Lecture 6
Markov Chain Monte Carlo - Gibbs Sampling

Guilherme J. M. Rosa

University of Wisconsin-Madison

Mixed Models in Quantitative Genetics

SISG, Seattle

22 – 24 July 2020

Table A.1 Continuous distributions

Distribution	Notation	Parameters	Density function	Mean, variance, and mode
Uniform	$\theta \sim \text{U}(a,b)$ $p(\theta) = \text{U}(\theta\|a,b)$	boundaries a,b with $b > a$	$p(\theta) = \frac{1}{b-a},\ \theta \in [a,b]$	$\text{E}(\theta) = \frac{a+b}{2}$, $\text{var}(\theta) = \frac{(b-a)^2}{12}$ no mode
Normal	$\theta \sim \text{N}(\mu,\sigma^2)$ $p(\theta) = \text{N}(\theta\|\mu,\sigma^2)$	location μ scale $\sigma > 0$	$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(\theta-\mu)^2\right)$	$\text{E}(\theta) = \mu$, $\text{var}(\theta) = \sigma^2$ $\text{mode}(\theta) = \mu$
Multivariate normal	$\theta \sim \text{N}(\mu,\Sigma)$ $p(\theta) = \text{N}(\theta\|\mu,\Sigma)$ (implicit dimension d)	symmetric, pos. def., $d \times d$ cov. matrix Σ	$p(\theta) = (2\pi)^{-d/2}\|\Sigma\|^{-1/2}$ $\times \exp\left(-\frac{1}{2}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)\right)$	$\text{E}(\theta) = \mu$, $\text{var}(\theta) = \Sigma$ $\text{mode}(\theta) = \mu$
Gamma	$\theta \sim \text{Gamma}(\alpha,\beta)$ $p(\theta) = \text{Gamma}(\theta\|\alpha,\beta)$	shape $\alpha > 0$ inverse scale $\beta > 0$	$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta},\ \theta > 0$	$\text{E}(\theta) = \frac{\alpha}{\beta}$ $\text{var}(\theta) = \frac{\alpha}{\beta^2}$ $\text{mode}(\theta) = \frac{\alpha-1}{\beta}$, for $\alpha \geq 1$
Inverse-gamma	$\theta \sim \text{Inv-gamma}(\alpha,\beta)$ $p(\theta) = \text{Inv-gamma}(\theta\|\alpha,\beta)$	shape $\alpha > 0$ scale $\beta > 0$	$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\beta/\theta},\ \theta > 0$	$\text{E}(\theta) = \frac{\beta}{\alpha-1}$, for $\alpha > 1$ $\text{var}(\theta) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$, $\alpha > 2$ $\text{mode}(\theta) = \frac{\beta}{\alpha+1}$
Chi-square	$\theta \sim \chi^2_\nu$ $p(\theta) = \chi^2_\nu(\theta)$	deg. of freedom $\nu > 0$	$p(\theta) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}\theta^{\nu/2-1}e^{-\theta/2},\ \theta > 0$ same as $\text{Gamma}(\alpha = \frac{\nu}{2}, \beta = \frac{1}{2})$	$\text{E}(\theta) = \nu$, $\text{var}(\theta) = 2\nu$ $\text{mode}(\theta) = \nu - 2$, for $\nu \geq 2$
Inverse-chi-square	$\theta \sim \text{Inv-}\chi^2_\nu$ $p(\theta) = \text{Inv-}\chi^2_\nu(\theta)$	deg. of freedom $\nu > 0$	$p(\theta) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}\theta^{-(\nu/2+1)}e^{-1/(2\theta)},\ \theta > 0$ same as $\text{Inv-gamma}(\alpha = \frac{\nu}{2}, \beta = \frac{1}{2})$	$\text{E}(\theta) = \frac{1}{\nu-2}$, for $\nu > 2$ $\text{var}(\theta) = \frac{2}{(\nu-2)^2(\nu-4)}$, $\nu > 4$ $\text{mode}(\theta) = \frac{1}{\nu+2}$
Scaled inverse-chi-square	$\theta \sim \text{Inv-}\chi^2(\nu, s^2)$ $p(\theta) = \text{Inv-}\chi^2(\theta\|\nu, s^2)$	deg. of freedom $\nu > 0$ scale $s > 0$	$p(\theta) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}s^\nu\theta^{-(\nu/2+1)}e^{-\nu s^2/(2\theta)},\ \theta > 0$ same as $\text{Inv-gamma}(\alpha = \frac{\nu}{2}, \beta = \frac{\nu}{2}s^2)$	$\text{E}(\theta) = \frac{\nu}{\nu-2}s^2$ $\text{var}(\theta) = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}s^4$ $\text{mode}(\theta) = \frac{\nu}{\nu+2}s^2$
Exponential	$\theta \sim \text{Expon}(\beta)$ $p(\theta) = \text{Expon}(\theta\|\beta)$	inverse scale $\beta > 0$	$p(\theta) = \beta e^{-\beta\theta},\ \theta > 0$ same as $\text{Gamma}(\alpha = 1, \beta)$	$\text{E}(\theta) = \frac{1}{\beta}$, $\text{var}(\theta) = \frac{1}{\beta^2}$ $\text{mode}(\theta) = 0$
Wishart	$W \sim \text{Wishart}_\nu(S)$ $p(W) = \text{Wishart}_\nu(W\|S)$ (implicit dimension $k \times k$)	deg. of freedom ν symmetric, pos. def. $k \times k$ scale matrix S	$p(W) = \left(2^{\nu k/2}\pi^{k(k-1)/4}\prod_{i=1}^k \Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1}$ $\times\|S\|^{-\nu/2}\|W\|^{(\nu-k-1)/2}$ $\times \exp\left(-\frac{1}{2}\text{tr}(S^{-1}W)\right)$, W pos. def.	$\text{E}(W) = \nu S$
Inverse-Wishart	$W \sim \text{Inv-Wishart}_\nu(S^{-1})$ $p(W) = \text{Inv-Wishart}_\nu(W\|S^{-1})$ (implicit dimension $k \times k$)	deg. of freedom ν symmetric, pos. def. $k \times k$ scale matrix S	$p(W) = \left(2^{\nu k/2}\pi^{k(k-1)/4}\prod_{i=1}^k \Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1}$ $\times\|S\|^{\nu/2}\|W\|^{-(\nu+k+1)/2}$ $\times \exp\left(-\frac{1}{2}\text{tr}(SW^{-1})\right)$, W pos. def.	$\text{E}(W) = (\nu - k - 1)^{-1}S$

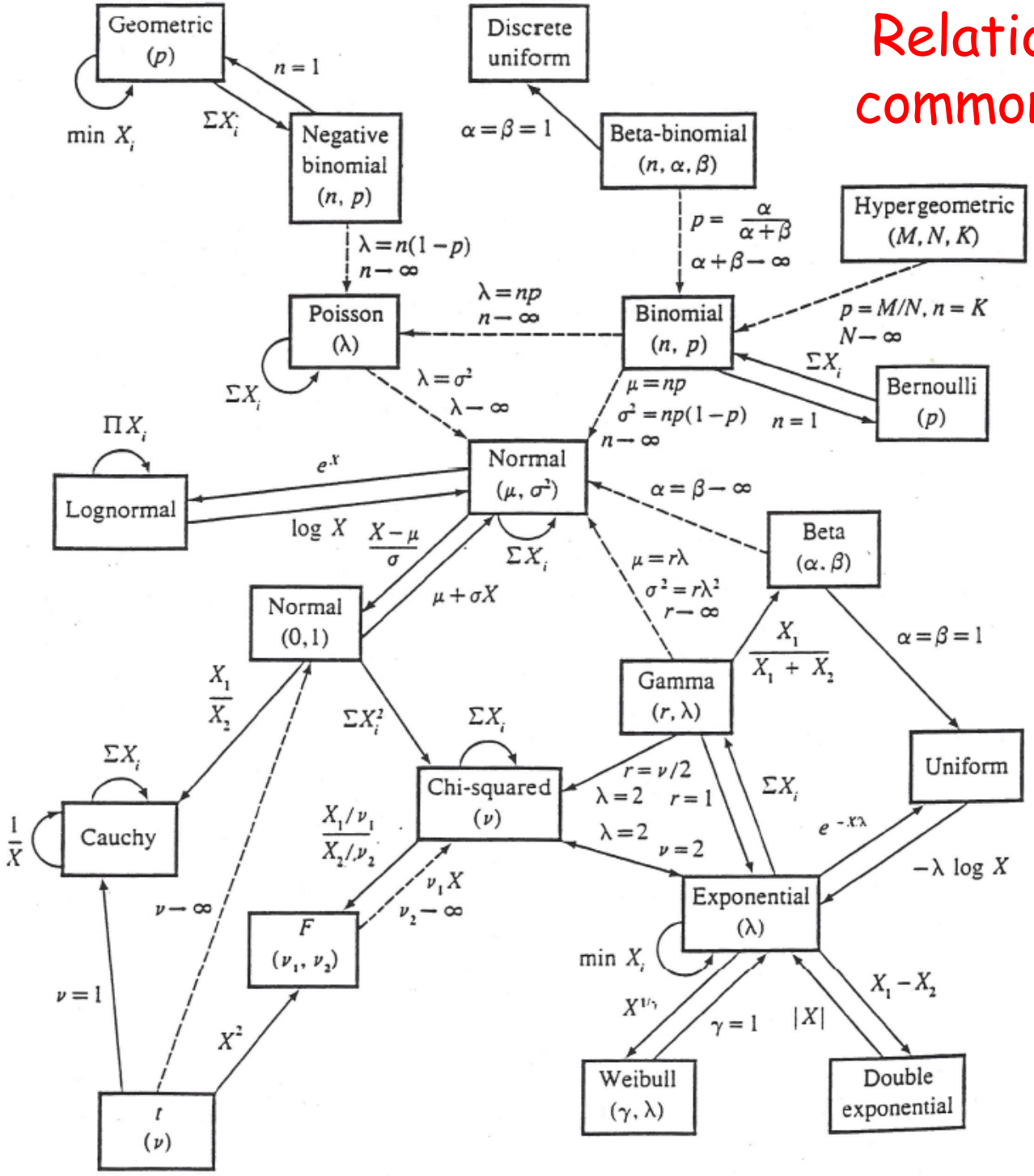
Table A.1 **Continuous distributions** *continued*

Distribution	Notation	Parameters	Density function	Mean, variance, and mode
Student-t	$\theta \sim t_\nu(\mu, \sigma^2)$ $p(\theta) = t_\nu(\theta\|\mu, \sigma^2)$ t_ν is short for $t_\nu(0,1)$	deg. of freedom $\nu > 0$ location μ scale $\sigma > 0$	$p(\theta) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}\sigma}(1 + \frac{1}{\nu}(\frac{\theta-\mu}{\sigma})^2)^{-(\nu+1)/2}$	$\mathrm{E}(\theta) = \mu$, for $\nu > 1$ $\mathrm{var}(\theta) = \frac{\nu}{\nu-2}\sigma^2$, for $\nu > 2$ $\mathrm{mode}(\theta) = \mu$
Multivariate Student-t	$\theta \sim t_\nu(\mu, \Sigma)$ $p(\theta) = t_\nu(\theta\|\mu, \Sigma)$ (implicit dimension d)	deg. of freedom $\nu > 0$ location $\mu = (\mu_1, .., \mu_d)$ symmetric, pos. def. $d \times d$ scale matrix Σ	$p(\theta) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}}\|\Sigma\|^{-1/2}$ $\times(1 + \frac{1}{\nu}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu))^{-(\nu+d)/2}$	$\mathrm{E}(\theta) = \mu$, for $\nu > 1$ $\mathrm{var}(\theta) = \frac{\nu}{\nu-2}\Sigma$, for $\nu > 2$ $\mathrm{mode}(\theta) = \mu$
Beta	$\theta \sim \mathrm{Beta}(\alpha, \beta)$ $p(\theta) = \mathrm{Beta}(\theta\|\alpha, \beta)$	'prior sample sizes' $\alpha > 0, \beta > 0$	$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$ $\theta \in [0,1]$	$\mathrm{E}(\theta) = \frac{\alpha}{\alpha+\beta}$ $\mathrm{var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $\mathrm{mode}(\theta) = \frac{\alpha-1}{\alpha+\beta-2}$
Dirichlet	$\theta \sim \mathrm{Dirichlet}(\alpha_1, .., \alpha_k)$ $p(\theta) = \mathrm{Dirichlet}(\theta\|\alpha_1, .., \alpha_k)$	'prior sample sizes' $\alpha_j > 0$; $\alpha_0 \equiv \sum_{j=1}^k \alpha_j$	$p(\theta) = \frac{\Gamma(\alpha_1+\cdots+\alpha_k)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)}\theta_1^{\alpha_1-1}\cdots\theta_k^{\alpha_k-1}$ $\theta_1, .., \theta_k \geq 0; \sum_{j=1}^k \theta_j = 1$	$\mathrm{E}(\theta_j) = \frac{\alpha_j}{\alpha_0}$ $\mathrm{var}(\theta_j) = \frac{\alpha_j(\alpha_0-\alpha_j)}{\alpha_0^2(\alpha_0+1)}$ $\mathrm{cov}(\theta_i, \theta_j) = -\frac{\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}$ $\mathrm{mode}(\theta_j) = \frac{\alpha_j-1}{\alpha_0-k}$

Table A.2 **Discrete distributions**

Distribution	Notation	Parameters	Density function	Mean, variance, and mode
Poisson	$\theta \sim \mathrm{Poisson}(\lambda)$ $p(\theta) = \mathrm{Poisson}(\theta\|\lambda)$	'rate' $\lambda > 0$	$p(\theta) = \frac{1}{\theta!}\lambda^\theta \exp(-\lambda)$ $\theta = 0, 1, 2, \ldots$	$\mathrm{E}(\theta) = \lambda$, $\mathrm{var}(\theta) = \lambda$ $\mathrm{mode}(\theta) = \lfloor\lambda\rfloor$
Binomial	$\theta \sim \mathrm{Bin}(n, p)$ $p(\theta) = \mathrm{Bin}(\theta\|n, p)$	'sample size' n (pos. integer) 'probability' $p \in [0,1]$	$p(\theta) = \binom{n}{\theta}p^\theta(1-p)^{n-\theta}$ $\theta = 0, 1, 2, \ldots, n$	$\mathrm{E}(\theta) = np$ $\mathrm{var}(\theta) = np(1-p)$ $\mathrm{mode}(\theta) = \lfloor(n+1)p\rfloor$
Multinomial	$\theta \sim \mathrm{Multin}(n; p_1, .., p_k)$ $p(\theta) = \mathrm{Multin}(\theta\|n; p_1, .., p_k)$	'sample size' n (pos. integer) 'probabilities' $p_j \in [0,1]$: $\sum_{j=1}^k p_j = 1$	$p(\theta) = \binom{n}{\theta_1\ \theta_2\cdots\theta_k}p_1^{\theta_1}\cdots p_k^{\theta_k}$ $\theta_j = 0, 1, 2, \ldots, n; \sum_{j=1}^k \theta_j = n$	$\mathrm{E}(\theta_j) = np_j$ $\mathrm{var}(\theta_j) = np_j(1-p_j)$ $\mathrm{cov}(\theta_i, \theta_j) = -np_ip_j$
Negative binomial	$\theta \sim \text{Neg-bin}(\alpha, \beta)$ $p(\theta) = \text{Neg-bin}(\theta\|\alpha, \beta)$	shape $\alpha > 0$ inverse scale $\beta > 0$	$p(\theta) = \binom{\theta+\alpha-1}{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{1}{\beta+1}\right)^\theta$ $\theta = 0, 1, 2, \ldots$	$\mathrm{E}(\theta) = \frac{\alpha}{\beta}$ $\mathrm{var}(\theta) = \frac{\alpha}{\beta^2}(\beta+1)$
Beta-binomial	$\theta \sim \text{Beta-bin}(n, \alpha, \beta)$ $p(\theta) = \text{Beta-bin}(\theta\|n, \alpha, \beta)$	'sample size' n (pos. integer) 'prior sample sizes' $\alpha > 0, \beta > 0$	$p(\theta) = \frac{\Gamma(n+1)}{\Gamma(\theta+1)\Gamma(n-\theta+1)}\frac{\Gamma(\alpha+\theta)\Gamma(n+b-\theta)}{\Gamma(\alpha+b+n)}$ $\times\frac{\Gamma(\alpha+b)}{\Gamma(\alpha)\Gamma(b)}, \quad \theta = 0, 1, 2, \ldots, n$	$\mathrm{E}(\theta) = n\frac{\alpha}{\alpha+\beta}$ $\mathrm{var}(\theta) = n\frac{\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Relationships among common distributions



Solid lines: transformations and special cases

Dashed lines: limits

(Leemis, 1986)

Bayesian Inference

- y: observed data; $y \sim p(y|\theta)$
- θ: parameters (all unobserved quantities)

$$p(\theta \mid y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y \mid \theta)}{p(y)}$$

$$p(\theta \mid y) \propto p(\theta)p(y \mid \theta)$$

posterior distribution

prior distribution

sampling distribution

Multi Parameter Models

$$\mathbf{y} \sim p(\mathbf{y} \mid \theta_1, \theta_2, \ldots, \theta_p)$$

$$p(\theta_1, \theta_2, \ldots, \theta_p \mid \mathbf{y}) \propto p(\theta_1, \theta_2, \ldots, \theta_p) p(\mathbf{y} \mid \theta_1, \theta_2, \ldots, \theta_p)$$

Marginal Posterior Distributions

$$p(\theta_1 \mid \mathbf{y}) \propto \int_{\theta \neq \theta_1} p(\theta_1, \theta_2, \ldots, \theta_p \mid \mathbf{y}) d\theta_{\theta \neq \theta_1}$$

Linear Mixed Models

Data: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$, with $\mathbf{u} \mid \sigma_u^2 \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$

Sampling model: $p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{u}, \sigma_\varepsilon^2) \propto (\sigma_\varepsilon^2)^{-n/2}$

$$\times \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})\right\}$$

Prior distribution: $p(\boldsymbol{\beta}, \sigma_u^2, \sigma_\varepsilon^2) = p(\boldsymbol{\beta})p(\sigma_u^2)p(\sigma_\varepsilon^2)$

(Note: independence was assumed a priori)

Joint posterior distribution:

$$p(\boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \sigma_\varepsilon^2 \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{u}, \sigma_\varepsilon^2)p(\mathbf{u} \mid \sigma_u^2)p(\boldsymbol{\beta}, \sigma_u^2, \sigma_\varepsilon^2)$$

$$\propto (\sigma_\varepsilon^2)^{-n/2}(\sigma_u^2)^{-q/2}p(\boldsymbol{\beta}, \sigma_u^2, \sigma_\varepsilon^2)$$

$$\times \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_\varepsilon^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \frac{1}{\sigma_u^2}\mathbf{u}^T\mathbf{A}^{-1}\mathbf{u}\right]\right\}$$

Marginal Posterior Distributions

Fixed effects vector:

$$p(\boldsymbol{\beta} \mid \mathbf{y}) = \int_{\mathbf{u}} \int_{\sigma_{\mathbf{u}}^2} \int_{\sigma_{\boldsymbol{\varepsilon}}^2} p(\boldsymbol{\beta}, \mathbf{u}, \sigma_{\mathrm{u}}^2, \sigma_{\boldsymbol{\varepsilon}}^2 \mid \mathbf{y}) d\sigma_{\boldsymbol{\varepsilon}}^2 d\sigma_{\mathbf{u}}^2 d\mathbf{u}$$

Note that integrating over a vector (e.g., vector **u**) implies integrating over each element in that vector, i.e.

$$p(\boldsymbol{\beta} \mid \mathbf{y}) = \int_{u_1} \int_{u_2} \cdots \int_{u_q} \int_{\sigma_{\mathbf{u}}^2} \int_{\sigma_{\boldsymbol{\varepsilon}}^2} p(\boldsymbol{\beta}, \mathbf{u}, \sigma_{\mathrm{u}}^2, \sigma_{\boldsymbol{\varepsilon}}^2 \mid \mathbf{y}) d\sigma_{\boldsymbol{\varepsilon}}^2 d\sigma_{\mathbf{u}}^2 du_q \ldots du_2 du_1$$

Single element of **β** (e.g. β_1):

$$p(\beta_1 \mid \mathbf{y}) = \int_{\beta_2} \int_{\beta_3} \cdots \int_{\beta_p} p(\boldsymbol{\beta} \mid \mathbf{y}) d\beta_p \ldots d\beta_3 d\beta_2$$

Marginal Posterior Distributions

Random effects vector:

$$p(\mathbf{u} \mid \mathbf{y}) = \int_{\beta} \int_{\sigma_u^2} \int_{\sigma_\varepsilon^2} p(\beta, \mathbf{u}, \sigma_u^2, \sigma_\varepsilon^2 \mid \mathbf{y}) d\sigma_\varepsilon^2 d\sigma_u^2 d\beta$$

Variance components:

$$p(\sigma_u^2 \mid \mathbf{y}) = \int_{\beta} \int_{\mathbf{u}} \int_{\sigma_\varepsilon^2} p(\beta, \mathbf{u}, \sigma_u^2, \sigma_\varepsilon^2 \mid \mathbf{y}) d\sigma_\varepsilon^2 d\mathbf{u} d\beta$$

$$p(\sigma_\varepsilon^2 \mid \mathbf{y}) = \int_{\beta} \int_{\mathbf{u}} \int_{\sigma_u^2} p(\beta, \mathbf{u}, \sigma_u^2, \sigma_\varepsilon^2 \mid \mathbf{y}) d\sigma_u^2 d\mathbf{u} d\beta$$

Marginal Posterior Distributions

Marginalization (i.e. integrals) in multi-dimensional models can be cumbersome and some times do not have analytical form

An alternative in this regard: Monte Carlo methods

Monte Carlo integration consists of sampling from the posterior distribution, and then using such sampled values to calculate features of interest on the (joint or marginal) posterior distribution

There are many algorithms that can be used to sample from a distribution; some are based on Markov chains, among which the Gibbs sampling is probably the most popular

PAUSE

- $p(\theta \mid y) \propto p(\theta)p(y \mid \theta)$
- Marginal posterior distributions require integrals that quite often are analytically intractable
- But we have MCMC! (next)

Next PAUSE, slide 18

Monte Carlo Methods

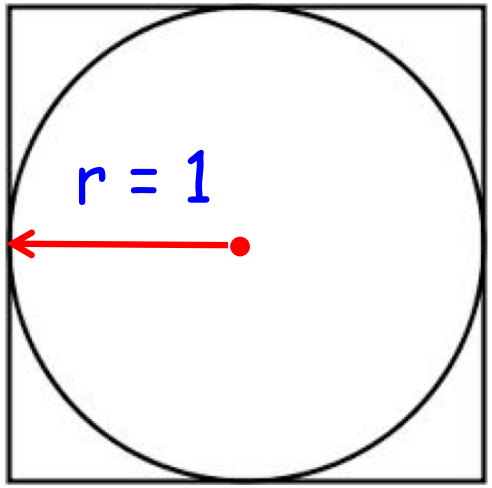
Any method which solves a problem by generating a series of random numbers and counting the incidences that obey specific property(ies)

The method is useful for obtaining numerical solutions to problems which are too complicated to solve analytically

The most common application of the Monte Carlo method is Monte Carlo integration

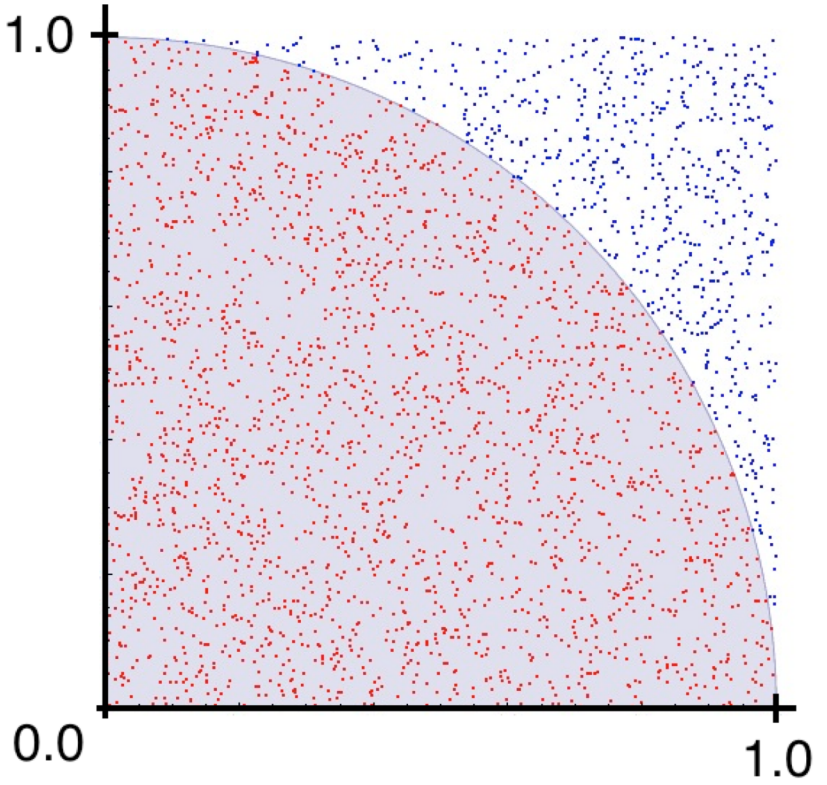
Monte Carlo Methods

Example: approximating the number π using a circle inscribed in a square



Area of circle = πr^2

Area of square = $4 r^2$



$x^2 + y^2 = r^2$

Monte Carlo Methods

Example: approximating the number π using a circle inscribed in a square

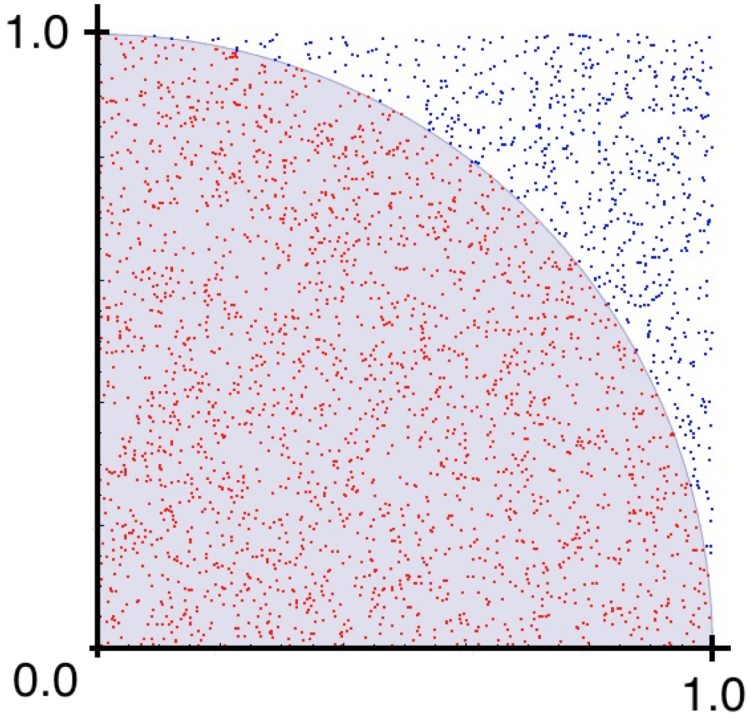
Sample x from Uniform(0,1)

Sample y from Uniform(0,1)

Check if point (x,y) is within the circle, i.e. $y^2 < 1 - x^2$

Repeat the process N times and count how many points (m) fall within the circle

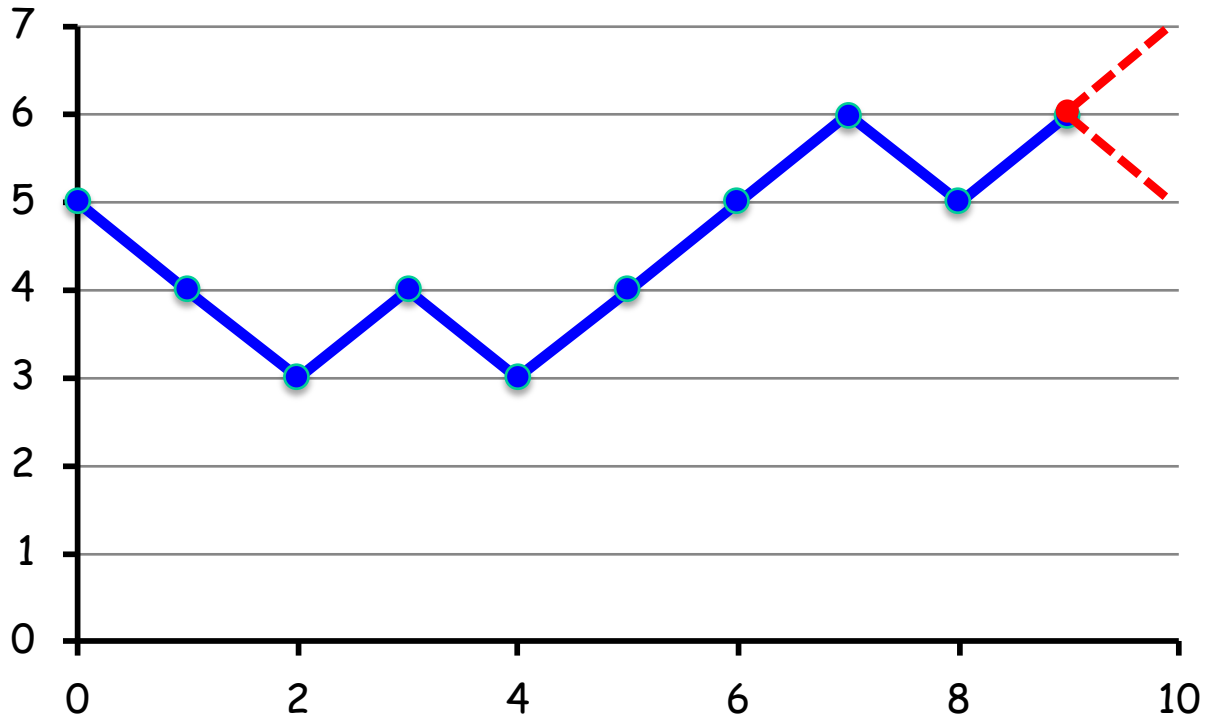
The ratio 4 x m/N is a Monte Carlo approximation for π



$x^2 + y^2 = r^2$

Markov Process

Markov process is a stochastic process that satisfies the Markov property (the memoryless property), i.e., predictions for the future of the process can be made based solely on its present state



Markov Process

Example: Suppose that weather on any given day can be classified into two states: sunny (S) or rainy (R)

Suppose also that, based on past experience, we know that:

Pr(Next day is S | Given today is R) = 0.50 and

Pr(Next day is S | Given today is S) = 0.90

Then, a transition matrix representing the probabilities of the weather moving from one state to another state can be expressed as:

$$P = \begin{array}{c} \begin{matrix} S & R \end{matrix} \\ \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} \end{array} \begin{matrix} S \\ R \end{matrix}$$

Markov Process

If the weather is sunny today (time 0), what is the chance that it will be sunny tomorrow (time 1) as well?

$$Pr(S_1 \mid S_0) = 0.90$$

What about two days from today?

$$Pr(S_2 \mid S_0) = Pr(S_2 \mid S_1) \times Pr(S_1 \mid S_0)$$
$$+ Pr(S_2 \mid R_1) \times Pr(R_1 \mid S_0)$$
$$= 0.9 \times 0.9 + 0.1 \times 0.5 = 0.86$$

Using the same approach to forecast weather on n-th day will approach the following 'equilibrium' probabilities as n increases:

$$Pr(S_n) = 0.833 \text{ and } Pr(R_n) = 0.167$$

PAUSE

- ⇨ Monte Carlo Methods (simulation!)
- ⇨ Markov chain
- ⇨ Next: Gibbs sampling

Next PAUSE, slide 26

Gibbs Sampling

$$\theta = (\theta_1, \theta_2, \ldots, \theta_r) \rightarrow p(\theta_i \mid \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_r)$$

$$\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_r^{(0)})$$

$$\theta_1^{(1)} \mid \theta_2^{(0)}, \theta_3^{(0)}, \ldots, \theta_r^{(0)}$$

$$\theta_2^{(1)} \mid \theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_r^{(0)}$$

$$\vdots$$

$$\theta_r^{(1)} \mid \theta_2^{(1)}, \theta_3^{(1)}, \ldots, \theta_{r-1}^{(1)}$$

Burn-in & Convergence

Tinning interval & Lag correlations

Sample size & Monte Carlo error

Monte Carlo Approximations

After convergence, each sampled vector is a sample from the joint posterior distribution, and so each sampled element (scalar) is a sample from the respective marginal posterior distribution

For each parameter (e.g., θ_1) we'll have then a series of values:

$$\theta_1^{(1)}, \theta_1^{(2)}, \theta_1^{(3)}, \ldots, \theta_1^{(N)}$$

from which features of its distribution (e.g., posterior mean) can be approximated, for example:

$$E[\theta_1 \mid \mathbf{y}] \cong \frac{1}{N} \sum_{j=1}^{N} \theta_1^{(j)}$$

Monte Carlo Approximations

Other often interesting features used to represent a marginal posterior distribution are: posterior variance (or standard deviation), posterior mode or median, percentiles, highest posterior density (HPD), etc.

Very useful property: If one is interested on the distribution of a function of the model parameters, samples from such a distribution can be obtained simply by applying that specific function to the sampled values of those parameters

For example, the posterior mean of the heritability can be obtained as:

$$E[h^2 \mid \mathbf{y}] \cong \frac{1}{N}\sum_{j=1}^{N} \frac{\sigma_u^{2(j)}}{\sigma_u^{2(j)} + \sigma_\varepsilon^{2(j)}}$$

Example: Linear Model

Data: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad \mathbf{e} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}\sigma_e^2)$

Sampling model: $p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma_e^2) \propto (\sigma_e^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$

Prior distribution: $p(\boldsymbol{\beta}, \sigma_e^2) = p(\boldsymbol{\beta})p(\sigma_e^2) \propto (\sigma_e^2)^{-1}$

Joint posterior distribution:

$$p(\boldsymbol{\beta}, \sigma_e^2 \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma_e^2)p(\boldsymbol{\beta}, \sigma_e^2)$$

$$\propto (\sigma_e^2)^{-(n+2)/2} \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

Example: Linear Model

Conditional distribution of location parameters:

$$p(\boldsymbol{\beta} \mid \sigma_e^2, \mathbf{y}) \propto \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}$$

Recall $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ and note that $\mathbf{y}-\mathbf{X}\boldsymbol{\beta} = \mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{X}\hat{\boldsymbol{\beta}}-\mathbf{X}\boldsymbol{\beta}$

such that:

$$(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \left[(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})-(\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})\right]^T\left[(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})-(\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})\right]$$

$$= \Big[\underbrace{(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})}_{\text{independent of } \boldsymbol{\beta}} - \underbrace{2(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})}_{\text{equal to zero}} + (\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}})\Big]$$

Example: Linear Model

Conditional distribution of location parameters:

Hence: $$p(\boldsymbol{\beta} \mid \sigma_e^2, \mathbf{y}) \propto \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{X}\boldsymbol{\beta} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{X}\boldsymbol{\beta} - \mathbf{X}\hat{\boldsymbol{\beta}})\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma_e^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T\mathbf{X}^T\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right\}$$

and so: $$\boldsymbol{\beta} \mid \sigma_e^2, \mathbf{y} \sim N\left(\hat{\boldsymbol{\beta}},\ (\mathbf{X}^T\mathbf{X})^{-1}\sigma_e^2\right)$$

where $$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Example: Linear Model

Conditional distribution of residual variance:

$$p(\sigma_e^2 \mid \boldsymbol{\beta}, \mathbf{y}) \propto (\sigma_e^2)^{-(n+2)/2} \exp\left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$

Hence: $\sigma_e^2 \mid \boldsymbol{\beta}, \mathbf{y} \sim \text{Inv-gamma}\left(\frac{n}{2}, \ \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right)$

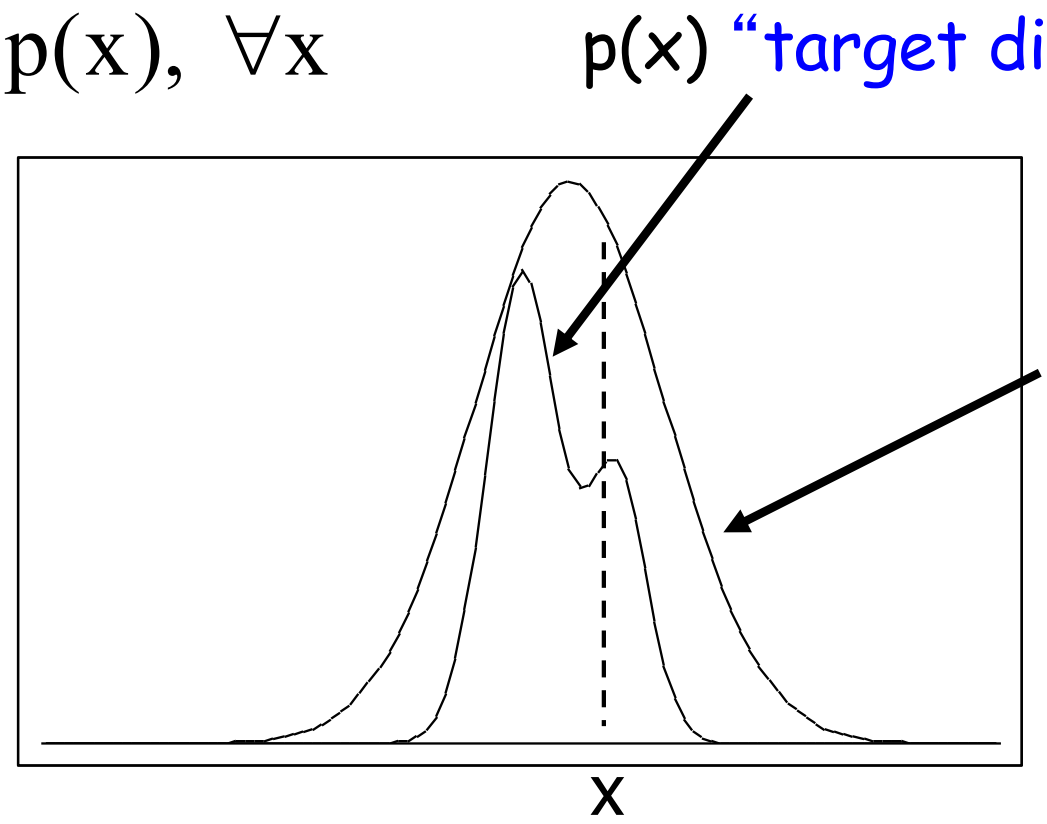
PAUSE

- ⇨ Gibbs sampling requires all conditional posterior distributions
- ⇨ What if not all are know distribution, with closed form?

Next PAUSE, slide 28 (end)

Rejection Sampling

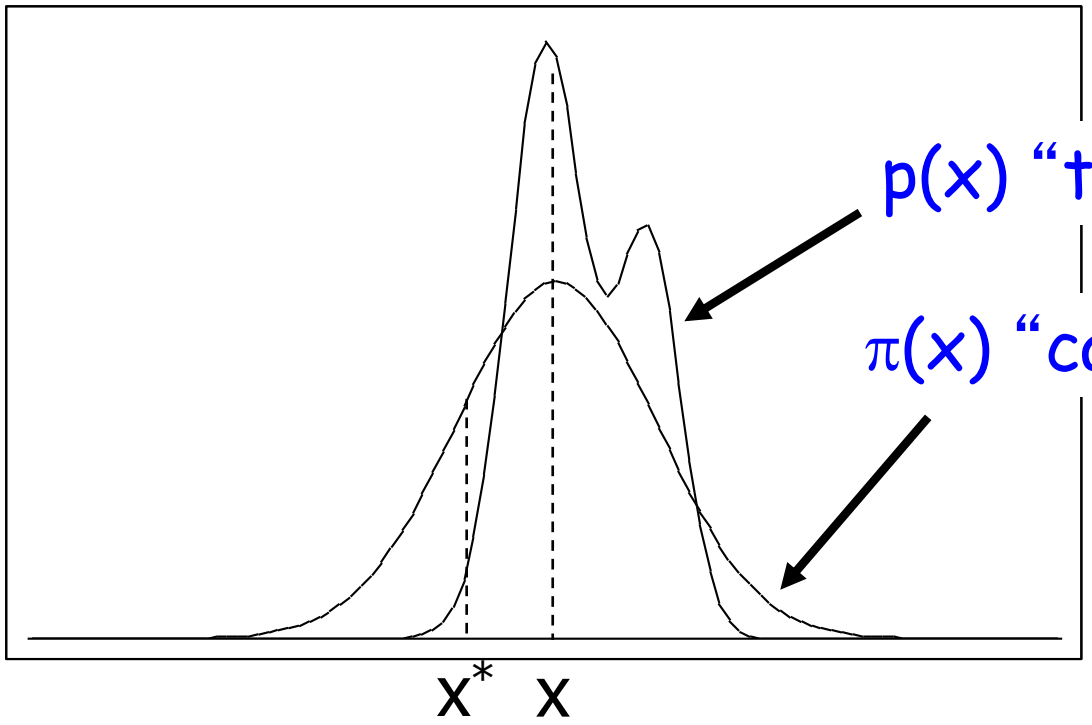
$$K\,f(x) \geq p(x), \quad \forall x$$



① Sample x from f(x)

② Decision: Probability of accepting x: $\alpha = \dfrac{p(x)}{Kf(x)}$

Metropolis-Hastings Algorithm



① x: current value; sample x^* from $\pi(x)$, e.g. $\pi(x) \sim N(x, \tau^2)$

② The chain moves from x to x^* with probability:

$$\alpha = \min\left[1, \frac{p(x^*)\pi(x)}{p(x)\pi(x^*)}\right]$$

Otherwise the chain remains at the current value