Configuration and site update.

Let V be a finite set of "sites" (or "indices") and let Λ be a set of "common states" which can be discrete or continuous. Introduce the **space of configurations** by

$$S := \{(x_v) : (x_v) \in \Lambda^V\}$$

A random sample X is the collection $(X_v : v \in V)$ of Λ-valued random variables X_v's, and the target distribution π becomes a joint probability distribution of X_v's. In discrete cases the state x_v at the site v can be updated according to the conditional probability

$$\pi(x_v \,|\, (x_w)_{w \neq v}) = \frac{\pi(x)}{\sum_{\lambda \in \Lambda} \pi((x; x_v \leftarrow \lambda))}$$

given the configuration $x = (x_w)_{w \neq v}$ restricted on $V \setminus \{v\}$. Here $(x; x_v \leftarrow \lambda)$ denotes the configuration replacing the state x_v by λ.

Gibbs sampler.

The Gibbs sampler is viewed as a special case of Metropolis-Hastings algorithm when one can take advantage of univariate proposal transition probability $\pi(x_v \mid (x_w)_{w \neq v})$ for one site v at time. Each move (site update) is always accepted in the algorithm.

Gibbs Algorithm (Gibbs Sampler)

1. Suppose that $X_t = x$ at time t.
2. Pick a site $v \in V$, and generate λ from the conditional distribution $\pi(x_v \mid (x_w)_{w \neq v})$.
3. Set $X_{t+1} = (x; x_v \leftarrow \lambda)$.

The choice of site v to update can be in a certain order (i.e., systematic scan) or random. While one whole cycle of sites is exhaustively updated for systematic scan, only a single random update is required in order for the Gibbs sampler to be ergodic.

R code: Gibbs algorithm with systematic scan.

Here we consider a Gibbs distribution $\pi(x)$ of the Potts model (which is introduced later in the topic) with $x = (x_1, \ldots, x_6)$ each taking a value on $\Lambda = \{1, \ldots, 10\}$ (`m=10`). Then we can simulate a Gibbs sampler for ten cycles of systematic scan (`tmax=10`).

```
source("potts.r")
potts(m=10,n=c(1,6),tmax=10,ask.break=T)
```

In general the Potts model is designed for a two-dimensional lattice V. For example, we can choose $V = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$ (`n=c(6,6)`). Then a configuration becomes

$$x = (x_{(1,1)}, x_{(1,2)}, \ldots, x_{(6,5)}, x_{(6,6)})$$

and it can be updated in the following demonstration.

```
potts(m=10,n=c(6,6),tmax=10,ask.break=T)
```

Transition probability of Gibbs sampler.

Suppose that Λ consists of finite common states. For example, we choose $\Lambda = \{-1, +1\}$ (as in Ising models introduced later). Then we can construct the transition probability at each site $v \in V$ by

$$\mathbf{P}_v(x, y) := \begin{cases} \dfrac{\pi(y)}{\sum_{\lambda \in \Lambda} \pi((x; x_v \leftarrow \lambda))} & \text{if } (x_w)_{w \neq v} = (y_w)_{w \neq v}; \\ 0 & \text{otherwise.} \end{cases}$$

A single transition probability $\mathbf{P}_v$ is not irreducible. We can consider a fixed ordering $v_1, \ldots, v_n$ of V, and produce a **systematic site update** by

$$\mathbf{P} = \mathbf{P}_{v_1} \cdots \mathbf{P}_{v_n}.$$

Alternatively we can introduce a distribution $\rho(v)$ over V, and define a **random site update** by

$$\mathbf{P} = \sum_{v \in V} \rho(v) \mathbf{P}_v.$$

In either of the update schemes $\mathbf{P}$ becomes irreducible. Aperiodicity is guaranteed since it is usually the case that $\mathbf{P}_v(x, x) > 0$.

Detailed balance of Gibbs sampler.

The single transition probability $\mathbf{P}_v$ satisfies the detailed balance

$$\begin{aligned}\pi(x)\mathbf{P}_v(x,y) &= \pi(x)\frac{\pi(y)}{\sum_{\lambda\in\Lambda}\pi((x;x_v\leftarrow\lambda))}\\ &= \pi(y)\frac{\pi(x)}{\sum_{\lambda\in\Lambda}\pi((y;y_v\leftarrow\lambda))} = \pi(y)\mathbf{P}_v(y,x)\end{aligned}$$

if $(x_w)_{w\neq v} = (y_w)_{w\neq v}$. It is reversible with stationary distribution π, but it is not ergodic. The systematic site update satisfies the detailed balance

$$\begin{aligned}\pi(x)\mathbf{P}(x,y) &= \pi(x)(\mathbf{P}_{v_1}\cdots\mathbf{P}_{v_n})(x,y)\\ &= \pi(y)(\mathbf{P}_{v_n}\cdots\mathbf{P}_{v_1})(y,x) = \pi(y)\tilde{\mathbf{P}}(y,x)\end{aligned}$$

and it has an ergodic transition probability $\mathbf{P}$ with time-reversed transition probability $\tilde{\mathbf{P}}$. While $\mathbf{P}$ of systematic site update may not be time-reversible (i.e., $\mathbf{P}\neq\tilde{\mathbf{P}}$ in general), the random site update will provide a time-reversible $\mathbf{P}$.

Gibbs distribution.

Provided a finite set Λ of common states, we can introduce a **Gibbs distribution** on the finite collection S of configurations by

$$\pi(x) = \frac{1}{z_\beta} \exp(-\beta H(x)), \quad x \in S$$

where

$$z_\beta = \sum_{x \in S} \exp(-\beta H(x))$$

is the normalizing constant. The parameter β and z_β are respectively called an **inverse temperature** and **partition function**. The function $H(x)$, called **Hamiltonian**, is interpreted as an **energy function** of the configuration x, where it abhors to retain a high energy when the temperature $T = 1/\beta$ is low. That is, $\pi(x)$ is distinctively smaller for the configuration x with higher $H(x)$ when β is also large.

Hamiltonian function.

Let $G = (V, E)$ be a finite undirected graph, and let $\Lambda = \{1, \ldots, L\}$ be a set of common states. In the language of statistical mechanics, a vertex $v \in V$ is called a **site**, and the site v is said to be a **neighbor** of w if $\{v, w\} \in E$. For each site v we can find coordinate-wise state x_v, which is called a **spin** particularly when $\Lambda = \{-1, +1\}$. Then a Gibbs distribution $\pi(x)$ may be determined by "nearest-neighbor interaction." $C \subset V$ is called a **clique** if $\{v, w\} \in E$ whenever $v, w \in C$. We denote the set of all cliques by $\mathcal{C}$. Then a Hamiltonian $H(x)$ can be formulated by

$$H(x) = \sum_{C \in \mathcal{C}} W_C(x), \tag{3.1}$$

where each W_C depends only on those states x_v from $v \in C$.

Gibbs random field (GRF).

Let S be a space of configurations, and let $X = (X_v)$ be a S-valued random variable. X is an **Markov random field** (MRF) with respect to a graph $G = (V, E)$ if for every $v \in V$ and $\mathbf{x} \in S$

1. $\mathbb{P}(X = x) > 0$;
2. $\mathbb{P}(X_v = x_v | X_w = x_w, w \neq v)$
   $= \mathbb{P}(X_v = x_v | X_w = x_w, \{w, v\} \in E)$.

Hammersley-Clifford theorem. $\mathbf{X}$ is an MRF with respect to G if and only if $\pi(\mathbf{x}) = \mathbb{P}(\mathbf{X} = \mathbf{x})$ is a Gibbs distribution with Hamiltonian function (3.1) with respect to G. In this sense a Gibbs distribution of this particular Hamiltonian function is also known as **Gibbs random field (GRF)**.

Site update of GRF.

In the GRF the conditional probability for site update is given by

$$\pi(x_v \,|\, (x_w)_{w \neq v}) = \frac{\exp(-\beta H_v(x))}{\sum_{\lambda \in \Lambda} \exp(-\beta H_v(x; x_v \leftarrow \lambda))}$$

Here the normalizing constant z_β is canceled and

$$H_v(x) = \sum_{v \in C, C \in \mathcal{C}} W_C(x)$$

is only the partial sum over neighboring cliques of v. Thus, the summation is over the singleton $\{v\}$, the edges $\{v, w\} \in E$, and all the other cliques which contains v.

Ising model.

Let $G = (V, E)$ be a graph representing a lattice (or a rectangular grid) and let $\Lambda = \{-1, +1\}$ be the states of spin "up" and "down." Since the lattice has cliques of singleton or edges, the Hamiltonian of GRF is given by

$$H(x) = -J \sum_{\{v,w\} \in E} x_v x_w - h \sum_{v \in V} x_v$$

In statistical mechanics J and h represent the strength of interaction between neighbors and that of external magnetic field. The GRF with this particular Hamiltonian is called an **Ising model**. The conditional probability of a state x_v at the site v is a function of

$$H_v(x) = -J x_v \sum_{w:\{v,w\} \in E} x_w - h x_v$$

Attractive spin system.

Assuming $J = 1$ and $h = 0$, the Ising model has the following site update:

- Set $x_v \leftarrow +1$ with probability
$$\pi(+1 \,|\, (x_w)_{w \neq v}) = \left[1 + \exp\left(-2\beta \sum_{w:\{v,w\} \in E} x_w \right) \right]^{-1};$$
- Set $x_v \leftarrow -1$ with probability
$$\pi(-1 \,|\, (x_w)_{w \neq v}) = \left[1 + \exp\left(2\beta \sum_{w:\{v,w\} \in E} x_w \right) \right]^{-1}.$$

The chance for the site v getting $+1$ increases as the number of neighbors being $+1$. In this sense the model is called an **attractive spin system**.

Critical temperature.

The Ising model undergoes a "phase transition" between an ordered and a disordered phase. When the temperature $T = 1/\beta'$ is low, spins are aligned, creating a large cluster of aligned states. As the temperature gets higher, the randomness takes over. In the Ising model it is known that this phase transition occurs at the critical temperature T_c, and that $T_c = [2/\ln(1+\sqrt{2})]J \approx 2.27J$ when the graph G is a 2-dimensional grid with $h = 0$.

Potts model.

Let $\Lambda = \{1, \ldots, M\}$. Then the Ising model is generalized with the Hamiltonian

$$\hat{H}(x) = -\hat{J} \sum_{\{v,w\} \in E} \delta(x_v, x_w) - \hat{h} \sum_{v \in V} \delta(x_v, f_v)$$

where

$$\delta(\lambda, \lambda') = \begin{cases} 1 & \text{if } \lambda = \lambda'; \\ 0 & \text{if } \lambda \neq \lambda'. \end{cases}$$

Here f is a configuration of external reference. This GRF is called a **Potts model**.

R code: Behavior around critital temperature.

The Ising model is a special case of the Potts model with $M = 2$, since their respective Hamiltonian function $H(x)$ and $\hat{H}(\hat{x})$ satisfy $H(x) \equiv 2\hat{H}(\hat{x})$ when $x_v = 1$ or -1 accordingly as $\hat{x}_v = 1$ or 2. Here we set $J = 1/2$ and obtain the inverse temperature $\beta = \log(1 + \sqrt{2}) \approx 0.88$ for the Potts model with $\hat{J} = 1$. Then we can simulate the phase transition for Ising model.

```
source("potts.r")
potts(m=2,tmax=1000,beta=0.8)
potts(m=2,tmax=1000,beta=0.88)
```

Change the grid size of model, and observe the behavior around the critical temperature.

```
potts(m=2,n=c(100,100), tmax=1000,beta=0.88)
```

R code: Potts models with external force.

An external force at each site is indicated by $0, 1, \ldots, M$ where the state 0 implies no external force at a particular site. Entire data of external force is represented in a matrix.

```
s1 = matrix(data=c(1,2,3,3,1,2,2,3,1,1,2,2,1,1,1,2), 4, 4)
potts(m=3, tmax=0, init.state=s1)
potts(m=3, beta=0.9, tmax=10000, ref.state=s1)
```

An $n \times m$ matrix of data `a` is created by the function `matrix()` with `data`. The result is a matrix with `data` assigned from the first column to the last column of the matrix. To examine the external force in color, use it as initial state.

R code: Exploration of Potts models.

The initial value `init.state` and the external force `ref.state` can be assigned in a form of matrix. When `init.state` is set to an integer from 1 to M, the value is set to every site at time $t = 0$.

- Choose a different number M of common states and inverse temperature β.
- See what the Gibbs sampler generates with or without existence of external force.

```
potts(m=2,tmax=1000,beta=0.9)
potts(m=2,tmax=0, init.state=f1)
potts(m=2,tmax=1000,beta=0.9,ref.state=f1)
potts(m=5,tmax=1000,beta=1.1)
potts(m=5,tmax=0, init.state=f2)
potts(m=5,tmax=1000,beta=1.1,ref.state=f2)
```

Activities and report questions.

1. Describe or formulate the detailed balance of $\mathbf{P}$ when the Gibbs sampler is formed by random site update. In particular discuss ergodicity and time-reversibility for $\mathbf{P}$.
2. Explain similarities and differences between MHA and Gibbs sampler, and discuss advantages and disadvantages of Gibbs sampler.
3. Explain a Hamiltonian function in the context of Gibbs distribution, and discuss the significance of Hammersley-Clifford theorem in Gibbs sampler.