Gibbs sampler to estimate population size using mark-recapture data

Posterior distribution

The parameters that we primarily interested in are δ and π. However, to allow for conjugacy, we will also be estimating the status of each individual (z_i) as well. Here is what we are after:

$$p(\delta, \pi, \{z_i\}|\{D_{it}\}) \propto p(\{D_{it}\}|\delta, \pi, \{z_i\})p(\{z_i\}|\delta, \pi)p(\delta, \pi)$$

Notice that detection $\{D_{it}\}$ does not depend on π and therefore

$$p(\{D_{it}\}|\delta, \pi, \{z_i\}) = p(\{D_{it}\}|\delta, \{z_i\})$$

Similarly, if an individual exists or not z_i does not depend on δ. This implies that

$$p(\{z_i\}|\delta, \pi) = p(\{z_i\}|\pi)$$

Putting all this together and assuming that δ and π are a priori independent, we get:

$$\propto p(\{D_{it}\}|\delta, \{z_i\})p(\{z_i\}|\pi)p(\delta)p(\pi)$$

$$\propto \left\{\prod_{i=1}^{\tilde{N}}\prod_{t=1}^{T} Bernoulli(D_{it}|\delta \times z_i)\right\}\left\{\prod_{i=1}^{\tilde{N}} Bernoulli(z_i|\pi)\right\} Beta(\delta|a_\delta, b_\delta)Beta(\pi|a_\pi, b_\pi)$$

Full conditional distributions

To create our own Gibbs sampler in R, we have to write the full conditionals, one parameter at a time.

1) FCD for detection parameter δ:

$$p(\delta|\ldots) \propto \left\{\prod_{i=1}^{\tilde{N}}\prod_{t=1}^{T} Bernoulli(D_{it}|\delta \times z_i)\right\} Beta(\delta|a_\delta, b_\delta)$$

$$\propto \left\{\prod_{i\in\Omega}\prod_{t=1}^{T} \delta^{D_{it}}(1-\delta)^{1-D_{it}}\right\} \delta^{a_\delta - 1}(1-\delta)^{b_\delta - 1}$$

where Ω is the set of all individuals for which $\mathrm{z_i} = 1$.

$$\propto \left\{\delta^{\sum_{i\in\Omega}\sum_{t=1}^{T} D_{it}}(1-\delta)^{\sum_{i\in\Omega}\sum_{t=1}^{T}[1-D_{it}]}\right\} \delta^{a_\delta - 1}(1-\delta)^{b_\delta - 1}$$

$$\propto \delta^{\left[\sum_{i\in\Omega}\sum_{t=1}^{T} D_{it}\right]+a_\delta - 1}(1-\delta)^{\left[\sum_{i\in\Omega}\sum_{t=1}^{T}[1-D_{it}]\right]+b_\delta - 1}$$

$$p(\delta|\ldots) = Beta\left(\left[\sum_{i\in\Omega}\sum_{t=1}^{T} D_{it}\right] + a_\delta, \left[\sum_{i\in\Omega}\sum_{t=1}^{T}[1-D_{it}]\right] + b_\delta\right)$$

2) FCD for proportion of potential individuals that truly exist π:

$$p(\pi|\dots) \propto \left\{\prod_{i=1}^{\tilde{N}} Bernoulli(z_i|\pi)\right\} Beta(\pi|a_\pi, b_\pi)$$

$$\propto \left\{\prod_{i=1}^{\tilde{N}} \pi^{z_i}(1-\pi)^{1-z_i}\right\} \pi^{a_\pi - 1}(1-\pi)^{b_\pi - 1}$$

$$\propto \left\{\pi^{\sum_{i=1}^{\tilde{N}} z_i}(1-\pi)^{\sum_{i=1}^{\tilde{N}} 1-z_i}\right\} \pi^{a_\pi - 1}(1-\pi)^{b_\pi - 1}$$

$$\propto \pi^{\left[\sum_{i=1}^{\tilde{N}} z_i\right]+a_\pi - 1}(1-\pi)^{\left[\sum_{i=1}^{\tilde{N}} 1-z_i\right]+b_\pi - 1}$$

$$p(\pi|\dots) = Beta\left(\left[\sum_{i=1}^{\tilde{N}} z_i\right] + a_\pi, \left[\sum_{i=1}^{\tilde{N}} 1 - z_i\right] + b_\pi\right)$$

3) FCD for the status of each individual:

Say we have 100 potential individuals. One would think that we need to sample 100 z_i parameters, one representing the status of each individual. In reality, however, the z_i's that correspond to individuals that we saw at least once (i.e., $D_{it} = 1$ for at least one t) are not truly random quantities. In other words, according to the assumptions in our model, we know these individuals truly exist (i.e., $z_i = 1$). The problem is when we never see a particular animal (i.e., when $D_{it} = 0$ for all t). In this case, either:

- the animal exists but we failed to detect them; OR
- the animal does not exist.

In other words, z_i is a random quantity only when $D_{it} = 0$ for all t. For this reason, we just write the full conditional distributions for these animals.

$$p(z_i|D_{i1} = 0, \dots, D_{iT} = 0, \dots) \propto \left\{\prod_{t=1}^{T} Bernoulli(D_{it}|\delta \times z_i)\right\} Bernoulli(z_i|\pi)$$

To use the equal sign, we can write this using a generic constant K:

$$p(z_i|D_{i1} = 0, \dots, D_{iT} = 0, \dots) = K\left\{\prod_{t=1}^{T} Bernoulli(D_{it}|\delta \times z_i)\right\} Bernoulli(z_i|\pi)$$

$$= K \times \left\{\prod_{t=1}^{T} (\delta \times z_i)^{D_{it}}[1 - (\delta \times z_i)]^{1-D_{it}}\right\} \times \{\pi^{z_i}(1-\pi)^{1-z_i}\}$$

Because we know that $D_{it} = 0$ for all t, this becomes:

$$= K \times [1 - (\delta \times z_i)]^T \times \{\pi^{z_i}(1-\pi)^{1-z_i}\} \text{ [eqn. 1]}$$

Now let's examine what is $p(z_i = 1|D_{i1} = 0, \dots, D_{iT} = 0, \dots)$. If we substitute $z_i = 1$ into eqn. 1, we get $K \times [1-\delta]^T \times \pi$. One way to interpret this equation is that the animal exists (given by probability π) but we just failed to observe it in every single survey (given by probability $[1-\delta]^T$). Similarly, what is $p(z_i = 0|D_{i1} = 0, \dots, D_{iT} = 0, \dots)$? If we substitute $z_i = 0$ into eqn. 1, we get $K \times (1-\pi)$.

It would seem like we are stuck because we do not know what our constant K is. However, we actually have all the information we need to determine K. We know that:

$$p(z_i = 0|D_{i1} = 0, \dots, D_{iT} = 0, \dots) + p(z_i = 1|D_{i1} = 0, \dots, D_{iT} = 0, \dots) = K(1-\delta)^T\pi + K(1-\pi) = 1$$

Therefore:

$$K = \frac{1}{(1-\delta)^T\pi + (1-\pi)}$$

This implies that:

$$p(z_i = 1|D_{i1} = 0, \dots, D_{iT} = 0, \dots) = K(1-\delta)^T\pi = \frac{(1-\delta)^T\pi}{(1-\delta)^T\pi + (1-\pi)}$$

If we want to sample z_i from its FCD, we generate a Bernoulli random variable with probability given by $\frac{(1-\delta)^T\pi}{(1-\delta)^T\pi+(1-\pi)}$. In other words:

$$z_i|D_{i1} = 0, \dots, D_{iT} = 0, \dots \sim Bernoulli\left(\frac{(1-\delta)^T\pi}{(1-\delta)^T\pi + (1-\pi)}\right)$$