Conditional Posterior Derivations for "Particle Gibbs Sampling for Regime-Switching State-Space Models"

1 Switching Population Model

1.1 Mathematical Model

Let $S_{j,t}$ and $C_{j,t}$ denote the number of stage j adult penguins and penguin chicks at year t, respectively, and $S_t = \sum_{j=3}^{J} S_{j,t}$ and $C_t = \sum_{j=1}^{J-2} C_{j,t}$ represent the total number of adults and chicks irrespective of age. Consider the following multinoulli switching transition process for modeling the evolution of penguin populations over T years:

$$r_t \sim \text{Multinoulli}(\boldsymbol{\gamma}), \tag{1}$$

$$S_{1,t} \sim \text{Binomial}(0.5C_{t-1}, \psi), \tag{2}$$

$$S_{j,t} \sim \text{Binomial}(S_{j-1,t-1}, \psi),\ j = 2, \dots, J, \tag{3}$$

$$S_{J,t} \sim \text{Binomial}(S_{J-1,t-1} + S_{J,t-1}, \psi), \tag{4}$$

$$C_{j,t} \sim \text{Binomial}(2S_{j,t}, \phi_{r_t}),\ j = 1, \dots, J-2, \tag{5}$$

where ψ denotes the survival rate and ϕ_{r_t} denotes the reproductive rate of regime r_t (hence, we assume that the reproductive rates switch). Importantly, penguins cannot be aged in the field and census counts capture only the total number of adults and chicks, irrespective of age. Errors in counting are assumed proportional to abundance and we model the total number in each class as:

$$\tilde{S}_t \sim \mathcal{N}(S_t, \sigma_s^2 S_t^2), \tag{6}$$

$$\tilde{C}_t \sim \mathcal{N}(C_t, \sigma_c^2 C_t^2). \tag{7}$$

The unknowns of the model are the regimes $r_t \in \{1, \dots, K\}$, the states $\mathbf{x}_t = [S_{1,t}, \dots, S_{J,t}, C_{1,t}, \dots, C_{J-2,t}]^\intercal$, the measurement/transition parameters $\boldsymbol{\theta} = [\psi, \phi_1, \dots, \phi_K, \sigma_s^2, \sigma_c^2]^\intercal$ and the regime transition probabilities $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_K]$.

1.2 Survival Conditional Posterior

Here, we derive $p(\psi|\boldsymbol{\theta}_{-\psi}, \boldsymbol{\gamma}, \mathbf{x}_{0:T}, r_{1:T}, \mathbf{y}_{1:T})$. We begin by pointing out that the random variable ψ is conditionally independent of all other random quantities when $\mathbf{x}_{0:T}$ is given. Then, our true goal is in finding $p(\psi|\mathbf{x}_{0:T})$. Clearly

$$p(\psi|\mathbf{x}_{0:T}) \propto p(\mathbf{x}_{0:T}|\psi)p(\psi).$$

We choose $p(\psi)$ to be a Beta distribution with parameters a_{ψ_0} and b_{ψ_0}, i.e., $\psi \sim \text{Beta}(a_{\psi_0}, b_{\psi_0})$. It can be shown that

$$p(\mathbf{x}_{0:T}|\psi) \propto \left(\prod_{t=1}^{T} \psi^{S_{1,t}} (1-\psi)^{0.5C_{t-1} - S_{1,t}} \left(\prod_{j=2}^{J-1} \psi^{S_{j,t}} (1-\psi)^{S_{j-1,t-1} - S_{j,t}} \right) \psi^{S_{J,t}} (1-\psi)^{S_{J,t-1} + S_{J-1,t-1} - S_{J,t}} \right)$$

$$= \prod_{t=1}^{T} \psi^{S_{1,t}} (1-\psi)^{0.5C_{t-1} - S_{1,t}} \psi^{\sum_{j=2}^{J-1} S_{j,t}} (1-\psi)^{\sum_{j=2}^{J-1} S_{j-1,t-1} - S_{j,t}} \psi^{S_{J,t}} (1-\psi)^{S_{J,t-1} + S_{J-1,t-1} - S_{J,t}}$$

$$= \prod_{t=1}^{T} \psi^{\sum_{j=1}^{J} S_{j,t}} (1-\psi)^{0.5C_{t-1}+\sum_{j=1}^{J} S_{j,t-1}-\sum_{j=1}^{J} S_{j,t}}$$
$$= \psi^{\sum_{t=1}^{T}\sum_{j=1}^{J} S_{j,t}} (1-\psi)^{\sum_{t=1}^{T}(0.5C_{t-1}+\sum_{j=1}^{J} S_{j,t-1}-\sum_{j=1}^{J} S_{j,t})}.$$

We can see that $p(\mathbf{x}_{0:T}|\psi)$ is proportional to a Binomial distribution with $N_S = \sum_{t=1}^{T}(0.5C_{t-1} + \sum_{j=1}^{J} S_{j,t-1})$ trials and probability of success ψ evaluated at $n_S = \sum_{t=1}^{T}\sum_{j=1}^{J} S_{j,t}$ occurrences. By conjugacy, the conditional posterior $p(\psi|\mathbf{x}_{0:T})$ can easily be determined and is a Beta distribution:

$$p(\psi|\mathbf{x}_{0:T}) = \text{Beta}(\psi; a_\psi, b_\psi),$$

where $a_\psi = a_{\psi_0} + n_S$ and $b_\psi = b_{\psi_0} + N_S - n_S$.

1.3 Reproductive Rates Conditional Posterior

Here we derive $p(\phi_k|\boldsymbol{\theta}_{-\phi_k}, \boldsymbol{\gamma}, \mathbf{x}_{0:T}, r_{1:T}, \mathbf{y}_{1:T})$. For simplicity, we start off by assuming no constraint on ϕ_k (we will take care of this later through simple modification of the prior distribution). We begin by pointing out that the random variable ϕ_k is conditionally independent of all other random quantities when given $\mathbf{x}_{1:T}$ and $r_{1:T}$. Thus, determining the full conditional posterior is equivalent to solving for $p(\phi_k|\mathbf{x}_{1:T}, r_{1:T})$. We choose $p(\phi_k)$ to be a Beta distirbution with parameters $a_{\phi_{k0}}$ and $b_{\phi_{k0}}$, i.e., we assume that $\phi_k \sim \text{Beta}(a_{\phi_{k0}}, b_{\phi_{k0}})$. To solve for $p(\phi_k|\mathbf{x}_{1:T}, r_{1:T})$ we apply Bayes' theorem to arrive at:

$$p(\phi_k|\mathbf{x}_{1:T}, r_{1:T}) \propto p(\mathbf{x}_{1:T}|\phi_k, r_{1:T})p(\phi_k),$$

where we have made the appropriate assumption that ϕ_k is independent of $r_{1:T}$. Therefore, all we need to do is to simplify the "likelihood term" $p(\mathbf{x}_{1:T}|\phi_k, r_{1:T})$. Let $\mathcal{T}_k = \{t|r_t = k\}$. By simple manipulation of the expression, we have that:

$$p(\mathbf{x}_{1:T}|\phi_k, r_{1:T}) \propto \prod_{t\in\mathcal{T}_k}\prod_{j=3}^{J} \phi_k^{C_{j-2,t}}(1-\phi_k)^{2S_{j,t}}$$
$$= \prod_{t\in\mathcal{T}_k} \phi_k^{C_t}(1-\phi_k)^{2S_t}$$
$$= \phi_k^{\sum_{t\in\mathcal{T}_k} C_t}(1-\phi_k)^{2\sum_{t\in\mathcal{T}_k} S_t}$$

We can see that $p(\mathbf{x}_{1:T}|\phi_k, r_{1:T})$ is proportional to a Binomial distribution with $N_{C,k} = 2\sum_{t\in\mathcal{T}_k} S_t$ trials with success probability ϕ_k, evaluated at $n_{C,k} = \sum_{t\in\mathcal{T}_k} C_t$ occurrences. By conjugacy, we will also find that the conditional posterior is a Beta distribution:

$$p(\phi_k|\mathbf{x}_{1:T}, r_{1:T}) = \text{Beta}(\phi_k; a_{\phi_k}, b_{\phi_k}),$$

where $a_{\phi_k} = a_{\phi_{k0}} + n_{C,k}$ and $b_{\phi_k} = b_{\phi_{k0}} + N_{C,k} - n_{C,k}$. In the case that we have parameter constraints, these would be incorporated into the prior using an indicator function. Let $\phi_0 = 0$ and $\phi_{K+1} = 1$. When taking the monotonicity constraints $\phi_0 < \phi_1 < \cdots < \phi_K < \phi_{K+1}$ into consideration, the prior $p(\phi_k)$ becomes

$$p(\phi_k) \propto \text{Beta}(\phi_k; a_{\phi_{k0}}, b_{\phi_{k0}})1_{\phi_k\in(\phi_{k-1},\phi_{k+1})},$$

which is a Beta distribution with truncated support. Then, under these monotonicity constraints, the conditional posterior distribution is a truncated Beta distribution:

$$p(\phi_k|\mathbf{x}_{1:T}, r_{1:T}, \phi_{-k}) \propto \text{Beta}(\phi_k; a_{\phi_k}, b_{\phi_k})1_{\phi_k\in(\phi_{k-1},\phi_{k+1})}.$$

1.4 Noise Parameters Conditional Posterior

Here, we aim to derive $p(\sigma_s^2|\boldsymbol{\theta}_{-\sigma_s^2}, \boldsymbol{\gamma}, \mathbf{x}_{0:T}, r_{1:T}, \mathbf{y}_{1:T})$ and $p(\sigma_c^2|\boldsymbol{\theta}_{-\sigma_c^2}, \boldsymbol{\gamma}, \mathbf{x}_{0:T}, r_{1:T}, \mathbf{y}_{1:T})$. Since the measurement distributions of $\tilde{S}_t$ and $\tilde{C}_t$ share the similar definition, as shown in (6) and (7), we will take the conditional posterior of σ_s^2 as an example. The conditional posterior of σ_c^2 can be obtained directly by replacing several variables.

Based on the population model, the $p(\sigma_s^2|\boldsymbol{\theta}_{-\sigma_s^2}, \boldsymbol{\gamma}, \mathbf{x}_{0:T}, r_{1:T}, \mathbf{y}_{1:T})$ can be simplified to $p(\sigma_s^2|\mathbf{x}_{0:T}, \mathbf{y}_{1:T})$. Again,

$$\begin{aligned} p(\sigma_s^2|\mathbf{x}_{0:T}, \mathbf{y}_{1:T}) &\propto p(\mathbf{y}_{1:T}|\mathbf{x}_{0:T}, \sigma_s^2) p(\sigma_s^2|\mathbf{x}_{0:T}) \\ &= p(\tilde{S}_{1:T}|S_{1:T}, \sigma_s^2) p(\sigma_s^2) \end{aligned}$$

where we assume that the noise parameter σ_s^2 has an inverse gamma prior with positive parameters $a_{\sigma_s 0}$ and $b_{\sigma_s 0}$, i.e., $p(\sigma_s^2) \sim \mathrm{IG}(\sigma_s^2; a_{\sigma_s 0}, b_{\sigma_s 0})$, $a_{\sigma_s 0} > 0$ and $b_{\sigma_s 0} > 0$. Then, the prior distribution is

$$p(\sigma_s^2) = \frac{b_{\sigma_s 0}^{a_{\sigma_s 0}}}{\Gamma(a_{\sigma_s 0})} {\sigma_s}^{-2(a_{\sigma_s 0}+1)} \exp(-b_{\sigma_s 0}/\sigma_s^2)\,.$$

Due to the independence of measurements, the likelihood $p(\tilde{S}_{1:T}|S_{1:T}, \sigma_s^2)$ can be factorized. Considering the Gaussian distribution of $\tilde{S}_t$ in (6), we have

$$\begin{aligned} p(\tilde{S}_{1:T}|S_{1:T}, \sigma_s^2) &= \prod_{t=1}^{T} p(\tilde{S}_t|S_t, \sigma_s^2) \\ &= \frac{\sigma_s^{-T}(2\pi)^{-T/2}}{\prod_{t=1}^{T} S_t} \exp\Big[-\frac{1}{2}\sum_{t=1}^{T}\frac{(\tilde{S}_t - S_t)^2}{\sigma_s^2 S_t^2}\Big] \end{aligned}$$

Therefore, the posterior is

$$\begin{aligned} p(\sigma_s^2|\mathbf{x}_{0:T}, \mathbf{y}_{1:T}) &\propto p(\tilde{S}_{1:T}|S_{1:T}, \sigma_s^2) p(\sigma_s^2) \\ &= \frac{b_{\sigma_s 0}^{a_{\sigma_s 0}} (2\pi)^{-T/2} \sigma_s^{-2(a_{\sigma_s 0}+1)-T}}{\Gamma(a_{\sigma_s 0}) \prod_{t=1}^{T} S_t} \exp\Big[-\frac{1}{2}\sum_{t=1}^{T}\frac{(\tilde{S}_t - S_t)^2}{\sigma_s^2 S_t^2} - \frac{b_{\sigma_s 0}}{\sigma_s^2}\Big] \\ &\propto \frac{b_{\sigma_s}^{a_{\sigma_s}}}{\Gamma(a_{\sigma_s})} \sigma_s^{-2(a_{\sigma_s}+1)} \exp(-b_{\sigma_s}/\sigma_s^2) \end{aligned}$$

In other words, the conditional posterior is also an inverse gamma distribution:

$$p(\sigma_s^2|\mathbf{x}_{0:T}, \mathbf{y}_{1:T}) \propto \mathrm{IG}(\sigma_s^2; a_{\sigma_s}, b_{\sigma_s})$$

where

$$\begin{aligned} a_{\sigma_s} &= a_{\sigma_s 0} + \frac{T}{2} \\ b_{\sigma_s} &= b_{\sigma_s 0} + \frac{1}{2}\sum_{t=1}^{T}\frac{(\tilde{S}_t - S_t)^2}{S_t^2} \end{aligned}$$