

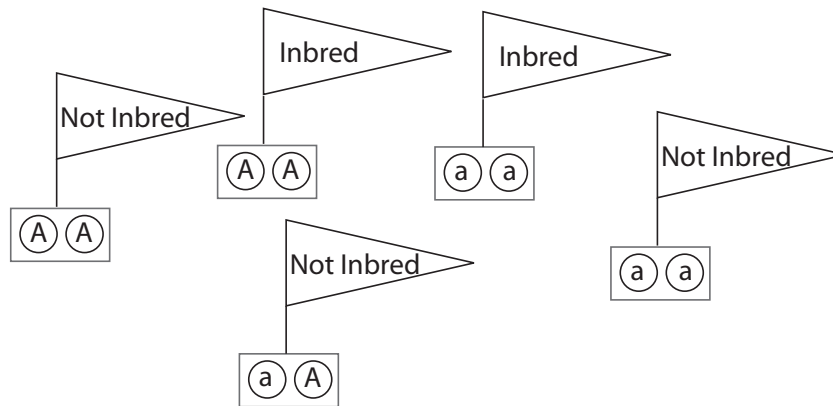
# Markov Chain Monte Carlo: more complex examples

Goals of this lecture:

- Consider MCMC in more than a single dimension
- Describe component-wise Metropolis-Hastings sampling
- Introduce the idea of latent variables and show how they lend themselves to Gibbs sampling—a special case of component-wise M-H sampling

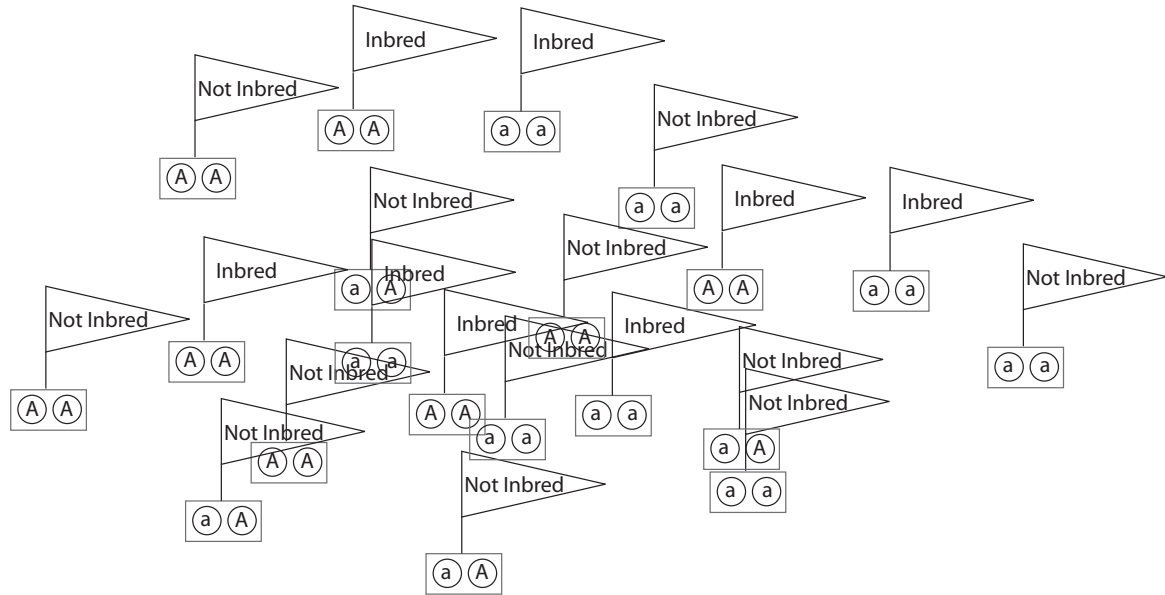
## Formulating the Model with Latent Variables:

Just think how easy it would be to estimate  $f$  and  $p$  if we knew whether every individual we sampled was inbred or not:



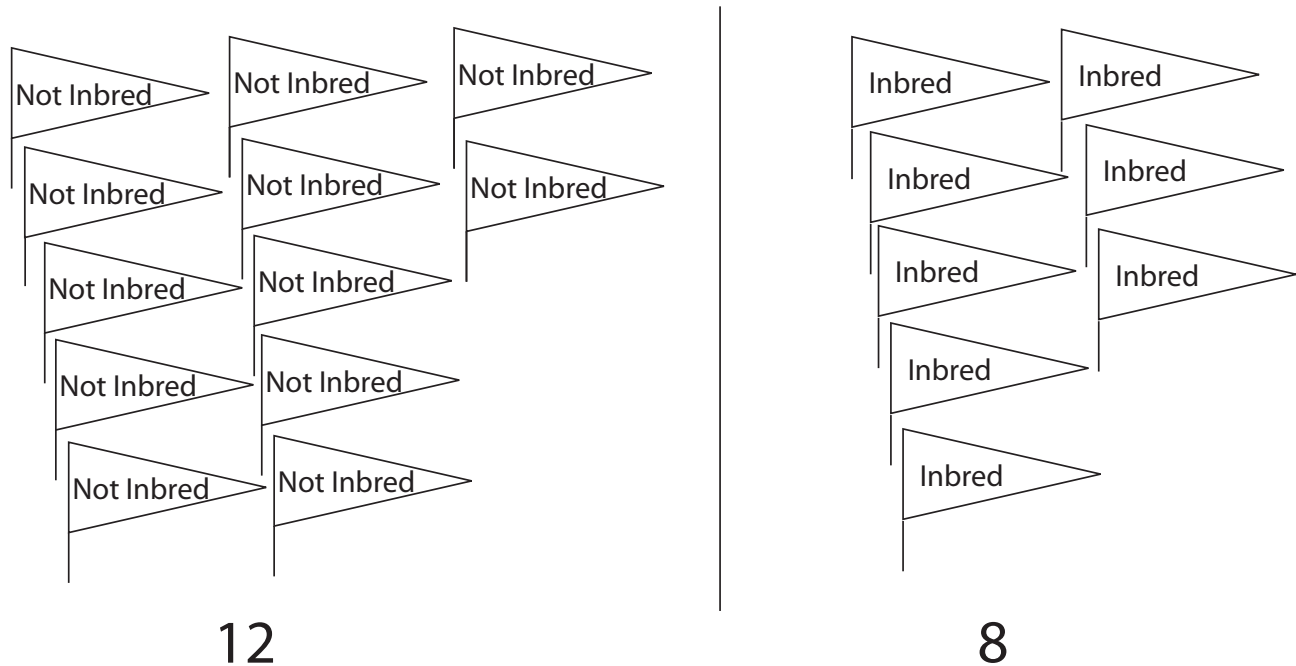
- Now,  $f$  is just a binomial proportion
- To estimate  $p$  you count both the alleles from Non-inbred individuals, and just one allele from Inbred individuals, and then it too is simply a binomial proportion.

In other words, if your data look like:

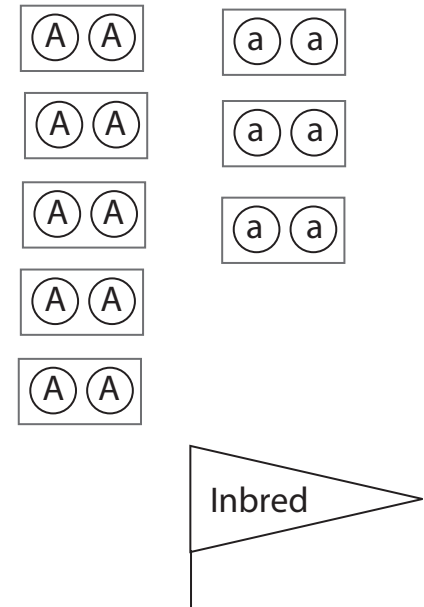
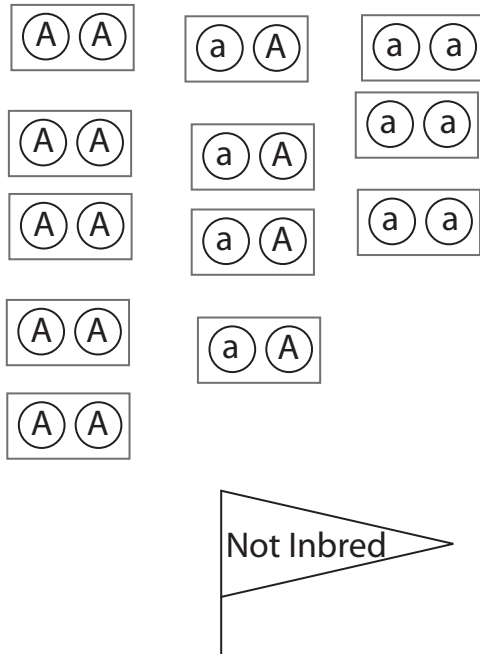


... then, to estimate  $f$ , you don't even need to think about the alleles carried by anyone. . .

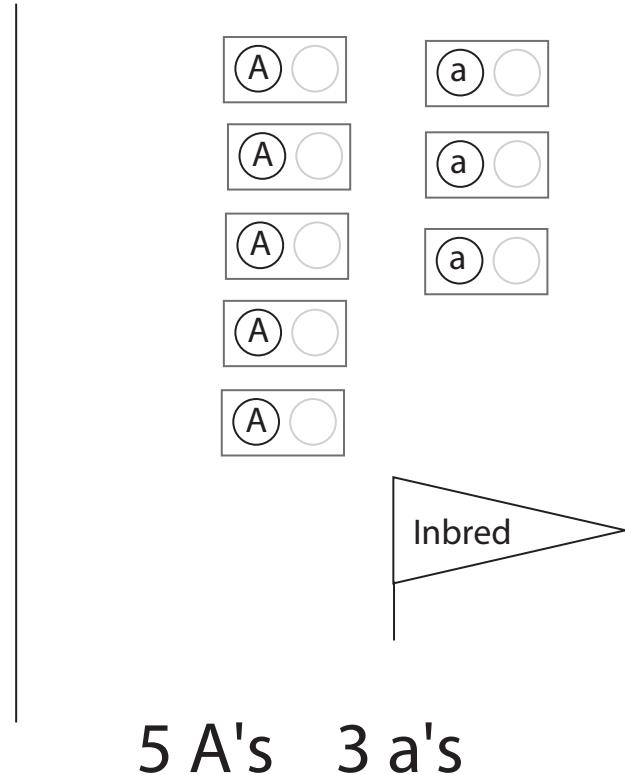
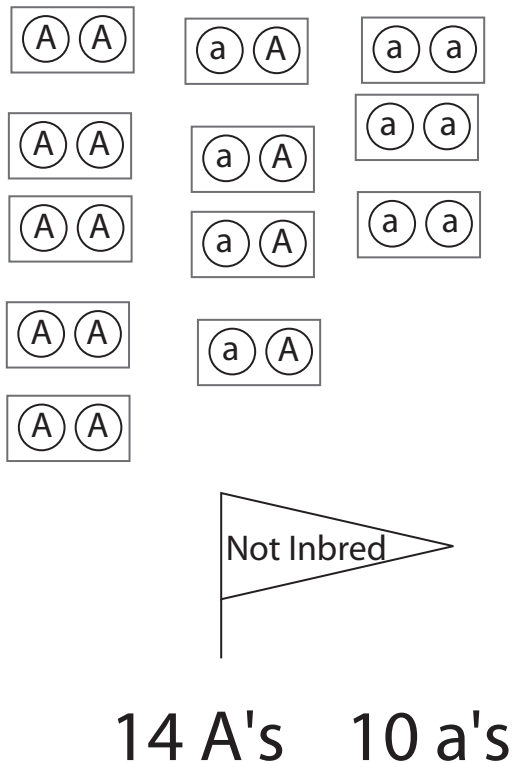
... you just have to “count the flags”:



To estimate  $p$  you just have to count alleles. . .



Keeping in mind each inbred individual contributes just a single gene copy...



## Joint Density of $p$ , $f$ , and $U$ :

- We denote the  $n$  latent variables as  $U = (U_1, \dots, U_n)$ .
- $u_i = 0$  indicates that  $i$  is Not Inbred.  $u_i = 1$  indicates  $i$  is Inbred.
- Let  $Y = (Y_1, \dots, Y_n)$  denote the genotypes

So, the joint density is:

$$P(p, f, U, Y) = P(p)P(f)P(U|f)P(Y|U, p)$$

and, so, the posterior is:

$$P(p, f, U|Y) = \frac{P(p, f, U, Y)}{\int_p \int_f \sum_{0 \leq U_1 \leq 1} \cdots \sum_{0 \leq U_n \leq 1} P(p, f, U, Y) dp df}$$

The normalizing constant is even nastier! It seems we have gained little, until we remember that we don't have to mess with the normalizing constant, ...but even then it's not immediately clear we have gained anything.

## Generic MCMC for $f$ and $p$ with latent variables:

- Recall, we wish to simulate  $f$  and  $p$  from their posterior distributions and so learn about  $P(f, p|Y)$ .
- However, now, our chain is moving about the space of  $f$ ,  $p$ , and  $U$ ...
  - it visits a sequence of states of the form  $(f^{(1)}, p^{(1)}, U^{(1)})$ ,  $(f^{(2)}, p^{(2)}, U^{(2)})$ ,  $(f^{(3)}, p^{(3)}, U^{(3)})$ , ...
- **Important Concept:** Obtaining the Marginal Posterior Distribution from MCMC output is simple:
  - From the sequence  $(f^{(1)}, p^{(1)}, U^{(1)})$ ,  $(f^{(2)}, p^{(2)}, U^{(2)})$ , ... you can just focus on collecting the values  $(f^{(1)}, p^{(1)})$ ,  $(f^{(2)}, p^{(2)})$ , ..., discarding the  $U^{(\cdot)}$ 's if you are not interested in their posterior distribution. (Note, though, that you might be interested in the individual  $U_i$ 's!).



## Naive Metropolis-Hastings for $f$ , $p$ , and $U$ :

1. Do an update for  $p$ :

(a) propose  $p^*$  from  $\text{Normal}(p, s_p)$

(b) accept or reject on the basis of

$$\frac{q(p|p^*)}{q(p^*|p)} \times \frac{P(p^*)P(f)P(U|f)P(Y|U, p^*)}{P(p)P(f)P(U|f)P(Y|U, p)}$$

2. Do an update for  $f$ :

(a) propose  $f^*$  from  $\text{Normal}(f, s_f)$

(b) accept or reject on the basis of

$$\frac{q(f|f^*)}{q(f^*|f)} \times \frac{P(p)P(f^*)P(U|f^*)P(Y|U, p)}{P(p)P(f)P(U|f)P(Y|U, p)}$$

3. Do an update for  $U$  ( $n$  separate updates, one for each  $U_i$ ).

For  $i$  in  $1, \dots, n$ :

(a) Propose a new  $U_i^*$  by flipping a coin (heads=0, tails=1).

(b) Accept or reject on the basis of

$$\begin{aligned} & \frac{q(U_i|U_i^*)}{q(U_i^*|U_i)} \times \frac{P(p)P(f)P(U^*|f)P(Y|U^*,p)}{P(p)P(f)P(U|f)P(Y|U,p)} \\ &= \frac{q(U_i|U_i^*)}{q(U_i^*|U_i)} \times \frac{P(p)P(f)P(U_i^*|f)P(Y|U_i^*,p)}{P(p)P(f)P(U_i|f)P(Y|U_i,p)} \end{aligned}$$

Note the cancellations in all these Hastings Ratios.

Computer Demo: `inbred_p` (“Silly” using (S), (f), and (p)) <sup>1</sup>

<sup>1</sup>which will segue into Gibbs sampling

## Gibbs Sampling:

Gibbs sampling is a special case of the component-wise M-H sampler in which the proposal distribution is the *full conditional distribution*.

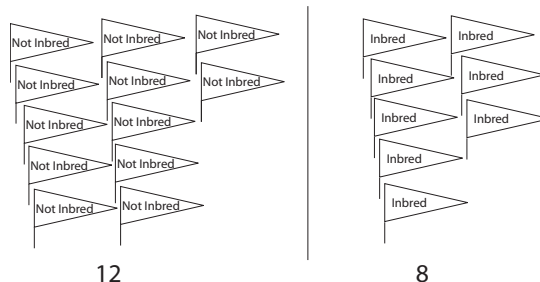
Gibbs Sampling Step for  $f$ :

- Recall the generic component-wise M-H update:

$$\frac{q(f|f^*)}{q(f^*|f)} \times \frac{P(p)P(f^*)P(U|f^*)P(Y|U, p)}{P(p)P(f)P(U|f)P(Y|U, p)}$$

The full conditional distribution for  $f$  is the distribution of  $f$  conditional upon the current values of all other variables in the model. This must be proportional to the product of all the factors in the joint density that have  $f$  in them:  $P(f)P(U|f)$

- The  $P(\mathbf{U}|f)$  portion of
- the joint density pertains to “counting up our flags”

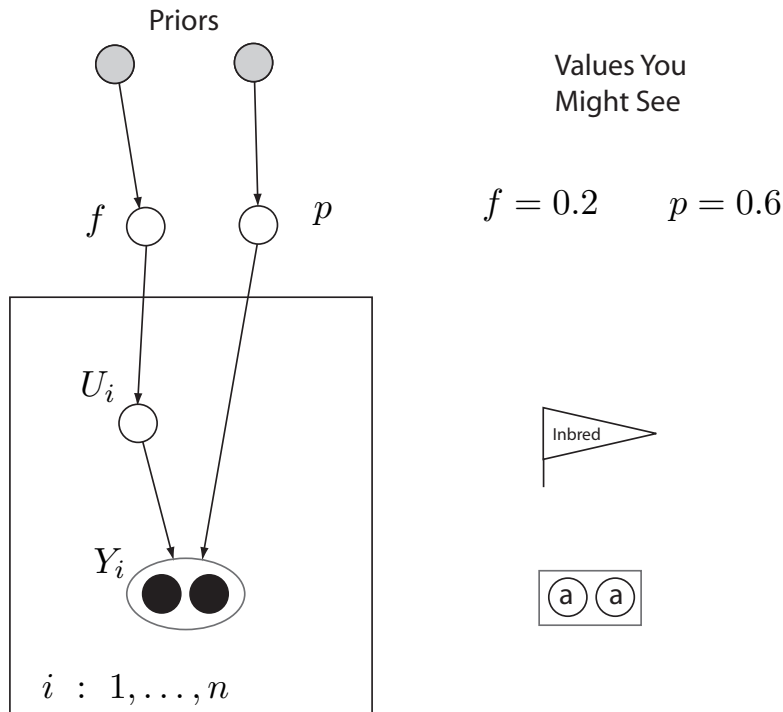


- $P(f)$  is the prior for  $f$
- So the full conditional for  $f$  is a beta distribution—it’s the “posterior” for  $f$  given the “data”  $\mathbf{U}$
- Letting the proposal  $q(\cdot|\cdot)$  be that full conditional gives us a Hastings Ratio of

$$\frac{P(f)P(\mathbf{U}|f)/C}{P(f^*)P(\mathbf{U}|f^*)/C} \times \frac{P(p)P(f^*)P(\mathbf{U}|f^*)P(\mathbf{Y}|\mathbf{U},p)}{P(p)P(f)P(\mathbf{U}|f)P(\mathbf{Y}|\mathbf{U},p)}$$

- Which is 1—This is always the case with Gibbs sampling—you always accept the proposal from the full conditional distribution.

## A Directed Graphical View of The Model... :

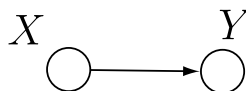


...makes it fairly easy to envision how Gibbs updates for  $p$  and  $U$  would proceed.

Computer Demo: `inbred_p` ("Gibbs" using (g))

## DAG notation and terminology:

○ “node”      ↓ “arrow”



$X$  is a parent of  $Y$

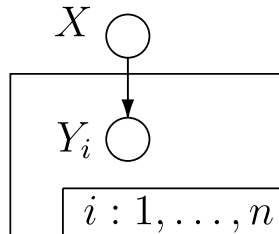
$Y$  is a daughter of  $X$

○  $X$  is an unobserved variable

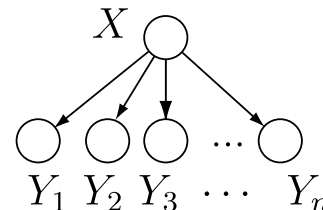
●  $X$  is an observed variable

◐  $X$  is a variable with an assumed value (for a prior)

$n$ -plates:



a shorthand for:



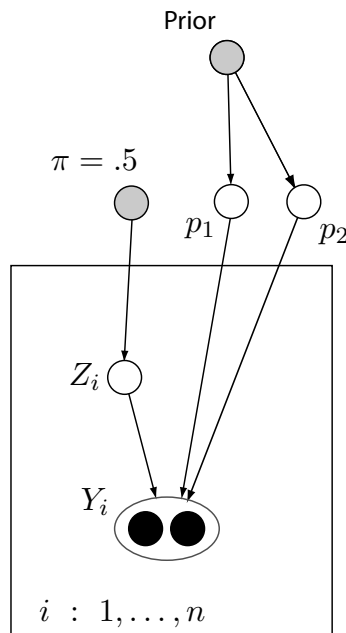
There are  $n$  variables  $Y_i$ , each conditionally independent given  $X$ .

Factorization:

$$P(\text{all}) = \prod_{\text{all nodes } i} P(\text{node } i | \text{parents of node } i)$$

## A reminder that all inference is conditional upon the underlying model... and a small step toward *structure*:

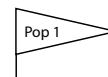
We could have attributed the departure from Hardy-Weinberg proportions to a mixture (in the proportions of  $\pi = .5$ ) of two populations with allele frequencies  $p_1$  and  $p_2$ , respectively.



Values You Might See

$$p_1 = 0.6$$

$$p_2 = .4$$



This is, in fact, the *structure* model with no admixture,  $K = 2$ , and a single locus.

Gibbs sampling is straightforward. . .

Computer Demo: `newhybs` via `RunNewHybs.sh`

## Wrap-Up:

### Main Points:

- In problems where it is useful, MCMC almost always proceeds by proposing changes to a small subset of the variables.
- There may be many different proposal types.
  - Each proposal type *must* satisfy detailed balance.
  - Each proposal type need not make an irreducible chain, BUT
  - All proposals taken together should form an irreducible chain.
- Latent variables may help in factorizing complex joint densities and lend themselves to Gibbs sampling.
- Simple proposals may be fast to implement, but may not lead to good mixing of the chain.
- Designing good proposal distributions is where one gets to perform “art” in implementing MCMC.