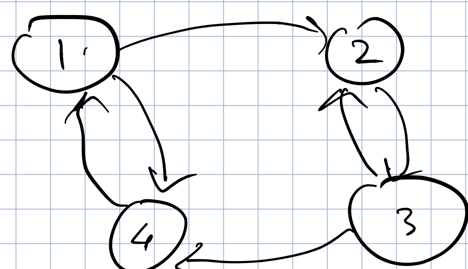
Markov chain theory continued

So far, we have learned about
1. Irreducibility
2. Stationarity

Now, another important concept is recurrence.

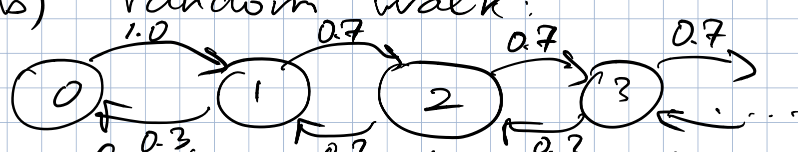
Example:

a) irreducible Markov chain on a finite state space



If a frog starts on island 1, then in will visit this island again with prob 1.

b) random walk:



infinite, countable state space: $\{0, 1, 2, \ldots\}$

it will wonder off to ∞

Prob (starting in 1 and returning

to 1 eventually) $\neq 1.0$

Def. Markov chain is <u>recurrent</u> if when it starts from any state the chain returns to the same with prob 1 (eventually). The chain is positive recurrent if the return time has finite expectation.

<u>Proposition</u> If Markov chain is irreducible and positive recurrent, then it has a unique stationary distribution.

Sad fact: there is no practical way of checking if a Markov chain is positive recurrent, except one:

Stationary distribution criterion:

Irreducible Markov chain is positive recurrent $\Longleftrightarrow$ it has a stationary distribution

<u>Theorem</u> (Ergodic Theorem)

$\{X_n\}$ - irreducible and positive recurrent Markov chain and some function f that takes X_n and sends to a number (e.g. $f(X_n) = \begin{cases} 1 & \text{if } X_n = 3 \\ 0 & \text{otherwise} \end{cases}$)

stationary distribution $\qquad f(X_n) = \mathbb{1}\{X_n=3\}$)

$\sum_{i\in E} |f(i)|\, \pi_i < \infty$. Then

state space

$$\underbrace{\frac{1}{N}\sum_{k=1}^{N} f(X_k)}_{} \longrightarrow \underbrace{\sum_{i\in E} f(i)\,\pi_i}_{} =$$

(e.g., proportion of time we visited 3 $\qquad \pi_3$ = stationary probability of state 3.

$= E_\pi[f(X)]$, where $X \sim \pi$

If $f(x) = \mathbb{1}_{\{x=3\}}$, then $\sum_{i\in E} f(i)\,\pi_i =$

$= \underset{0}{f(1)}\,\pi_1 + \underset{0}{f(2)}\,\pi_2 + \underset{1}{f(3)}\,\pi_3 + \dots$

This means that we can approximate stationary distribution of an irreducible and positive recurrent Markov chain by simulating and averaging

Markov chain Monte Carlo

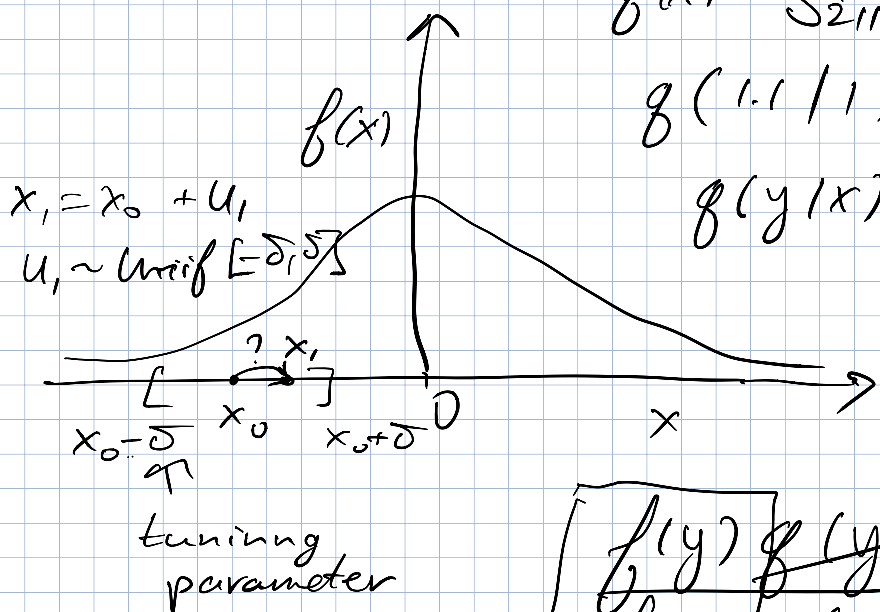
Big idea: Suppose I want to study some distribution: I want to know its mean, variance, tails, quantiles, etc. Artificially, design a Markov chain in such a way that the distribution of interest is the stationary distribution of the Markov chain and then simulate and use ergodic theorem

M-H-lab

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$q(1.1 \mid 1) = q(1 \mid 1.1)$$

$$q(y \mid x) = q(x \mid y)$$



$$\boxed{\frac{f(y)\,\cancel{q(y \mid x)}}{f(x)\,\cancel{q(x \mid y)}}}$$

$$\frac{\cancel{\frac{1}{\sqrt{2\pi}}}}{\cancel{\frac{1}{\sqrt{2\pi}}}} \; \frac{e^{-\frac{y^2}{2}}}{e^{-\frac{x^2}{2}}} =$$

$$= e^{[x^2 - y^2]/2}$$

Combining Markov kernels

Suppose we have 2 M-H algorithms

P_1 M-H 1 : $y = x_n + \varepsilon_n, \quad \varepsilon_n \sim Unif[-\delta_1, \delta_1]$

e.g. $\delta_1 = 5.0$

P_2 M-H 2 : $y = x_n + \varepsilon_n, \quad \varepsilon_n \sim Unif[-\delta_2, \delta_2)$

e.g. $\delta_2 = 0.5$

x_0 = initial value

$x_1 \leftarrow$ use M-H1

$x_2 \leftarrow$ use M-H2

$x_3 \leftarrow$ use M-H1

$x_4 \leftarrow$ use M-H2

⋮

Sequential combination or scan of kernels

kernel = transition prob matrix

$$\underline{\pi^T P} = \underbrace{\pi^T P_1}_{\pi^T} \cdot P_2 = \underbrace{\pi^T P_2}_{\pi^T} = \underline{\pi^T}$$

Another idea:

x_0 = initial value

$x_1 \leftarrow$ flip a biased coin, which will tell me whether to use M-H1 or M-H2

$x_2 \leftarrow$ ——— " ———

⋮

random combination or scan

$P = \alpha_1 P_1 + \alpha_2 P_2$, where $0 < \alpha_1, \alpha_2 < 1$

$\alpha_1 + \alpha_2 = 1$

$$\boxed{\bar{\pi}^T P} = \alpha_1 \underbrace{\bar{\pi}^T P_1}_{\bar{\pi}^T} + \alpha_2 \underbrace{\bar{\pi}^T P_2}_{\bar{\pi}^T} = \alpha_1 \bar{\pi}^T + \alpha_2 \bar{\pi}^T =$$

$$= \underbrace{(\alpha_1 + \alpha_2)}_{1} \bar{\pi}^T = \boxed{\bar{\pi}^T}$$

Multivariate target distributions

$\bar{x} = (x_1, \dots, x_m) \in \mathbb{R}^m = \mathbb{R}^1 \times \mathbb{R}^1 \times \dots \times \mathbb{R}^1$

$f(\bar{x})$ – target distribution

One M-H idea:

$$\bar{y} = \bar{x}_n + \bar{\varepsilon}_n, \quad \varepsilon_n \sim MVN(\bar{0}, \Sigma)$$

Doesn't work very well, because it will reject a lot unless you come up with a really good Σ matrix

In practice, we often update one component at a time

M-H 1 : updates component 1 in $\bar{x} = (x_1, x_2, \dots, x_m)$ (e.g. using uniform random walk

M-H 2 : $\bar{x}_{current} \to \bar{x}_{proposed}$, where these vectors differ only in component 2

M-H 2 : —— $\|$ —— component 2, only, in component

⋮

—— $\|$ —— component m

Special M-H algorithm =

= Gibbs sampling

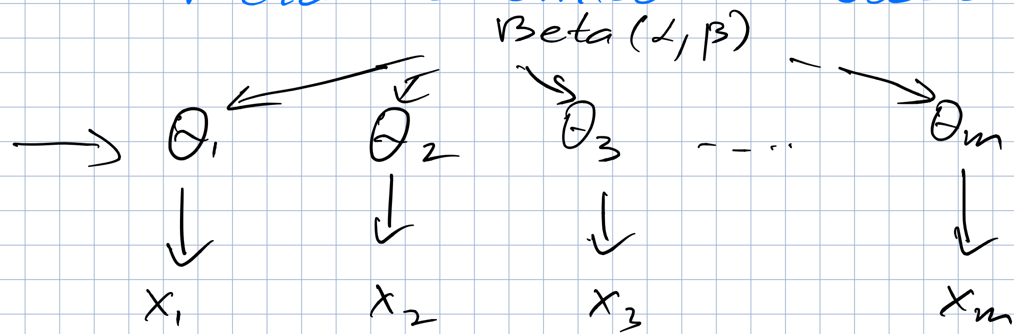
Gibbs 1 $\qquad x_1 \mid x_2, x_3, \dots, x_m$

Gibbs 2 $\qquad x_2 \mid x_1, x_3, \dots, x_m$

M-H 3 $\qquad$ update x_3 using M-H

$\vdots$

Beta-binomial model



$x_i \sim Bin(n_i, \theta_i)$

α, β – hyperpameters, unknown and we would like to estimate them togeter with $\theta_1, \theta_2, \dots, \theta_m$

likelihood: $Pr(\bar{x} \mid \bar{\theta}, \cancel{\alpha, \beta}) =$

$$= Pr(x_1 \mid \theta_1) \times Pr(x_2 \mid \theta_2) \times \dots \times Pr(x_m \mid \theta_m)$$

$$= \binom{n_1}{x_1} \theta_1^{x_1} (1-\theta_1)^{n_1 - x_1} \times \dots$$

constants with respect to $\vec{\theta}, \alpha, \beta$, so we drop them

We will use Gibbs sampling to update all θs, but α and β will be updated using M-H steps with the following proposal:

$$\alpha, \beta > 0$$

$$\alpha^{\text{proposed}} = \alpha^{\text{current}} \cdot e^{\lambda_\alpha (U_1 - \frac{1}{2})},$$

where $U_1 \sim \text{Unif}[0,1]$.

$$\beta^{\text{proposed}} = \beta^{\text{current}} \cdot e^{\lambda_\beta (U_2 - \frac{1}{2})}$$

$U_2 \sim \text{Unif}[0,1]$