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

Genotype Frequencies and Inbreeding:

Starting with the last exercise of Session 3: one locus with two alleles, A and a, at frequencies p and $1-p$, respectively, and "inbreeding coefficient" f.

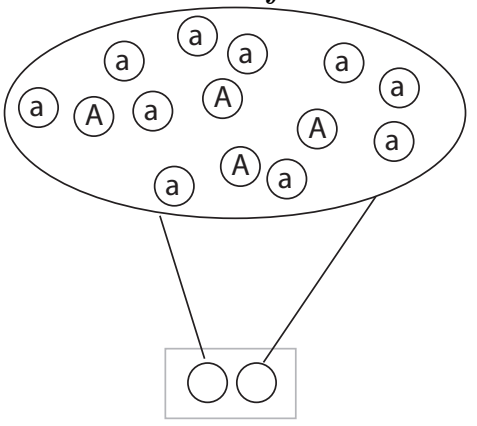
Probabilities of the three genotypes are:

- $P(AA) = fp + (1-f)p^2$
- $P(Aa \text{ or } aA) = (1-f)2p(1-p)$
- $P(aa) = f(1-p) + (1-f)(1-p)^2$

Since "inbreeding" is used to describe a lot of (related) things, let's briefly review what this model is saying...

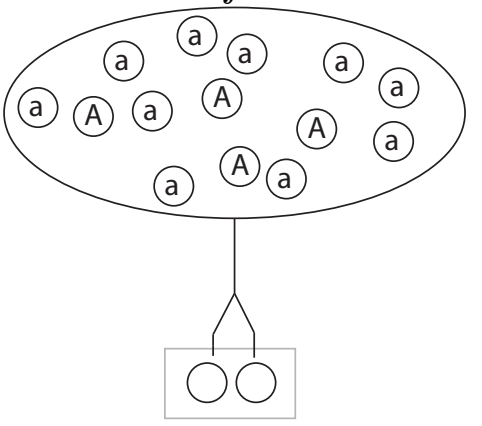
Inbreeding model:

Not-inbred with probability $1 - f$



- $P(AA) = p^2$
- $P(Aa \text{ or } aA) = 2p(1-p)$
- $P(aa) = (1-p)^2$

Inbred with probability f



- $P(AA) = p$
- $P(Aa \text{ or } aA) = 0$
- $P(aa) = (1-p)$

$$P(n_{AA}, n_{Aa}, n_{aa}|p, f) = C \times [fp + (1-f)p^2]^{n_{AA}} \times [(1-f)2p(1-p)]^{n_{Aa}} \times [f(1-p) + (1-f)(1-p)^2]^{n_{aa}}$$

What Our Data Would Look Like:

- We have a sample of n individuals total.
  - n_{AA} are homozygous for the A allele
  - n_{Aa} are heterorozygous
  - n_{aa} are homozygous for the a allele
- Clearly, $n = n_{AA} + n_{Aa} + n_{aa}$

A concrete example we will use throughout this lecture is $n = 50$ with:

$$n_{AA} = 30 \qquad n_{Aa} = 10 \qquad n_{aa} = 10$$

which are roughly the expected values if $p = .7$ and $f = .5$.

Bayesian Model for Estimating f and p:

To go about estimating f and p in a Bayesian fashion, we must fully specify the model. This means that we need

- Priors for f and p. We will choose uniform priors $f \sim \texttt{Beta}(1, 1)$ and $p \sim \texttt{Beta}(1, 1)$
- The data themselves. These are n_{AA}, n_{Aa}, and n_{aa}.
- The likelihood: $P(n_{AA}, n_{Aa}, n_{aa}|p, f)$, which is given on the previous page

The posterior distribution is then prior $\times$ likelihood, divided by the ("nasty") normalizing constant

$$P(p, f|n_{AA}, n_{Aa}, n_{aa}) = \frac{P(f)P(p)P(n_{AA}, n_{Aa}, n_{aa}|p, f)}{\int_{f,p} P(f)P(p)P(n_{AA}, n_{Aa}, n_{aa}|p, f)df\,dp}$$

Computing the normalizing constant is difficult, but with MCMC, *we don't have to!*

Recall, to perform MCMC it suffices to know the target distribution up to a constant of proportionality.

Our target distribution is

$$P(p, f|n_{AA}, n_{Aa}, n_{aa}) \propto P(f)P(p)P(n_{AA}, n_{Aa}, n_{aa}|p, f)$$

So long as $0 < f < 1$ and $0 < p < 1$.

Otherwise it is 0.

Two-Dimensional M-H Sampler:

We can compute the target distribution (up to a constant) easily for any f and p. So, to simulate from this posterior distribution we can just implement a Metropolis-Hastings sampler.

Following the examples of Session 3, for p we will choose a normal proposal distribution centered on the current value with standard deviation of s_p:

$$q(p^*|p) \equiv \text{Normal}(p, s_p)$$

And, we will use the same for f:

$$q(f^*|f) \equiv \text{Normal}(f, s_f)$$

Applying these proposal distributions in sequence gives us a simple way to simulate proposed values, (p^*, f^*), from the current values (p, f).

A "sweep" of our MCMC algorithm would look like:

1. Propose a new value, (p^*, f^*) for (p, f)
   - propose p^* from $\mathsf{Normal}(p, s_p)$
   - propose f^* from $\mathsf{Normal}(f, s_f)$
2. Accept or reject the proposed value (p^*, f^*) with probability R which is the minimum of 1 and the Hasting's Ratio:

$$R = \min\left\{1, \frac{q(p|p^*)q(f|f^*)}{q(p^*|p)q(f^*|f)} \times \frac{P(p^*)P(f^*)P(n_{AA}, n_{Aa}, n_{aa}|p^*, f^*)}{P(p)P(f)P(n_{AA}, n_{Aa}, n_{aa}|p, f)}\right\}$$

If you accept the proposed value, set the current value to the proposed value. Otherwise leave the current values unchanged.

Computer Demo: `inbred_p -n 30 10 10` (starts on "Jointly" (j))

Component-wise Metropolis Hastings Sampler :

- In any MCMC implementation, the proposal distribution *need not propose changes to* **every** *variable/parameter in the model*.
- In fact, there are few "real-world" problems requiring MCMC in which you would use a single proposal distribution in which changes were proposed to all the variables in the model.
- **Important Concept:**
  - Any proposal distribution, regardless of how many or how few variables it proposes changes to, is valid, so long as the proposal is accepted or rejected in a way that satisfies detailed balance w.r.t. the target distribution.
  - These different flavors of the "propose-reject/accept" step may be combined in series in whatever manner is desired, so long as they produce an irreducible, aperiodic chain[1].

[1] Nonetheless, some ways are better than others and will lead to a better-mixing chain

Simple Component-wise M-H Sampler for p and f:

Simplest scenario has a sweep as follows:

1. Do an update for p:
   (a) propose p^* from $\text{Normal}(p, s_p)$
   (b) Accept or reject the proposed value p^* with probability $\min\{1, \alpha\}$, where:

$$\alpha = \frac{q(p|p^*)}{q(p^*|p)} \times \frac{P(p^*)P(f)P(n_{AA}, n_{Aa}, n_{aa}|p^*, f)}{P(p)P(f)P(n_{AA}, n_{Aa}, n_{aa}|p, f)}$$

2. Do an update for f:
   (a) propose f^* from $\text{Normal}(f, s_f)$
   (b) Accept or reject the proposed value f^* with probability $\min\{1, \alpha\}$, where:

$$\alpha = \frac{q(f|f^*)}{q(f^*|f)} \times \frac{P(p)P(f^*)P(n_{AA}, n_{Aa}, n_{aa}|p, f^*)}{P(p)P(f)P(n_{AA}, n_{Aa}, n_{aa}|p, f)}$$

Simple Component-wise M-H Sampler for p and f, cont'd:

Some very important points:

- The proposals are each a little simpler (though just slightly...) than jointly proposing changes to (p, f)
- Neither step 1 nor step 2 of the sweep creates an irreducible chain (obviously, if you never update p, for example, your chain could never reach every possible value of p).
- However, taken together, steps 1 and 2 create an irreducible chain.

Computer Demo: `inbred_p` (Component-wise using (c) from info window)

Simple Component-wise M-H Sampler for p and f, cont'd:

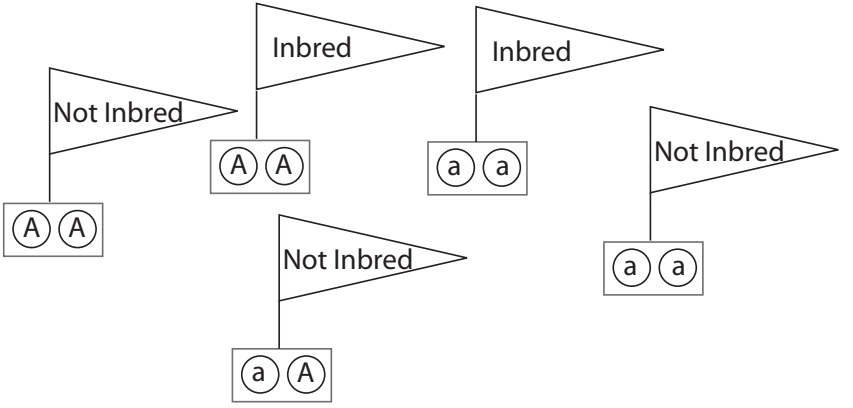
Two more important points

- The form of the Hastings ratio is a little simpler when we have proposed changing just a subset of the variables.
- However, in this case the target density remains just as complex, because it does not factorize into a separate part for f and a part for p.
- Since we are changing just a small part of the model at a time, it seems like we could spend some more energy on making each separate proposal distribution more "intelligent."

The final two points above get us to thinking about Gibbs sampling, which we will return to after a brief discussion of latent variables...

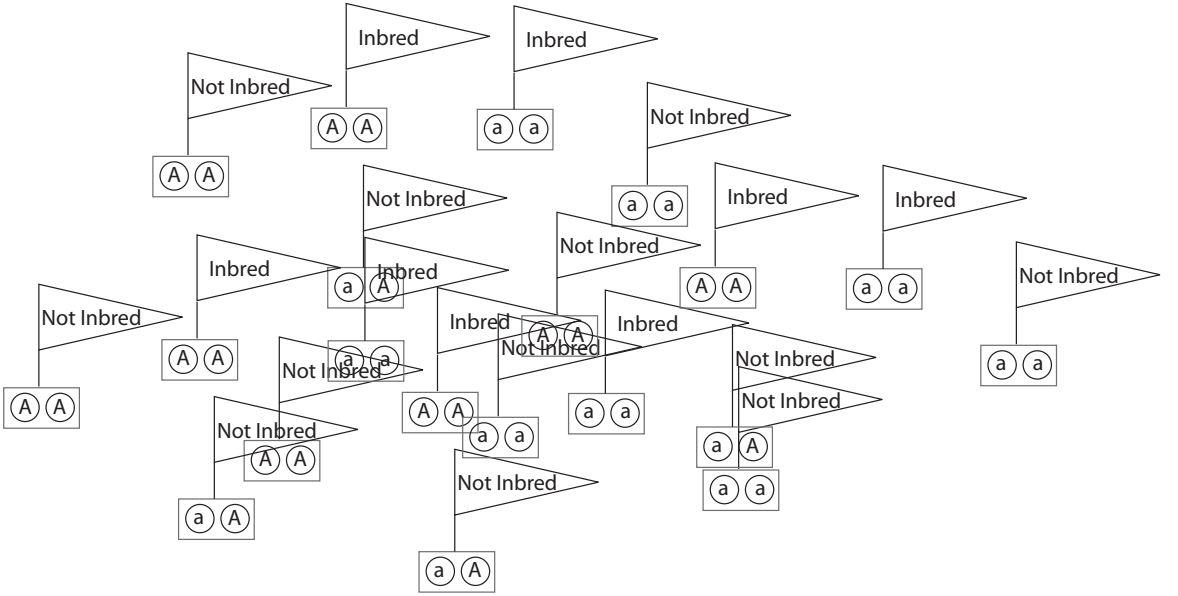
Formulating the Model with Latent Variables:

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



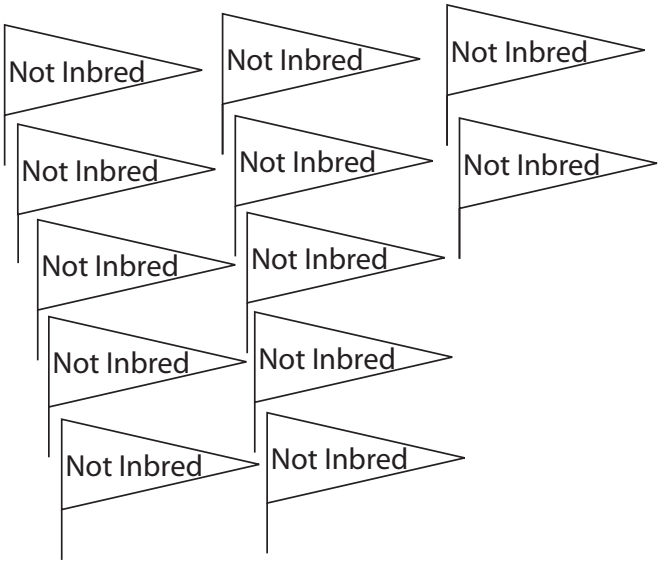
- Now, f is just a binomial proportion
- To estimate p you count both the alleles from Non-inbred individuals, and just one allele from Inbred individuals, and then it too is simply a binomial proportion.

In other words, if your data look like:

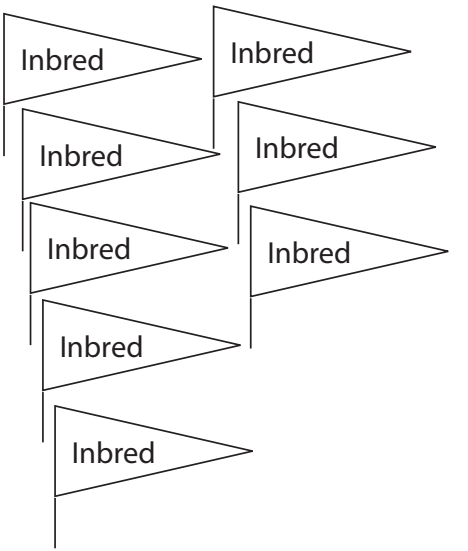


. . . then, to estimate f, you don't even need to think about the alleles carried by anyone. . .

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

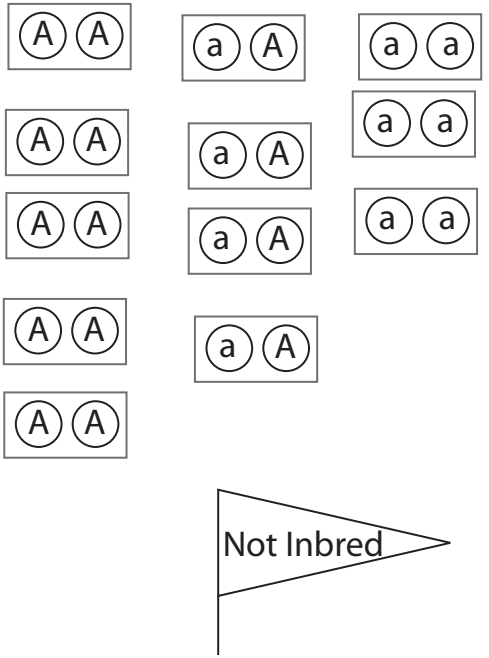


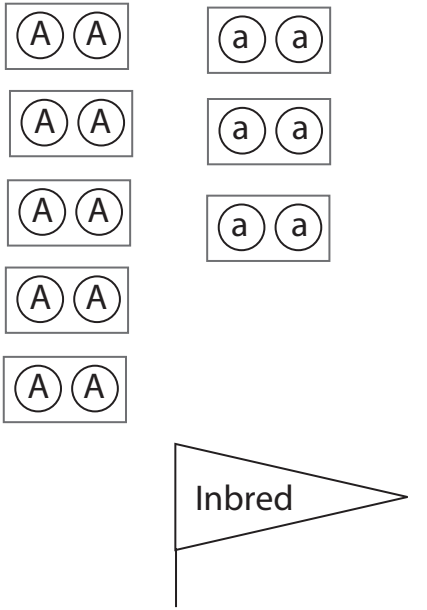
12



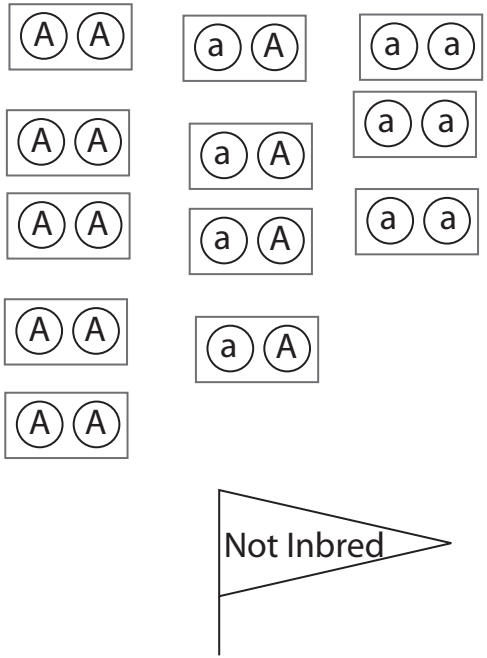
8

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

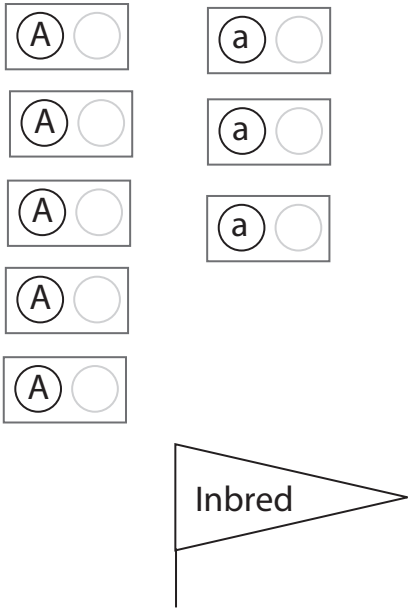




Keeping in mind each inbred individual contributes just a single gene copy. . .



14 A's 10 a's



5 A's 3 a's

Joint Density of p, f, and U:

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

So, the joint density is:

$$P(p, f, \boldsymbol{U}, \boldsymbol{Y}) = P(p)P(f)P(\boldsymbol{U}|f)P(\boldsymbol{Y}|\boldsymbol{U}, p)$$

and, so, the posterior is:

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

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 $P(f, p|\boldsymbol{Y})$.
- However, now, our chain is moving about the space of f, p, *and* $\boldsymbol{U}\ldots$
  - it visits a sequence of states of the form $(f^{(1)}, p^{(1)}, \boldsymbol{U}^{(1)})$, $(f^{(2)}, p^{(2)}, \boldsymbol{U}^{(2)})$, $(f^{(3)}, p^{(3)}, \boldsymbol{U}^{(3)}), \ldots$

- **Important Concept**: Obtaining the Marginal Posterior Distribution from MCMC output is simple:
  - From the sequence $(f^{(1)}, p^{(1)}, \boldsymbol{U}^{(1)})$, $(f^{(2)}, p^{(2)}, \boldsymbol{U}^{(2)}), \ldots$ you can just focus on collecting the values $(f^{(1)}, p^{(1)})$, $(f^{(2)}), \ldots$, discarding the $\boldsymbol{U}^{(\cdot)}$'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

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

2. Do an update for f:
   (a) propose f^* from $\text{Normal}(f, s_f)$
   (b) accept or reject on the basis of

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

3. Do an update for $\boldsymbol{U}$ (n separate updates, one for each U_i).
   For i in $1, \ldots, n$:
   (a) Propose a new U_i^* by flipping a coin (heads=0, tails=1).
   (b) Accept or reject on the basis of

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

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

Note the cancellations in all these Hastings Ratios.

Computer Demo: `inbred_p` ("Silly" using (S), (f), and (p)) [2]

[2] 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*.

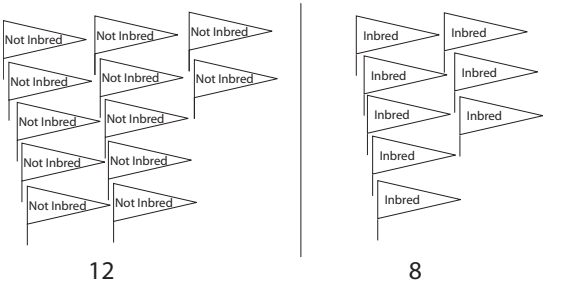
Gibbs Sampling Step for f:

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

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

  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: $P(f)P(\boldsymbol{U}|f)$

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

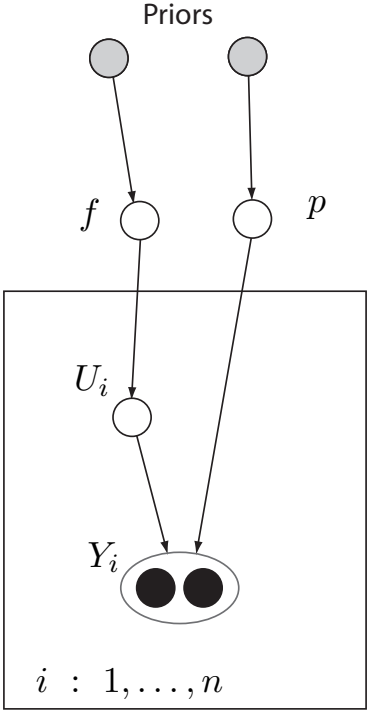


- $P(f)$ is the prior for f
- So the full conditional for f is a beta distribution—it's the "posterior" for f given the "data" $\boldsymbol{U}$
- Letting the proposal $q(\cdot|\cdot)$ be that full conditional gives us a Hastings Ratio of

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

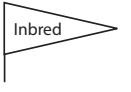
- Which is 1—This is always the case with Gibbs sampling—you always accept the proposal from the full conditional distribution.

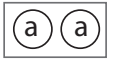
A Directed Graphical View of The Model. . . :



Values You Might See

$f = 0.2 \qquad p = 0.6$

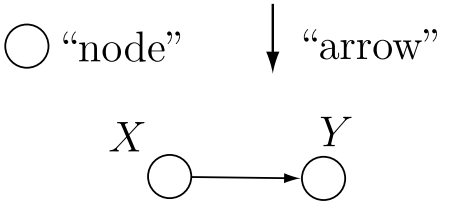




. . . makes it fairly easy to envision how Gibbs updates for p and $\boldsymbol{U}$ would proceed.

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

DAG notation and terminology:



X is a parent of Y
Y is a daughter of X

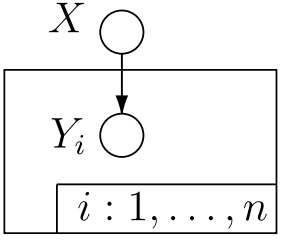
X is an unobserved variable

X is an observed variable

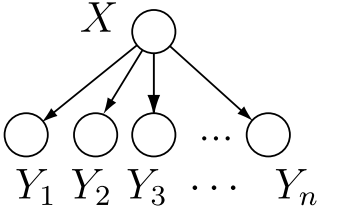
X is a variable with an assumed value (for a prior)

---

n-plates:



a shorthand for:



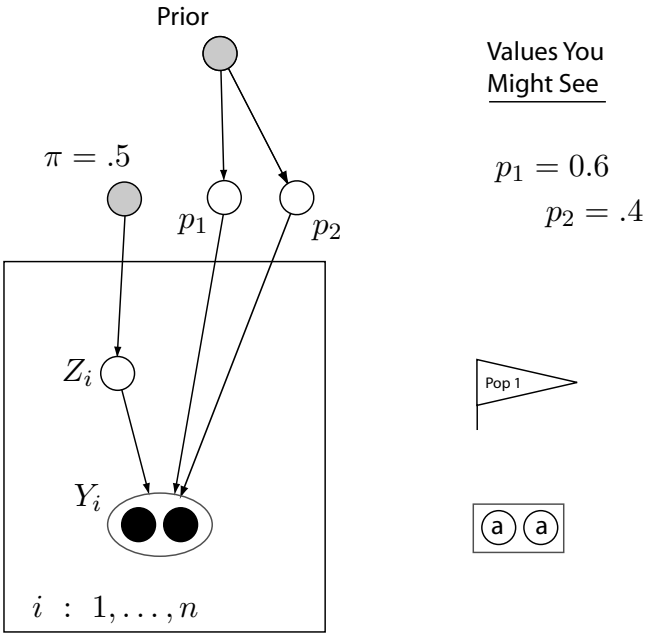
There are n variables Y_i, each conditionally independent given X.

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 $p2$, respectively.



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.