Bayesian Genetic Mark-Recapture Methods For Estimating Seasonal River Run Size Of Stock Populations

Yiran Wang, Martin Lysy, Audrey Béliveau

Department of Statistics and Actuarial Science
University of Waterloo

June 9, 2021

UNIVERSITY OF WATERLOO

UNIVERSITY OF WATERLOO
FACULTY OF MATHEMATICS
Department of Statistics and Actuarial Science

Table of Contents

Motivating Application

Figure: Sockeye Salmon (Photo by Kevin Phillips)

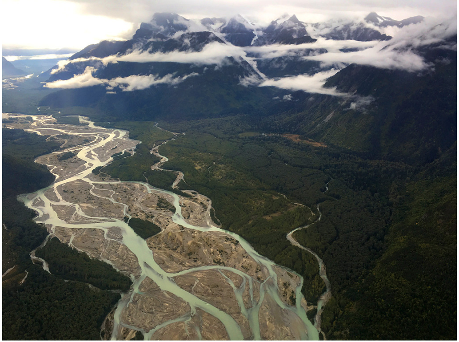

Figure: Taku River (Photo by Paul Vecsei)

Motivating Application

Genetic Mark-Recapture (GMR) Method

Genetic Data:

Motivating Application

Genetic Mark-Recapture (GMR) Method

Genetic Data:
Estimated Proportions for species of interest

Motivating Application

Genetic Mark-Recapture (GMR) Method

Genetic Data:
Estimated Proportions for species of interest

Count Data:

Motivating Application

Genetic Mark-Recapture (GMR) Method

Genetic Data:
Estimated Proportions for species of interest

Count Data:
Counts for some species

Motivating Application

Genetic Mark-Recapture (GMR) Method

Genetic Data:
Estimated Proportions for species of interest

$$(\mu_1, \ldots, \mu_K)$$
$$(\sigma_1, \ldots, \sigma_K)$$

Count Data:
Counts for some species

Motivating Application

Genetic Mark-Recapture (GMR) Method

Genetic Data:
Estimated Proportions for species of interest

$$(\mu_1, ..., \mu_K)$$
$$(\sigma_1, ..., \sigma_K)$$

Count Data:
Counts for some species

$$N_1$$

Motivating Application

Genetic Mark-Recapture (GMR) Method

Genetic Data:
Estimated Proportions for species of interest

$$(\mu_1, ..., \mu_K)$$
$$(\sigma_1, ..., \sigma_K)$$

Count Data:
Counts for some species

$$N_1$$

$$N = \frac{N_1}{\mu_1}$$

Motivating Application

Genetic Stock Identification (GSI) Data

n_t, $t = 1, ..., T$: Sample size of genetic tissues in week t
$(\mu_{k,t})$: In-sample posterior stock proportion estimate in week t for stock k
$(\sigma_{k,t})$: In-sample posterior standard deviation in week t for stock k

Motivating Application

Genetic Stock Identification (GSI) Data

n_t, $t = 1, ..., T$: Sample size of genetic tissues in week t
$(\mu_{k,t})$: In-sample posterior stock proportion estimate in week t for stock k
$(\sigma_{k,t})$: In-sample posterior standard deviation in week t for stock k

Weir Count Data

N_1: Total aggregate count for lake-type stocks

Motivating Application

Genetic Stock Identification (GSI) Data

n_t, $t = 1, ..., T$: Sample size of genetic tissues in week t
$(\mu_{k,t})$: In-sample posterior stock proportion estimate in week t for stock k
$(\sigma_{k,t})$: In-sample posterior standard deviation in week t for stock k

Weir Count Data

N_1: Total aggregate count for lake-type stocks

Run Weight Data

w_t: The Sockeye Salmon run weight (relative abundance) in week t, with $\sum_{t=1}^{T} w_t = 1$

Method of Moments (MM)

In 2001, W. J. Gazey developed a method of moments estimator to estimate total abundance of specific Salmon stocks in the Alsek River system[2]. This method has been used to estimate the abundance of Sockeye Salmon in the Taku River system by Pestal et al.[1]:

$$\hat{N} = \frac{N_1}{\sum_{t=1}^{T} w_t \mu_{1,t}}, \tag{1}$$

with

$$\widehat{\text{Var}}(\hat{N}) = \sum_{t=1}^{T} (w_t \sigma_{1,t})^2 \left[\frac{\hat{N}}{\sum_{i=1}^{T} w_i \mu_{1,i}} \right]^2. \tag{2}$$

Method of Moments (MM)

In 2001, W. J. Gazey developed a method of moments estimator to estimate total abundance of specific Salmon stocks in the Alsek River system[2]. This method has been used to estimate the abundance of Sockeye Salmon in the Taku River system by Pestal et al.[1]:

$$\hat{N} = \frac{N_1}{\sum_{t=1}^{T} w_t \mu_{1,t}}, \tag{1}$$

with

$$\widehat{\text{Var}}(\hat{N}) = \sum_{t=1}^{T} (w_t \sigma_{1,t})^2 \left[\frac{\hat{N}}{\sum_{i=1}^{T} w_i \mu_{1,i}} \right]^2. \tag{2}$$

- **The variance may be significantly underestimated.**

Proposed Model

1. Let $X_{k,t}$ denote the number of fish of stock k in the GSI sample in week t, then:

$$(X_{1,t}, X_{2,t}, \dots, X_{K,t}) \sim \text{Multinom}(n_t; \pi_{1,t}, \pi_{2,t}, \dots, \pi_{K,t}), \tag{3}$$

   where $\pi_{k,t}$ is the true proportion of stock k in the population in week t.

2. The observed GSI proportions $\mu_{k,t}$ are described as

$$\mu_{k,t} \sim TN\left(\frac{X_{k,t}}{n_t}, \sigma^2_{k,t}\right) \tag{4}$$

   where TN represents a truncated normal distribution with a range of $[0, \infty)$. The choice of the normal distribution is motivated by the Bernstein-von Mises theorem.

Proposed Model

We can then derive the total number of Sockeye Salmon as

$$N = \frac{N_1}{\sum_{t=1}^{T} w_t \pi_{1,t}} \tag{5}$$

and stock-specific abundances as $N_k = N \sum_{t=1}^{T} w_t \pi_{k,t}$, for $k = 2, ..., K$.

Choice of Prior

Conjugate Dirichlet Prior (CDP)

$$(\pi_{1,t}, \ldots, \pi_{K,t}) \sim \text{Dirichlet}(\theta_1, \ldots, \theta_K) \tag{6}$$

When $\theta_k = 1$, it is a uniform prior; when $\theta_k = \frac{1}{2}$, it is a Jeffreys prior[3].

Choice of Prior

Transformed Normal Prior (TNP)[4]

$$\pi_{k,t} = \frac{\exp(\tau_{k,t})}{\sum_{i=1}^{K} \exp(\tau_{i,t})} \tag{7}$$

$$\tau_{k,t} \overset{ind}{\sim} N(0, \xi^2). \tag{8}$$

We consider various values for ξ: 0.5, 1, 2, ..., 5 as well as a uniform hyperprior $\xi \sim U(0, 10)$.

Choice of Prior

Time Series Prior (TSP)

$$\tau_{k,t} = \rho \tau_{k,t-1} + \epsilon_{k,t} \tag{9}$$

$$\tau_{k,1} \sim N\left(0, \frac{s^2}{1-\rho^2}\right) \tag{10}$$

$$\epsilon_{k,t} \sim N(0, s^2) \tag{11}$$

$$\rho \sim \text{Unif}(-1, 1) \tag{12}$$

For s, we consider the values 0.5, 1, 2, ..., 5 as well as a uniform hyperprior $s \sim U(0,5)$. We also consider the case where s and ρ vary per stock and week with independent priors.

Simple Simulation

- Settings:
  - 5 weeks, 3 stocks.
  - 500 GSI datasets with arbitrary preset values of $\pi_{k,t}$.
  - The true value of N was set as 80,000 and the weir count data N_1 was calculated from
    $$N = \frac{N_1}{\sum_{t=1}^{T} w_t \pi_{1,t}}$$
    as a function of N, $\pi_{k,t}$ and w_t, where $w_t = 0.2$ for $t = 1, ..., 5$.

Simple Simulation

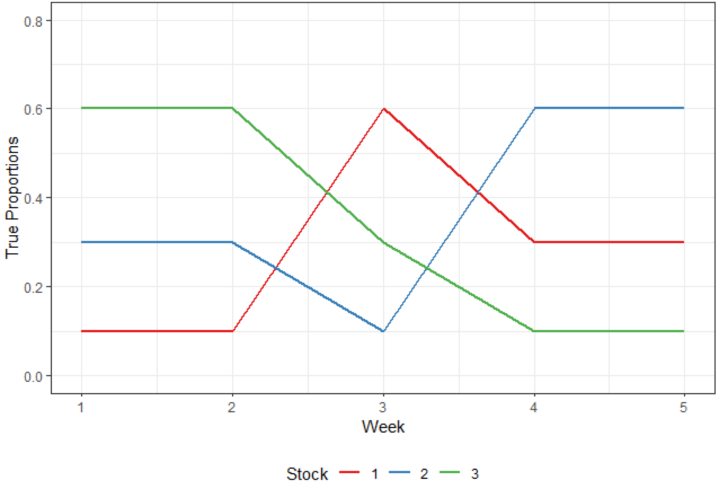

Figure: Parameter values ($\pi_{k,t}$) of the simple simulation.

Simple Simulation

- Results:
  - Close to the true value: absolute RBs<0.01, RRMSEs<0.07.
  - CP: 34.2% (MM) vs. 94.6% $\sim$ 96% (Proposed)
  - SD of $\hat{N}$: Proposed $= 4 \times$ MM
- The MM estimator underestimates the amount of uncertainty in the estimate.
- The MM estimator does not account for the fact that $\mu_{k,t}$ is the stock proportion in the GSI samples, instead of in the population.

Data-Based Simulation

- Settings:
  - 12 weeks, 17 stocks (4 lake-types vs. 13 river-types).
  - 500 GSI datasets with the observed $\mu_{k,t}$ in the GSI data as the values of $\pi_{k,t}$.
  - The true value of N was set as 60,000 based on the MM estimator in the PSC report[1].
  - The values of w_t and $\sigma_{k,t}$ were set as the observed values from the Taku River dataset.

Data-Based Simulation

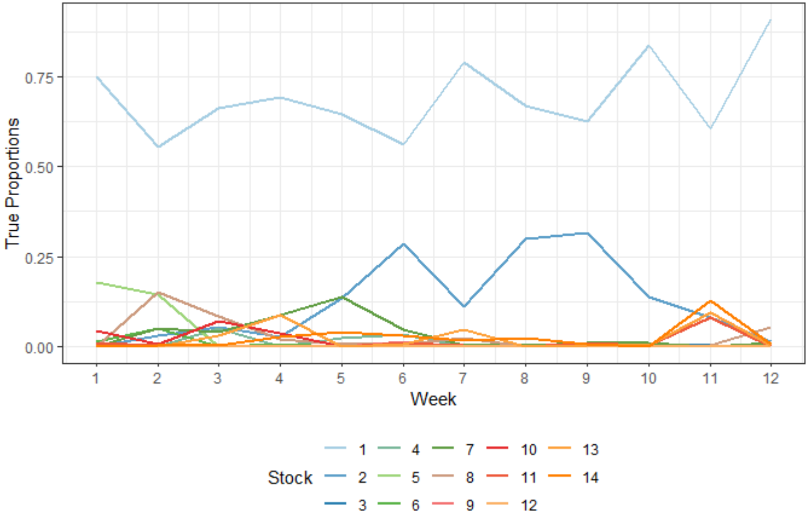

Figure: Parameter values ($\pi_{k,t}$) of the data-based simulation.

Data-Based Simulation

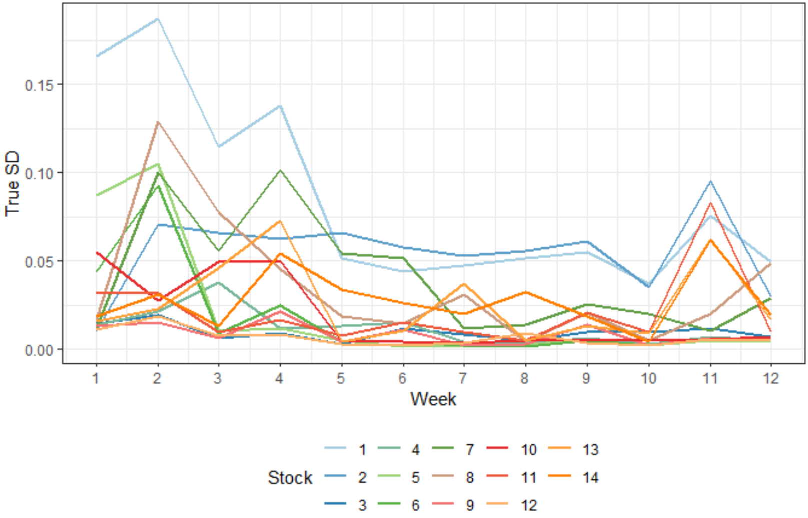

Figure: Standard deviation values ($\sigma_{k,t}$) of the data-based simulation.

Data-Based Simulation

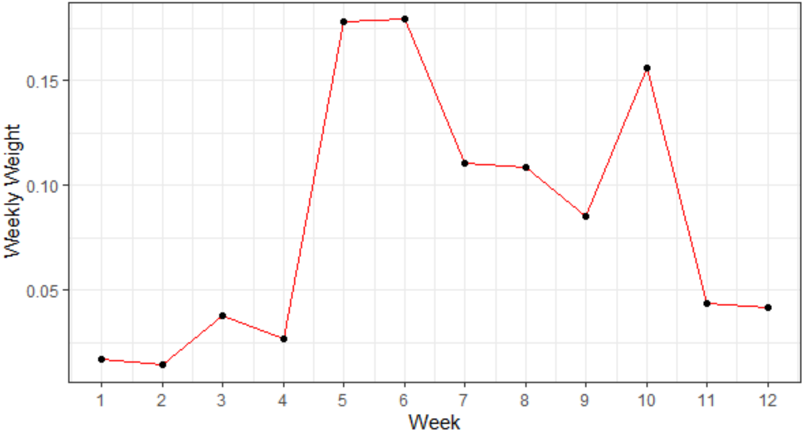

Figure: Run weights (w_t) of the data-based simulation.

Data-Based Simulation

Table 1: Results of the data-based simulation.

Setting	Relative Bias	RRMSE	SD	CP
MM	**0.00**	**0.032**	1,570	**0.88**
DCP $\theta_k = 1$	0.18	0.1868	2,334	0.00
DCP $\theta_k = 0.5$	0.09	0.092	2,042	0.18
TNP $\xi = 0.5$	0.25	0.253	2,621	0.00
TNP $\xi = 1$	0.06	0.064	1,983	0.58
TNP $\xi = 2$	**-0.01**	**0.030**	1,760	**0.93**
TNP $\xi = 3$	**-0.03**	**0.038**	1,694	**0.85**
TNP $\xi = 4$	**-0.03**	0.045	1,658	0.78
TNP $\xi = 5$	**-0.04**	0.049	1,634	0.69
TNP $\xi \sim \mathrm{U}(0, 10)$	**-0.05**	0.054	1,633	0.63
TSP $s \sim \mathrm{U}(0, 5)$	**-0.04**	0.052	1,640	0.64

Discussion

- The proposed method is particularly sensitive to the choice of prior, may due to some of the $\pi_{k,t}$'s being very close to zero.
- One possible explanation could be that when simulating the data, we did not constrain $\sum_{k=1}^{K} \mu_{k,t} = 1$ for $t = 1, .., T$.
- The TNP model may still be reasonable to use if various values of ξ are considered and model selection criteria such as the DIC or cross-validation are applied.
- Future Work:
  - Propose priors that would not be so sensitive.
  - Use spline method to explore the dynamics through time.
  - ...

References

Pestal, G., Schwarz, C.J. and Clark, R.A.(2020).
Taku River Sockeye Salmon Stock Assessment Review and Updated 1984-2018 Abundance Estimates.
Pacific Salmon Commission Technical Report No. 43.

Gazey, W. J.(2010).
GSI Sample Size Requirements for In-river Run Reconstruction of Alsek Chinook and Sockeye Stocks.
Pacific Salmon Commission, Vancouver, British Columbia.

Tuyl, F.(2017).
A Note on Priors for the Multinomial Model.
The American Statistician, 71(4), 298-301.

Gelman, A. (1995).
Method of moments using Monte Carlo simulation.
Journal of Computational and Graphical Statistics, 4(1), 36-54.

Thank You