Background for groupedBBMH R Package

Sumonkanti Das

2025-06-12

Overview of the groupedBBMH R Package

The **groupedBBMH** R package implements a **nested beta-binomial hierarchical model** designed for **group-testing data in biosecurity surveillance**, where the primary goal is to efficiently detect rare contamination events. In routine inspections of imported agricultural consignments, only group-level binary outcomes—positive or negative—are observed, without information on the exact number of contaminated items. This package extends traditional beta-binomial modeling to incorporate **imperfect test sensitivity and specificity**, offering both **exact inference** using a Metropolis-Hastings (MH) algorithm and **fast approximate estimation** via a beta-binomial approximation. It supports estimation of contamination levels, quantifies the risk of undetected contamination, and aids in risk-based decision-making through model-based simulation.

Case Study: Frozen Prawn Importation To illustrate the method, consider a biosecurity context involving the importation of frozen prawns into Australia. Annually, around 800 consignments (batches) are imported, with each batch containing approximately $B = 8000$ bags. A simple random sample of $b = 13$ bags is tested per batch. From each selected bag, a group of $m = 5$ prawns (out of $M = 40$) is randomly selected, leading to $n = bm = 65$ sampled prawns per batch. Each group of 5 prawns is tested using a PCR test with sensitivity $\Delta = 0.70$ and specificity $\Lambda = 0.99$. The only observation per batch is the number of groups testing positive, out of the $b = 13$ tested.

Let X_{ij} represent the number of contaminated prawns in bag j of batch i. Instead of directly observing X_{ij}, we observe Y_{ij}, where $Y_{ij} = 1$ if any prawn in the group is contaminated, and 0 otherwise.

Let define the total contamination as $T_{Xi} = \sum_{j=1}^{B} \sum_{k=1}^{M} X_{ijk}$, and the sample contamination as $t_{xi(m)} = \sum_{j=1}^{b} \sum_{k=1}^{m} X_{ijk}$. Similarly, $T_{Yi(M)} = \sum_{j=1}^{B} Y_{ij}$ and $t_{yi(m)} = \sum_{j=1}^{b} Y_{ij}$ are the total and sampled number of contaminated bags, respectively.

A key quantity of interest is **leakage**, defined as:

$$L_i = T_{Xi} \cdot \mathbb{I}(t_{yi(m)} = 0),$$

representing the number of contaminated prawns entering undetected when no sampled group tests positive. We focus on:

- the **expected leakage** $\mathbb{E}(L_i)$,
- the **probability of leakage** $\Pr(L_i > 0)$,
- and posterior inference on T_{Xi} when $t_{yi(m)} = 0$.

Now let p_i denote the true prevalence of contamination in batch i, and define $\phi_{i(m)} = 1 - (1 - p_i)^m$ as the probability that a pool of m ($1 \leq m \leq M$) prawns is contaminated. Assuming prawns are randomly distributed among bags, we have:

$$X_{ij} \mid p_i \overset{\text{i.i.d.}}{\sim} \text{Bin}(M, p_i), \quad p_i \overset{\text{i.i.d.}}{\sim} \text{Beta}(\alpha, \beta).$$

Thus, the group-level test outcome Y_{ij} follows:

$$Y_{ij} \mid p_i \sim \text{Bernoulli}(\phi_i), \quad \phi_i = 1 - (1 - p_i)^m,$$

and the observed number of positive groups in the sample is:

$$t_{yi} \mid p_i \sim \text{Binomial}(b, \phi_i).$$

To incorporate **imperfect testing**, we define the effective probability of a positive test as:

$$\tilde{\phi}_{\Delta\Lambda}(p_i) = \Delta\phi(p_i) + (1 - \Lambda)(1 - \phi(p_i)).$$

Under perfect specificity ($\Lambda = 1$), this simplifies to:

$$\tilde{\phi}_{\Delta}(p_i) = \Delta\phi(p_i).$$

When $\beta \gg \alpha$, the contamination prevalence p_i is approximately Gamma distributed, leading to:

$$\tilde{\phi}_{\Delta\Lambda} \sim (1 - \Lambda) + \text{Beta}\left(\alpha, \frac{\beta}{m(\Delta + \Lambda - 1)}\right).$$

Two important special cases:

- **Perfect testing** ($\Delta = 1$, $\Lambda = 1$): $\tilde{\phi} \sim \text{Beta}(\alpha, \beta/m)$,
- **Perfect specificity** ($\Lambda = 1$): $\tilde{\phi} \sim \text{Beta}(\alpha, \beta/(m\Delta))$.

In either case, t_{yi} approximately follows a Beta-Binomial distribution:

$$t_{yi} \sim \text{Beta-Binomial}(b, \alpha, \beta/(m\Delta)).$$

Threshold-Based Risk Estimation In the context of the prawn biosecurity study, suppose a regulatory threshold is set such that contamination levels below a certain prevalence cut-off are considered acceptable for import. Suppose a regulatory cut-off k is introduced, such that contamination prevalence below k is considered acceptable. The effective prevalence becomes:

$$p_i = p_i \cdot \mathbb{I}(p_i > k), \quad p_i \sim \text{Beta}(\alpha, \beta).$$

Using this truncated model, the **probability of leakage** becomes:

$$\Pr[L_i > 0] = \frac{B_{(k_1,1)}(\alpha, \frac{\beta}{m\Delta} + b)}{B(\alpha, \frac{\beta}{m\Delta})} - \frac{B_{(k,1)}(\alpha, \beta + MB)}{B(\alpha, \beta)},$$

with $k_1 = \Delta(1 - (1 - k)^m)$. Under perfect testing:

$$\Pr[L_i > 0] = \frac{B_{(k,1)}(\alpha, \beta + mb)}{B(\alpha, \beta)} - \frac{B_{(k,1)}(\alpha, \beta + MB)}{B(\alpha, \beta)}.$$

When $k = 0$, this reduces to the standard beta-binomial case.

The **expected leakage** under threshold k is:

$$\mathbb{E}(L_i) = (B-b)M \cdot \mathbb{E}\left[(1-p_i)^{bm\Delta} \cdot p_i \cdot \mathbb{I}(p_i > k)\right],$$

which simplifies to:

$$\mathbb{E}(L_i) = (B-b)M \cdot \frac{B_{(k,1)}(\alpha+1, \beta+bm\Delta)}{B(\alpha,\beta)}.$$

Model-Based Simulation to Assess Leakage Risk Using the posterior estimates of Beta parameters (α, β), we simulate contamination and testing outcomes for $D = 1000$ consignments of frozen prawns. Each consignment consists of $B = 8000$ bags, each containing $M = 40$ prawns. For inspection, $b = 13$ bags are randomly selected, and $m = 5$ prawns per selected bag are tested using PCR.

For each consignment i, the contamination prevalence $p_i \sim \text{Beta}(\alpha, \beta)$ is used to simulate the number of contaminated prawns per bag:

$$X_{ij} \sim \text{Binomial}(M, p_i), \quad Y_{ij} = \mathbb{I}(X_{ij} > 0),$$

where Y_{ij} indicates whether bag j is contaminated. The true and sampled contamination counts are:

$$T_{yi} = \sum_{j=1}^{B} Y_{ij}, \quad t_{yi} = \sum_{j=1}^{b} Y_{ij}.$$

To simulate testing, b bags are randomly sampled, and pooled samples of $m = 5$ prawns per bag are tested using PCR with imperfect sensitivity $\Delta = 0.70$. Contamination detection is adjusted accordingly.

Test performance is evaluated by comparing true contamination (T_{yi}) with the observed number of positive test results (t_{yi}):

- **True Positive (TP):** $T_{yi} > 0$ and $t_{yi} > 0$
- **True Negative (TN):** $T_{yi} = 0$ and $t_{yi} = 0$
- **False Negative (FN):** $T_{yi} > 0$ and $t_{yi} = 0$

False negatives lead to **leakage**, meaning contaminated consignments pass undetected.

This simulation provides estimates of:

- **Probability of leakage**: $\Pr(T_{Xi} > 0 \text{ and } t_{yi} = 0)$
- **Expected leakage**: average number of undetected contaminated prawns per consignment

These metrics quantify the effectiveness of group testing and support biosecurity decision-making.

Installation

You can install the development version of `groupedHG` like so:

```
install.packages("devtools")
devtools::install_github("sumon148/groupedBBMH")
```