Sampling and Markov Chain Monte Carlo

Lec 12. Sampling methods and Markov Chain Monte Carlo

1. Monte Carlo estimates of expectations 11.1.1 11.2.1 11.1.2 - 11.1.5
   rejection
   importance
2. Directed graphical models & ancestral sampling
3. MCMC Intro, Stationary Distributions, Ergodicity
4. Sampling via Inverting the CDF
5. Transformations of Sampling Variables.

Sampling

1. Inverting the PDF
2. Transformations of sampling variables
3. MCMC
4. Gibbs
5. Importance / Rejection

Monte Carlo estimates of expectations

why samples ?

- Represent distributions that have no simple analytical form
- Easily estimate expectations

Task: You have model for RV z given by $p(z)$

You want to know the expected value of $f(z)$

$$\boxed{\text{Ideal: } \bar{f} = \mathbb{E}_{z\sim p(z)} f(z) = \int p(z) f(z)\, dz}$$

① integral hard
② $p(z)$ hard to know

Monte Carlo Estimate:

$$\boxed{\hat{f} = \frac{1}{S} \sum_{i=1}^{S} f(z^s) \text{, where } z \overset{iid}{\sim} p(z)}$$

- Question:
  - What is the expected value of MC estimate?

$$
\begin{aligned}
E(\hat{f}) &= E\Big(\frac{1}{S}\sum_{s=1}^{S} f(z^s)\Big) \\
&= \frac{1}{S}\sum_{s=1}^{S} E(f(z^s)) \\
&= \frac{1}{S}\sum_{s=1}^{S} \mathbb{E}_{z^s\sim p(z)} (f(z^s)) \\
&= \frac{1}{S}\sum_{s=1}^{S} \int p(z)\cdot f(z)\, dz \\
&= \frac{1}{S}\sum_{s=1}^{S} \bar{f} \\
&= \bar{f}
\end{aligned}
$$

$\Rightarrow$ MC estimate is Unbiased !

- What is the Variance of MC estimate?

Conclusion: $\mathrm{Var}[\hat{f}] = \frac{1}{S}\mathrm{Var}[f(z)]$

$\Rightarrow$ So for the large S. MC will be "close" to the ideal $\bar{f}$!

How to sample from models with Multiple Variable?

Directed graphical model, ancestral sampling. joint samples → marginal, conditional

- Suppose we have model of many r. v., we wish to sample from joint distribution $p(z_1, z_2, \cdots, z_T)$ e.g. draw from Markov models.

- Suppose further we know conditional independence assumption, and can use these to define a graph (must be directed & no cycle). vertices/nodes are random variables. Directed edges represents conditional dependence assumption.

- $z_1 \to z_2 \to z_3 \to z_4$

$$p(z_{1:4}) = \prod_{i \in V} p(z_i \mid z_{pa(i,E)})$$

V: vertex list
E: edge list
$pa(i,E)$: "parents" of node i:
$\{j : (j,i) \in E\}$

Ex: - For the 1^{st} order Markov model

$z_1 \to z_2 \to z_3 \to z_4$

$E = \{(1,2), (2,3), (3,4)\}$

$$p(z_{1:4}) = p(z_1) \cdot p(z_2 \mid z_1) \cdot p(z_3 \mid z_2) \cdot p(z_4 \mid z_3)$$

- For the 2^{nd} order Markov

$$\begin{tikzcd} z_1 \arrow[r] \arrow[rr, bend right] & z_2 \arrow[r] \arrow[rr, bend right] & z_3 \arrow[r] & z_4 \end{tikzcd}$$

$E = \{(1,2), (1,3), (2,3), (2,4), (3,4)\}$

$$p(z_{1:4}) = p(z_1) \cdot p(z_2 \mid z_1) \cdot p(z_3 \mid z_2, z_1) \cdot p(z_4 \mid z_3, z_2)$$

- We call this represents a $\boxed{\text{directed graphical model}}$.
- useful for sampling and computing the
  - margins $p(z_t)$
  - joints $p(z_1, z_2)$
  - conditions $p(z_2 \mid z_1)$

- $\boxed{\text{Question}}$: How to sample from directed graphical model?
  - Assume we have way to sample from $p(z_i \mid z_{pa(i,E)})$, then node indices in order $1, 2, \cdots, T$ s.t. for any index $j \in \{1, 2, \cdots, T\}$ if i is a parent of j, then $i < j$

- Ancestral Sampling:

  for i in $1, 2, \cdots, T$:

  $\quad z_i \sim p(z_i \mid z_{pa(i,E)})$

  return $[z_1, z_2, \cdots, z_T]$

  Guaranteed to sample from joint $p(z_{1:T})$

- $\boxed{\text{Question}}$: What can we do with samples from a joint distribution?
  - Compute a MC expectation
  - Sample from a marginal

    Given S samples $z_{1:T}^{(1)}, \cdots, z_{1:T}^{(S)}$, we can get samples of z_t by just keeping $z_t^{(1)}, z_t^{(2)}, \cdots, z_t^{(S)}$
  - Sample from a conditional

    Given S samples $z_{1:T}^{(1)}, \cdots, z_{1:T}^{(S)}$, we can get samples from $p(z_t \mid z_u = k)$ by keeping $\{z_t^{(S)} = z_u^{(S)} = k\}$ and discarding others

Markov Chain Monte Carlo (MCMC)

- $\boxed{\text{Goal: Want sample } z \text{ from a target distribution } p^*(z)}$
- But we may only know the pdf up to a constant (doesn't depend on z)

$$p^*(z) = \underset{\uparrow \atop \text{unknown}}{C} \cdot \underset{\uparrow \atop \text{known function, easy to evaluate}}{\tilde{p}(z)}$$

$$\log p^*(z) = \log C + \log \tilde{p}(z)$$

Ex: $\text{posterior} = \dfrac{\text{likelihood} \times \text{prior}}{\text{evidence}} = C \cdot \tilde{p}(z)$

where likelihood $\times$ prior is easy to know, but evidence is a hard integral

- Markov Chain Monte Carlo

Markov Chain ↓ Draw samples sequentially instead of iid

Monte Carlo ↓ Want to draw samples from target dist $p^*(z)$

ex: previous setup:

$$x^1 \sim p^*(x)$$
$$x^2 \sim p^*(x)$$
$$\vdots$$
$$x^n \sim p^*(x)$$

Now:

$$x^1 \sim T(x \mid x_0)$$
$$x^2 \sim T(x \mid x_1)$$
$$\vdots$$
$$x^n \sim T(x \mid x_{n-1})$$

$$z_1 \xrightarrow{T} z_2 \xrightarrow{T} \cdots, z_s$$

$T \rightarrow$ proposed distribution

If we are careful about choosing proposed dist T, then margin $\sum p(z_s) = p^*(z)$

- Goal: We want $p^*(z)$ to be the stationary distribution of the Markov Chain
  (Choose T so the stationary distribution of the chain is $p^*(z)$)
- Stationary distribution:

For discrete:

$$\underbrace{p(z_{t+1}=k)}_{\text{next time step}} = \sum_{j=1}^{k} \underbrace{p(z_t=j)}_{\text{p. of being in state } j \text{ at time } t} \cdot \underbrace{T(z_{t+1}=k \mid z_t=j)}_{\text{p. of transition from state } j \text{ to } k} \qquad z \in \{1, 2, \cdots, k\}$$

(Current timestep)

For continuous:

$$p(z_{t+1}=k) = \int_{z_t \in \Omega} p(z_t)\, T(z_{t+1} \mid z_t)\, dz_t$$

- $\boxed{\text{Question}}$ When does a unique stationary distribution exist?
- Answer: When the markov chain is $\boxed{\text{ergodic}}$, which means if we start in state i at $t=0$, then for some time $T_0 > 0$, we have for every state k

$$p(z_T = k \mid z_0 = i) > 0$$

for all $T \geq T_0$

- Intuition: Need to be able to get from any state to any other state.

- key technical conditions for ergodicity:

① T must be irreducible : path exists between any two states

② T must be aperiodic : no cycles

Sampling via Inverting the CDF

- Consider real valued random var X with pdf : $p(x)$ and cdf $C(X)$
- If cdf function F is invertible analytically, we can sample using the simple transformation

   ① $U \sim \text{Unif}([0,1])$

   ② $X \leftarrow F^{-1}(U)$

Sampling via transformations

Lec 13 Random Walk proposals and Metropolis-Hastings Algorithm

1. Markov Transitions that Propose then Accept / Reject
2. Random walk proposals
3. Detailed Balanced, Proof of Random Walk's Stationary Distribution
4. Metropolis and Metropolis-Hastings algorithms.

▲对于任意给定的目标分布 $\pi(x)$，我们如何找到它为唯一平稳分布的 Markov chain，并且基于 Monte Carlo 采样的方法，实现对其进行采样

Markov Transitions that Propose then Accept / Reject

Review

- Goal: Sample from target distribution with pdf $\boxed{p^*(z)}$ over real $z \in \Omega \subseteq \mathbb{R}$

- Challenge: Do not know how to evaluate $p^*(z)$. only known up to multiplicative constant

  $\tilde{p}(z)$ is evalateable

  $p^*(z) = c\,\tilde{p}(z)$ is NOT (c is unknown)

- Idea: ratio of PDFs of two possible z values can be evaluated exactly

$$\frac{p^*(z^1)}{p^*(z^2)} = \frac{c\,\tilde{p}(z^1)}{c\,\tilde{p}(z^2)} = \frac{\tilde{p}(z^1)}{\tilde{p}(z^2)}$$

- Therefore. We want to design Markov proposal that some how uses ratios

Markov Transition Design: Propose then accept / reject.

- Transition distribution T takes current "state" value z_t, and produces new "state" value z_{t+1}

- Idea: Propose new value $z' \sim Q(z' \mid z_t)$, then decide to accept with Prob $A(z', z_t)$

$$z_{t+1} = \begin{cases} z' & \text{if } u < \underbrace{A(z', z_t)}_{\text{是个概率}} \\ z_t & \text{otherwise} \end{cases}$$

where $u \sim \text{Uniform}(0,1)$

- Example:

Start: $z_1 = 0.3$

1) Sample $z' \sim \underline{Q(\cdot \mid z_1)}$ $\quad z' = -1.8$
2) Sample $u \sim \text{Unif}(0,1)$ $\quad u = 0.3$
3) Evaluate $\underline{A(z', z_t)} = 0.8$. $\quad u < A$. so ACCEPT

$z_2 = -1.8$

1) Sample $z' \sim Q(\cdot \mid z_2)$ $\quad z' = 2.7$
2) Sample $u \sim \text{Unif}(0,1)$. $\quad u = 0.91$
3) Eval $A(z', z_t) = 0.01$ $\quad u > A$. so REJECT

$z_3 = -1.8$

⋮ keep going to get $z_3 \ldots, z_S$

- A is an accept probability threshold

We know $0 \le A \le +\infty$

$A = 0.0$ means no chance of accept ALWAYS REJECT

$A \in (0, 0.5)$ means can accept, but likely reject

$A \in (0.5, 1]$ means can reject but likely accept

$A \ge 1$ means ALWAYS ACCEPT

So sometimes we write

$$p(\text{accept} \mid z', z_t) = \min(A(z', z_t), 1)$$

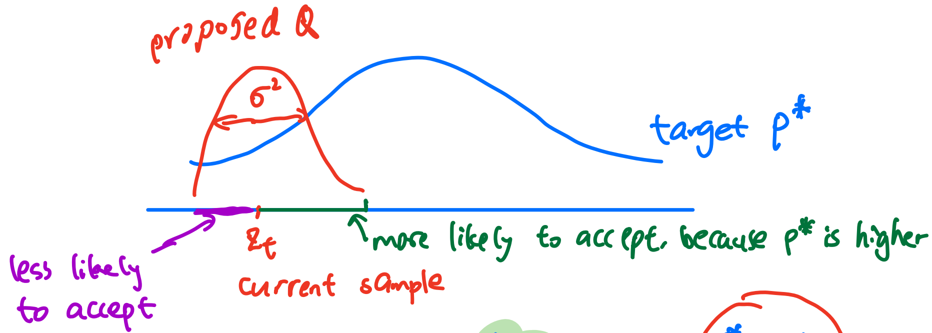
any value ≥ 1 is just set to 1

Random walk proposals: Intuition

- Question: How to choose $Q(\cdot \mid z_t)$?
- Random walk: $Q(z' \mid z_t) = \text{Normal}(z' \mid z_t, \sigma^2)$ hyperparameter

Reason is: easy to sample; likely to propose "nearby"

- Intuition:



Let's try $A(z', z_t) = \dfrac{\tilde{p}(z')}{\tilde{p}(z_t)} = \dfrac{p^*(z')}{p^*(z_t)}$ → Can be computed exactly !

- Given current z_t, and proposed value z'

$p^*(z')$ ___?___ than $p^*(z_t)$	ratio $\dfrac{p^*(z')}{p^*(z_t)}$	Proposal will be
much smaller	$<< 0.5$	likely rejected
a bit smaller	$(0.5, 1)$	likely accepted
same or larger	≥ 1	certainly accepted

Detailed Balanced, Proof of Random Walk's Stationary Distribution

- Intuitively, random walk proposed Q plus p^* ratio A makes sense. but how to show this will create Markov chain with stationary distribution p^* ?

- Proof strategy
  1) Show T satisfies detailed balance
  2) Show that any T satisfies detailed balance will have p^* stationary distribution.

- Detailed Balance $z_a \leftrightarrows z_b$ "reversible"
  z is random variable with random space Ω z_a, z_b two possible values.

$$\boxed{p^*(z_a) T(z_b | z_a) = p^*(z_b) \cdot T(z_a | z_b)}$$

- Q: Can we show Markov chain with transition T that satisfies detailed balance for p^* has p^* as its stationary ?

To show DB for p^* has p^* as its stationary

Def of stationary : $p^*(z_{t+1}) = \int p^*(z_t) \cdot T(z_{t+1} | z_t)\, dz_t$ ①

Def of DB : $p^*(z_a) \cdot T(z_b | z_a) = p^*(z_b) \cdot T(z_a | z_b)$ ②

由②，①可改写成：

$$p^*(z_{t+1}) = \int p^*(z_{t+1}) \cdot T(z_t | z_{t+1})\, dz_t$$

$$= p^*(z_{t+1}) \underbrace{\int T(z_t | z_{t+1})\, dz_t}_{=1}$$

$$= p^*(z_{t+1})$$ T is a pdf and must integrate to 1

Thus, p^* is stationary distribution because it meets required definition.

To show Markov chain with transition T that satisfies DB

- pdf for T when $z_a \neq z_b$

$$T(z_b | z_a) = P(\text{accept} | z_b, z_a) \cdot Q(z_b | z_a)$$

$$= \min\left(1, \frac{p^*(z_b)}{p^*(z_a)}\right) N(z_b | z_a, \sigma^2)$$

Def of DB:

$$p^*(z_a) \cdot T(z_b | z_a) = p^*(z_b) \cdot T(z_a | z_b)$$ 同样可以化成左边形式

$$\Rightarrow p^*(z_a) \cdot \min\left(1, \frac{p^*(z_b)}{p^*(z_a)}\right) \cdot N(z_b | z_a, \sigma^2) = p^*(z_b) \cdot T(z_a | z_b)$$

$$\Rightarrow \min(\cancel{p^*(z_a), p^*}(z_b)) \cdot N(z_b | z_a, \sigma^2)$$

$$= \min(\cancel{p^*(z_b), p^*}(z_a)) \cdot N(z_b | z_b, \sigma^2)$$

$$\Rightarrow N(z_b | z_a, \sigma^2) = N(z_b | z_b, \sigma^2)$$ True.

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} \cdot \underline{(z_b - z_a)^2}}$$ symmetric

- $z_a = z_b$, obviously.

Metropolis and Metropolis-Hastings algorithms.

METROPOLIS MCMC Algorithm

Assumes that Q valid PDF over Ω
easy to sample
easy to evaluate PDF
must be symmetric

$$Q(z_a|z_b) = Q(z_b|z_a)$$

for all $z_a \in \Omega$, $z_b \in \Omega$

$\tilde{p}$

Initialize $z_1 \in \Omega$
for t in $1, 2, \cdots S-1$
1) $z' \sim Q(\cdot|z_t)$
2) $u \sim \mathrm{Unif}([0,1])$
3) $z_{t+1} = \begin{cases} z' & \text{if } u \le \dfrac{\tilde{p}(z')}{\tilde{p}(z_t)} = A(z', z_t) \\ z_t & \text{otherwise} \end{cases}$

return $[z_1, z_2, \cdots z_S]$

if S large enough, can consider $z_{S-B}, \cdots z_S$ as samples of p^*

METROPOLIS-HASTINGS MCMC Algorithm

Assumes only that
Q is valid PDF over Ω
easy to sample
& easy to evaluate PDF

Q don't need to be symmetric!

Initialize $z_1 \in \Omega$
for t in $1, 2, \cdots S-1$
1) $z' \sim Q(\cdot|z_t)$
2) $u \sim \mathrm{Unif}([0,1])$
3) $z_{t+1} = \begin{cases} z' & \text{if } u < \dfrac{\tilde{p}(z')}{\tilde{p}(z_t)} \dfrac{Q(z_t|z')}{Q(z'|z_t)} \\ z_t & \text{otherwise} \end{cases}$

can show this A also satisfies DETAILED BALANCE

return $[z_1, z_2, \cdots z_S]$

Lec 14 Gibbs sampling 11.3

1. Overview of sampling Vector Random Variable
2. Gibbs Sampling Algorithm
3. Proof sketch of Gibbs Sampling Correctness
4. Pro/con Comparision v.s. random walk

Overview of sampling Vector Random Variable

- Goal now: Sample a vector-valued random variable Z from a target distribution p^*

  $Z = [z_1, z_2, \cdots, z_D]$

- Q: How to draw our samples?

  $[z_1^{(s)}, z_2^{(s)}, \cdots, z_D^{(s)}] \sim p^*$

  1) Transform simplier
  2) Random Walk (Metropolis MCMC)
  3) Gibbs sampling

- Try to use Metropolis-Hastings for vector-valued random variables. There're two options.

Option 1: Propose joint sample, accept or reject entire vector

1次更新一个时间段的所有entry.

for t in $2, 3, \cdots S$:

$\quad z_1', \cdots, z_D' \sim Q(\cdot \mid z^{(t-1)}, \cdots z_D^{(t-1)})$

$\quad u \sim \text{Unif}([0,1])$

$$z_1^{(t)} \cdots z_D^{(t)} = \begin{cases} z_1', \cdots, z_D' & \text{if } u < A(z', z^{(t-1)}) \\ z^{(t-1)}, \cdots, z_D^{(t-1)} & \end{cases}$$

Option 2: Separate proposed for each dimension

1次更新一个时间段的一个entry (维度)

for t in $2, 3, \cdots S$:

$\quad$ for d in $1, 2 \cdots, D$:

$$z_d' \sim Q_d(\cdot \mid \underbrace{z_1^{(t)} \cdots z_{d-1}^{(t)}}_{\text{new value for dim} < d}\ z_d^{(t-1)} \underbrace{\cdots z_D^{(t-1)}}_{\text{old value for dim} > d})$$

$\quad\quad u \sim \text{Unif}([0,1])$

$$z_d^{(t)} = \begin{cases} z'_d & \text{if } u < A\left(\dfrac{z_1^{(t)} \cdots z_{d-1}^{(t)}, z'_d, \cdots z_D^{(t-1)},}{z_1^{(t)} \cdots z_{d-1}^{(t)}, z_d^{(t-1)}, \cdots z_D^{(t-1)}} \right) \\ z_d^{(t-1)} & \end{cases}$$

- Option 1: can make big changes, but could have high rejections
  2: may accept more often, but slow to make big changes.

- Gibbs Sampling Algorithm

- Goal: Sample $z_1^{(s)}, \ldots, z_D^{(s)} \sim p^*$
- Challenge: Cannot sample using analytical methods.
- Insight: Sample from joint are hard, but samples from conditional might be easier

一维一维地采样

So, set up markov chain where:

$$z^{(t)} \xrightarrow[T_{gibbs}]{} z^{(t+1)}$$

Transition T_{Gibbs} has D steps: p* 可以考虑成 MH 中的 Q 函数

1) $z_1^{(t+1)} \sim p^*(z_1 \mid z_2^t, \cdots z_D^t)$
2) $z_2^{(t+1)} \sim p^*(z_2 \mid z_1^{t+1}, z_3^{t+1}, \cdots z_D^t)$
3) $z_3^{(t+1)} \sim p^*(z_3 \mid z_1^{t+1}, z_2^{t+1}, z_4^t, \cdots, z_D^t)$

$\vdots$

z_D^{t+1}

Then we can get the vector z for $t+1$ moment.

- If we repeatly sample using T_{Gibbs}, can show that p^* is stationary distribution

- Example:
  How to sample from $p^*(z_1, z_2) = N\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \middle| \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 \\ 0.8 & 2 \end{bmatrix}\right)$

Recall formula for conditionals given joint Gaussian:

$$z \sim N(\mu, \Sigma), \text{ where } \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

then $z_1 \mid z_2 \sim N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(z_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right) = N(0.4z_2, 0.68)$

$z_2 \mid z_1 \sim N\left(\mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(z_1 - \mu_2), \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}\right) = N(0.8, 1.36)$

Therefore, Gibbs sampler would be:

Initialize $z^{(1)} = [0, 0]$

for t in $2, 3, \cdots S$:

$z_1^{(t)} \sim$ Sample_$\underline{\text{1D_NORM}}$ $(0.422, 0.68)$

$z_2^{(t)} \sim$ Sample_$\underline{\text{1D_NORM}}$ $(0.8, 1.36)$

Return $[z^{(1)}, z^{(2)}, \cdots, z^{(S)}]$

Proof sketch of Gibbs Sampling correctness

- 为什么通过这种采样方法采得的样本能服从目标分布 $p(z)$? & why Gibbs is a MCMC chain with stationary p^*?
- $\boxed{\text{Prove 1: Gibbs is a special case of Metropolis-Hastings.}}$
- Consider using the Gibbs transition distribution as the proposal distribution. Q of Metro-hastings we'll focus <u>not</u> on all D steps of T_{Gibbs}, but just one.

Consider only the <u>first dim's</u> conditional transition:

old z_1^t z_2^t

proposed: z_1' z_2^t

sampled copied

Main: proposal only sample $d=1$, other dim is copied!

$$z_1' \sim Q_1(z_1 \mid z_{1\to 2}^{(t)}) : p^*(z_1 \mid z_2^{(t)})$$

then accept threshold is

$$A(z', z^{(t)}) = \frac{p^*(z')}{p^*(z^t)} \frac{Q(z^{(t)} \mid z')}{Q(z' \mid z^{(t)})} \leftarrow \text{MH}$$

$$= \frac{p^*(\overset{c}{z_2'}) \cdot \cancel{p^*(z_1' \mid \overset{c}{z_2'})}}{p^*(\underset{c}{z_2^t}) \cdot \cancel{p^*(z_1^t \mid \underset{c}{z_2^t})}} \frac{\cancel{p^*(z_1^{(t)} \mid \overset{c}{z_2})}}{\cancel{p^*(z_1' \mid \underset{c}{z_2^{(t)}})}} \qquad z_2' = z_2^{(t)} = c$$

$$= 1$$

$\boxed{\text{Punchline: We would always accept a proposal from the Gibbs Conditional}}$

- prove 2: T_{Gibbs} has stationary distribution p^*
- strategy: Show that T_{Gibbs} satisfies Detailed balance, wrt p^*. This implies p^* is the stationary.

T has Stationary p^* $\underset{\not\leftarrow}{\rightarrow}$ T satisfy detailed balance

- Let fix $D=2$. Focus on transition only for dim $d=1$

Detailed balance condition says:

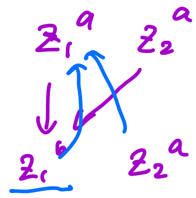
for all z_1^a, z_2^a, z_1^b:

original detail balance:

$$\boxed{p^*(z_a)\,T(z_b|z_a) = p^*(z_b)\cdot T(z_a|z_b)}$$

a 开始, b 结束 = b 开始, a 结束

$$p^*(z_1^a, z_2^a)\,T(z_1^b|z_2^a) = p^*(z_1^b, z_2^a)\,T(z_1^a|z_2^a)$$



$$\Rightarrow \cancel{p^*(z_2^a)}\cdot p^*(z_1^a|z_2^a)\,T(z_1^b|z_2^a)$$
$$= \cancel{p^*(z_2^a)}\cdot p^*(z_1^b|z_2^a)\cdot T(z_1^a|z_2^a)$$

$$\Rightarrow p^*(z_1^a|z_2^a)\,T(z_1^b|z_2^a) = p^*(z_1^b|z_2^a)\,T(z_1^a|z_2^a)$$

Proved!

Thus, DB Condition is true for T_{Gibbs}, and p^* is the markov Chain's stationary distribution

Pro/con Comparision v.s. random walk

	practical effectiveness (sample $z^{(s)}$ is really from p^*)	required properties of p^*	ease of derivation for a new model	hyperparams
(1) Transformation Methods	EXCELLENT	Can prove $T(u) \sim p^*$ HARD	CHALLENGING to IMPOSSIBLE	none
2) Random Walk MCMC	TERRIBLE in worst case, OK in best case	$\tilde{p}$ evaluatable EASY	EASY	std dev σ
(3) Gibbs MCMC	FAIR-to-GOOD	Can sample from conditionals $p^*(z_d \mid z_1 \cdots z_{d-1}, z_{d+1} \cdots z_D)$ MODERATE	DOABLE with EFFORT for most directed graphical models	none

Punchline: For most "interesting" models with many variables (large $D \gg 1$)
random walk is unlikely to be useful
b/c will take way too long to converge
even though convergence guaranteed eventually

$$\boxed{\text{Using Gibbs sampling is one of our best options for effective sampling}}$$

现在需要对$p^*(z)$用MCMC进行近似采样.
选择一个合适的T使Markov chain的平稳分布
为$p^*(z)$. 如何选择T? 这个T要满足
Detailed balance 即可.

Detailed balance: (进a出b = 进b出a

1维:

$$p^*(z_a)\cdot T(z_b \mid z_a) = p^*(z_b)\cdot T(z_a \mid z_b)$$

2维: z_1^a, z_2^a, z_1^b

$$p^*(z_1^a, z_2^a)\, T(z_1^b \mid z_1^a) = p^*(z_1^b, z_1^a)\cdot T(z_1^a \mid z_1^a)$$