An EM algorithm for maximum likelihood estimation of the BLN probability mass function parameters

Robert Vogel

2022-08-04

Pejman has found that the Binomial Logit Normal (BLN) distribution is affective at modeling Allele Specific Expression (ASE) count data. Without loss of generality, let the alternative allele counts be a random variable such that $X \sim \text{BLN}(n, \mu, \sigma^2)$, with total counts n, mean μ, variance σ^2, and probability mass function denoted $p_X\left(x; n, \mu, \sigma^2\right)$. We assume that n is known and are interested in inferring the BLN parameters μ, σ^2 from a data set $\mathcal{D} = \{(x_i, n_i)\}_{i=1}^N$, with each $x_i \in \mathbb{Z}_{\geq 0}$, $n_i \in \mathbb{Z}_{\geq 0}$, and $x_i \leq n_i$ by maximum likelihood. Under the standard i.i.d. assumption of datum points from data set $\mathcal{D}$ the log-likelihood is defined as follows

$$\ell(\boldsymbol{\theta}; \mathbf{x}|\mathbf{n}) = \sum_{i=1}^{N} \log\left(p_X\left(x_i; n_i, \mu, \sigma^2\right)\right) \tag{1}$$

with $\boldsymbol{\theta} = [\mu, \sigma^2]$, and the MLE estimates of $\hat{\boldsymbol{\theta}}^{\text{MLE}}$ is found by the optimization

$$\hat{\boldsymbol{\theta}}^{\text{MLE}} = \underset{\boldsymbol{\theta}}{\text{argmax}} \quad \ell(\boldsymbol{\theta}; \mathbf{x}|\mathbf{n}). \tag{2}$$

A closed form analytical solution to (2) is hard or impossible on account of the BLN distribution being a compound distribution. To see this recall our formulation of the BLN probability mass function.

Definition 1. *Recall the variable and parameter definitions above. Let $S \sim \mathcal{N}\left(\mu, \sigma^2\right)$ whose probability density is denoted $f_S\left(s; \mu, \sigma^2\right)$ and $p_{X|S}\left(x; n, logit^{-1}\left(s\right)\right)$ be the probability mass function of the Binomial distribution. Given these definitions the BLN probability mass function ($p_X\left(x; n, \mu, \sigma^2\right)$) may be expressed as*

$$p_X\left(x; n, \mu, \sigma^2\right) = \int_{s=-\infty}^{\infty} p_{X|S}\left(x; n, logit^{-1}\left(s\right)\right) f_S\left(s; \mu, \sigma^2\right) ds.$$

By Def. 1 we see that the log-likelihood (1) is a sum of log BLN integrals, which precludes an easy-to-calculate closed form analytical solution for (2). We then seek to perform the optimization (2) numerically. Inspection of Def. 1 reveals that S may be interpreted as a continuous latent variable. Such interpretation is advantageous as the numerical optimization of (2) may conducted by the Expectation Maximization (EM) algorithm. In what follows I introduce EM at a high level, and derive the steps of the EM algorithm for the BLN distribution.

1 MLE inference by EM

I've used the discussion of EM in [1] to develop EM for BLN parameter inference. EM faciliates MLE based parameter inference by augmenting the data set to include the latent parameters S_i. Given this new "total data set" EM optimizes the total log-likelihood

$$\ell(\boldsymbol{\theta}; \mathbf{x}, \mathbf{S}|\mathbf{n}) = \sum_{i=1}^{N} \log\left(p_{X_i|S_i}\left(x; n_i, \operatorname{logit}^{-1}\left(S_i\right)\right)\right) + \log\left(f_{S_i}\left(s_i; \mu, \sigma^2\right)\right) \tag{3}$$

by an iterative application of the Expectation ("E-step") followed by the Maximization ("M-step") steps. There are several strategies to determine convergence, I conclude that the algorithm has converged when the difference of the data log-likelihood (1) given the j^{th} and $j^{th}+1$ parameter sets is less than a specified tolerance.

In the E step we calculate $Q\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$, a quantity that is less than or equal to the data log-likelihood (1) with equality at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(j)}$, by the expectation

$$Q\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right) = \mathbb{E}_{\mathbf{S}|\hat{\boldsymbol{\theta}}^{(j)}, \mathbf{x}, \mathbf{n}}\left[\ell(\boldsymbol{\theta}; \mathbf{x}, S|\mathbf{n})\right]. \tag{4}$$

Let's stop and appreciate what we have accomplished so far. First, by augmenting the data set with the latent variable $\mathbf{S}$ the log function of the total log-likelihood (3) is applied to the Binomial pmf and the Gaussian pdf as opposed to the integral of the two functions as in the data log-likelihood (1). Second, by calculating the expectation of the total log-likelihood we eliminate the dependency on the latent variable and are left with a function exclusively in terms of $\boldsymbol{\theta}$. Next, the M step takes advantage of these two accomplishments to establish equations that updated parameter values that are theoretically guaranteed to increase the data log-likelihood of (1).

The M step updates the parameters of the j^{th} iteration by optimizing Q, that is

$$\hat{\boldsymbol{\theta}}^{(j+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right). \tag{5}$$

Analytically, this amounts to finding an extremum of $Q\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ and determining that the extremum is a maximum. In what follows I present my calculations of the E step (4) and M step (5) for inferring the parameters of the BLN.

In (4) the expctation is taken over the the distribution function of $\mathbf{S}|\hat{\boldsymbol{\theta}}^{(j)}, \mathbf{x}, \mathbf{n}$, a distribution that we do not *a priori* know. Consequently, prior to the E step we determine this distribution.

Claim 1 (Conditional distribution of **S**)**.** *Given the problem setup and Def. 1, I denote the conditional distribution* $S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i$ *by*

$$\hat{\gamma}_i^{(j)}(s_i) := Pr\left(S_i = s_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i\right)$$

such that

$$\hat{\gamma}_i^{(j)}(s_i) = \frac{p_{X_i|S_i}\left(x_i; n_i, logit^{-1}(s_i)\right) f_{S_i}\left(s_i; \hat{\mu}^{(j)}, \widehat{\sigma^2}^{(j)}\right)}{\int_{s_i=-\infty}^{\infty} p_{X_i|S_i}\left(x_i; n_i, logit^{-1}(s_i)\right) f_{S_i}\left(s_i; \hat{\mu}^{(j)}, \widehat{\sigma^2}^{(j)}\right)}.$$

Proof. By Bayes' theorem,

$$\Pr\left(S_i = s_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i\right) = \frac{\Pr\left(X_i = x_i|\hat{\boldsymbol{\theta}}^{(j)}, n_i, s_i\right)\Pr\left(S_i = s_i|\hat{\boldsymbol{\theta}}^{(j)}, \mathbf{n}\right)}{\int_{s=-\infty}^{\infty}\Pr\left(X_i = x_i|\hat{\boldsymbol{\theta}}^{(j)}, n_i, s_i\right)\Pr\left(S_i = s_i|\hat{\boldsymbol{\theta}}^{(j)}, \mathbf{n}\right)} \tag{6}$$

where we recognize $\Pr\left(X_i = x_i|\hat{\boldsymbol{\theta}}^{(j)}, n_i, s_i\right)$ and $\Pr\left(S_i = s_i|\hat{\boldsymbol{\theta}}^{(j)}, \mathbf{n}\right)$ are the binomial and Gaussian distributions form Def. 1, respectively. □

Claim 2 (E step)**.** *Given the problem setup, Claim 1 and the total log-likelihood* (3)*, the E step* (4) *for the BLN distribution is*

$$Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right) = -\frac{N}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[(s_i-\mu)^2\right]$$

where

$$\mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[(s_i-\mu)^2\right] = \int_{s_i=-\infty}^{\infty}(s_i-\mu)^2\,\hat{\gamma}_i^{(j)}(s_i)\,ds_i.$$

and $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ *denotes the terms of* $Q\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ *that depend on the elements of* $\boldsymbol{\theta}$*.*

Proof. First, recall by (4) the E step consists of finding $Q\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ by the expectation of the total log-likelihood over the distribution function for $\mathbf{S}|\hat{\boldsymbol{\theta}}^{(j)}, \mathbf{x}, \mathbf{n}$. Then by substitution of the total log-likelihood (3) into (4) and application of the i.i.d. assumption

$$Q\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right) = \sum_{i=1}^{N} \underbrace{\mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[\log\left(p_{X_i|S_i}\left(x; n_i, \text{logit}^{-1}\left(S_i\right)\right)\right)\right]}_{A_i} + \underbrace{\mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[\log\left(f_{S_i}\left(s_i; \mu, \sigma^2\right)\right)\right]}_{B_i}. \quad (7)$$

Each A_i in (7) does not depend on either μ or σ^2 and consequently does not contribute to $\nabla_{\boldsymbol{\theta}} Q\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$. Taking this into account, we denote the effective A_i term as

$$a_i = 0 \quad (8)$$

Now if we expand B_i,

$$B_i = -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[\left(s_i - \mu\right)^2\right]. \quad (9)$$

and subsequently recognize that $\log\left(2\pi\sigma^2\right)$ the $\log\left(2\pi\right)$ term will be eliminated after differentiation the effective B_i term is

$$b_i = -\frac{1}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[\left(s_i - \mu\right)^2\right]. \quad (10)$$

Then we define the effective Q as

$$Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right) := \sum_{i=1}^{N} a_i + b_i \quad (11)$$

which by substitution of (8) and (10) into (11) completes the proof. □

Next we calculate the parameter updates of the M step by substituting $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ of Claim 2 for Q in (5). I perform a two-step procedure for maximizing Q, find a stationary point of $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ and then show that the stationary point is a maximum (not a minimum). I begin with finding a stationary point.

Claim 3 (A single stationary point)**.** *Given the problem setup, Claim 2, and* (5) *the stationary point of* $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ *is denoted* (μ_0, σ_0^2) *such that*

$$\mu_0 = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[s_i\right]$$

and

$$\sigma_0^2 = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[\left(s_i - \hat{\mu}^{(j+1)}\right)^2\right].$$

Proof. Recall that stationary points of a function are those in which the gradient is zero. Let us then take the gradient of $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ defined in Claim 2

$$\partial_\mu Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right) = -\frac{N\mu}{2\sigma^2} + \frac{1}{2\sigma^2}\sum_{i=1}^{N} \mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[s_i\right] \tag{12}$$

and

$$\partial_{\sigma^2} Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right) = -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{N} \mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[(s_i-\mu)^2\right]. \tag{13}$$

Setting (12) and (13) to zero and solving the system of equations completes the proof. □

Claim 3 shows that Q has a single stationary point (μ_0, σ_0^2). To determine whether this stationary point is a maximum I ask whether $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ is a concave function. To determine concavity I analyze the matrix of second derivatives of $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$, the so-called Hessian matrix $\mathbf{H}$ with $H_{ij} := \partial_{\theta_l}\partial_{\theta_k} Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$. By definition of $\mathbf{H}$ it is immediately clear that $\mathbf{H}$ is a square $p \times p$ symmetric matrix with Eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$. The trace of a $p \times p$ matrix is

$$\operatorname{tr}(\mathbf{H}) = \sum_{i=1}^{p} \lambda_i \tag{14}$$

and determinant

$$\det(\mathbf{H}) = \prod_{i=1}^{p} \lambda_i. \tag{15}$$

We will use properties (14) and (15) in what follows.

Claim 4 (The Hessian matrix)**.** *The Hessian matrix of* $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ *from Claim 2 has elements*

$$\begin{aligned}
H_{11} &= -\frac{N}{\sigma^2} \\
H_{12} &= -\frac{1}{(\sigma^2)^2}\mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[s_i-\mu\right] \\
H_{22} &= \frac{N}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^{N} \mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[(s_i-\mu)^2\right]
\end{aligned}$$

Proof. Let $H_{11} = \partial_\mu^2 Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$, $H_{12} = \partial_{\sigma^2}\partial_\mu Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$, and $H_{22} = \partial_{\sigma^2}^2 Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$. Differentiating $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ for each H_{ij} completes the proof. □

Claim 5 (Concave or convex)**.** *There exists distinct sets of 2-tuples* (μ, σ^2) *for which* $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ *of Claim 2 is concave or convex.*

Proof. Recall that if $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ is concave or convex the Hessian matrix of $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ will be negative or postive semi-definite for any value of μ and σ^2, respectively. Recall that each Eigenvalue of a negative or positive semi-definite matrix satisfies $\lambda_i \leq 0$ or $\lambda_i \geq 0$, respectively. As $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ depends on two parameters, the Hessian is a 2×2 matrix, and by equations (14) and (15)

$$\operatorname{tr}(\mathbf{H}) = \lambda_1 + \lambda_2 \leq 0 \tag{16}$$

$$\det(\mathbf{H}) = \lambda_1 \lambda_2 \geq 0 \tag{17}$$

if $\mathbf{H}$ is negative semi-definite. If positive semi-definite both the trace and determinants are greater than 0.

By the definition of the matrix trace and Claim 4

$$\operatorname{tr}(\mathbf{H}) = -\frac{N}{\sigma^2} \underbrace{\left(1 - \frac{1}{2\sigma^2}\right)}_{A} - \underbrace{\frac{1}{(\sigma^2)^3} \sum_{i=1}^{N} \mathbb{E}_{S_i|\hat{\boldsymbol{\theta}}^{(j)}, x_i, n_i}\left[(s_i - \mu)^2\right]}_{B}. \tag{18}$$

By definition of the variance we know that B of (18) is zero or positive. However, A of (18) is 0 or positive for $\sigma^2 \geq 1/2$ and negative for $\sigma^2 < 1/2$. Under the independent ranges of A and B it is feasible that $|A| < B$ and $A < 0$, and consequently the trace of (18) is positive. It remains to be determined if such values are in the range of (18). $\square$

Claim 5 is troubling, as we cannot easily establish whether the stationary point of Claim 3 maximizes or minimizes $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$. To make progress, I next assess whether the stationary points are a local maximum or minimum.

Claim 6 (Local maximum)**.** *The stationary point of Claim 3 is a local maximum.*

Proof. To establish whether the stationary point of Claim 3 is a local maximum, consider the second-order Taylor expansion of $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)$ about $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0^2)$

$$Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right) = Q_{eff}\left(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}^{(j)}\right) + \nabla_{\boldsymbol{\theta}} Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \delta\boldsymbol{\theta} + \frac{1}{2}\delta\boldsymbol{\theta}^T \mathbf{H}(\boldsymbol{\theta} = \boldsymbol{\theta}_0)\,\delta\boldsymbol{\theta} \tag{19}$$

with $\delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0$. As the gradient evaluated at a stationary point is zero, I eliminate the first-order term and rearrange to see that

$$Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right) - Q_{eff}\left(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}^{(j)}\right) = \delta\boldsymbol{\theta}^T \mathbf{H}(\boldsymbol{\theta} = \boldsymbol{\theta}_0)\,\delta\boldsymbol{\theta}. \tag{20}$$

If $Q_{eff}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(j)}\right) - Q_{eff}\left(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}^{(j)}\right) < 0$ then Q_{eff} evaluated at the stationary point is a local maximum. By (20) when the LHS is less than zero, so is the RHS. Recall that a square matrix $\mathbf{A}$ in which $x^T A x < 0$ is said to be negative definite. Therefore, if the $\mathbf{H}\left(\boldsymbol{\theta} = \boldsymbol{\theta}_0\right)$ is negative definite then $Q_{eff}\left(\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}^{(j)}\right)$ is a local maximum.

First, the elements of the Hessian matrix evaluated at the stationary point is determined by substituting Claim 3 into Claim 4

$$H_{11} = -\frac{N}{\sigma_0^2}, \tag{21}$$

$$H_{12} = 0, \tag{22}$$

$$H_{22} = -\frac{N}{(\sigma^2)^2} \tag{23}$$

As (21) and (23) are less than zero, the trace less than zero – the first criterion of negative definite matrices is satisfied. Next, by symmetry of $\mathbf{H}$ and substituting (21), (22), and (23) into the determinant,

$$\det\left(\mathbf{H}(\boldsymbol{\theta} = \boldsymbol{\theta}_0)\right) = H_{11}H_{22} - H_{12}^2 = H_{11}H_{22} > 0. \tag{24}$$

By (24) the second criterion of negative definite matrices is satisfied, and therefore $\mathbf{H}(\boldsymbol{\theta} = \boldsymbol{\theta}_0)$ is negative definite. □

Claim 7 (M step)**.** *If $\hat{\boldsymbol{\theta}}^{(j)}$ is sufficiently close to $\hat{\boldsymbol{\theta}}^{(j+1)}$ that the Taylor expansion of Claim 6 is a good approximation, then the stationary points of Claim 3 are sufficient for the M step of the EM algorithm.*

Proof. By Claim 6 the stationary points of Claim 3 are a local maximum of Q and satisfy the criterion of (5). □

The claims above are describe the rationale of my implementation of EM for inferring the parameters to the BLN distribution from data. In what follows, we test my implementation of this algorithm using the Python programming language and simulation data.

2 Application

I conducted simulation experiments to validate my EM implementation. In these simulations I determine whether the algorithm may infer diverse BLN parameter sets for a diverse number of samples. That is, for each parameter set consisting of $\mu \in \{-2, -1, 0, 1, 2\}$, $\sigma^2 \in \{0.2, 0.5, 0.75, 1, 1.5, 2, 2.5\}$, and total ASE counts of 100, I simulate $N_{\text{samples}} = \{10, 17, 31, 56, 100, 177, 316, 562, 1000\}$ observations and infer parameters 10 times. I then report the mean ± standard deviation over replicate sets for each parameter combination as a function of N_{samples}.

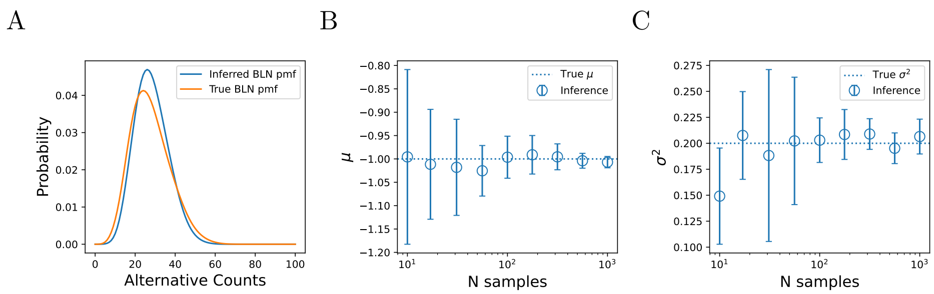

Figure 1: (A) The BLN pmf of alternative counts given the true ($\mu = -1, \sigma^2 = 0.2$) and inferred parameters for $N_{\text{samples}} = 10$ of an arbtrary replicate simulation / inference experiment. The mean $\pm$ standard deviation of the BLN mean parameter μ (B) and variance parameter σ^2 (C) for different number of samples.

Figure 1 shows the results of a single parameter set. First, by visual inspection of Figure 1A, we see that the BLN pmf of alternative allele counts of the true and inferred parameter sets is similar. Quantitative analysis corroborates this qualitative assessment for the mean Figure 1B and variance Figure 1C. Moreover, we see that the standard deviation of inferred parameters over replicate experiments decreases with the number of samples simulated. Together, these results suggest that the EM algorithm and its implementation with the Python programming language is working as expected. Figure 2, Figure 3, and Figure 4 display results for $\mu = -1, 0, 2$, respecitvely. In each μ figure panel I perform tests for $\sigma^2 = 0.5, 1, 1.5, 2.5$.

3 Conclusion

The implementation of EM for BLN parameter inference is sufficient for simulation data.

References

[1] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*, volume 2. Springer Series in Statistics New York, 2009.

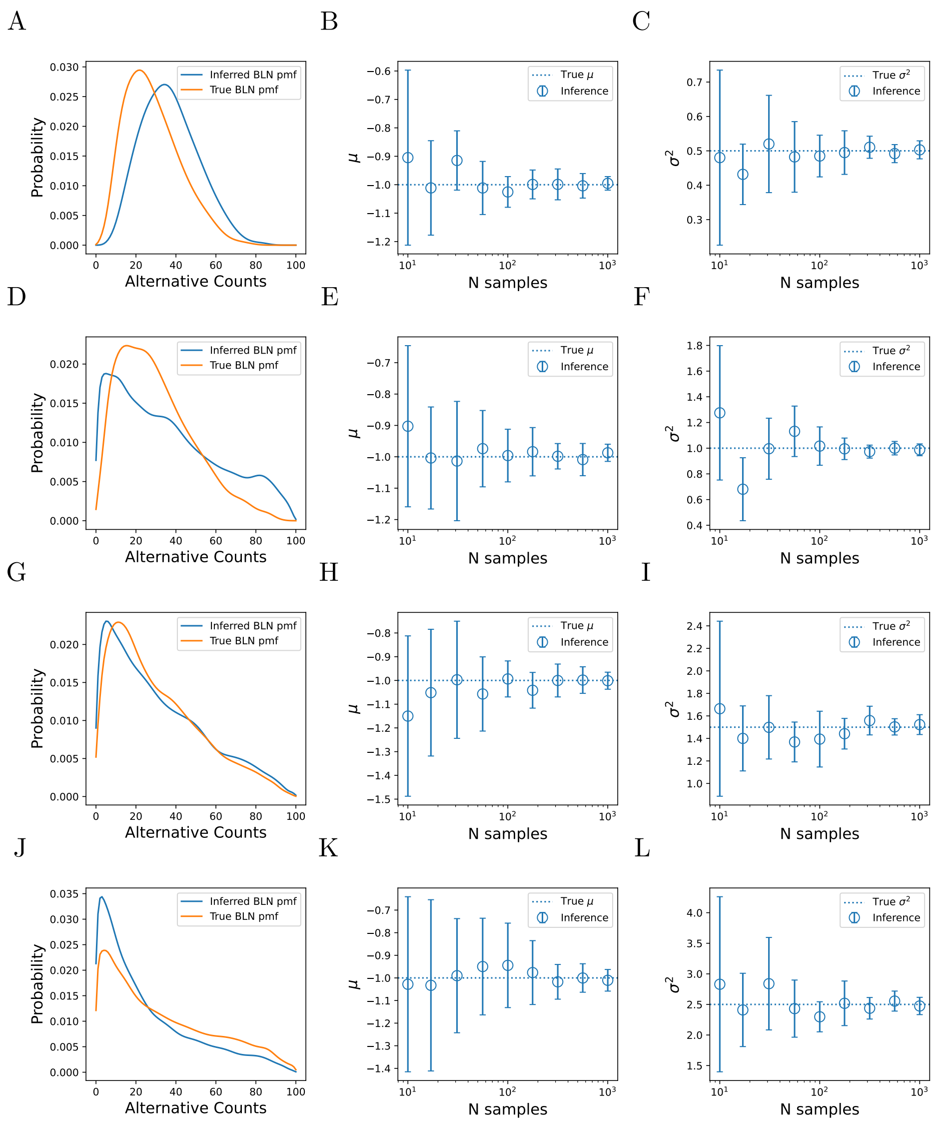

Figure 2: (A,D,G,J) The BLN pmf of alternative counts given the true and inferred parameters for $N_{\text{samples}} = 10$ of an arbtrary replicate simulation / inference experiment. The mean $\pm$ standard deviation of the BLN mean parameter μ (B,E,H,K) and variance parameter σ^2 (C,F,I,L) for different number of samples. Here, each row displays results for BLN variances = 0.5, 1, 1.5, and 2.5.

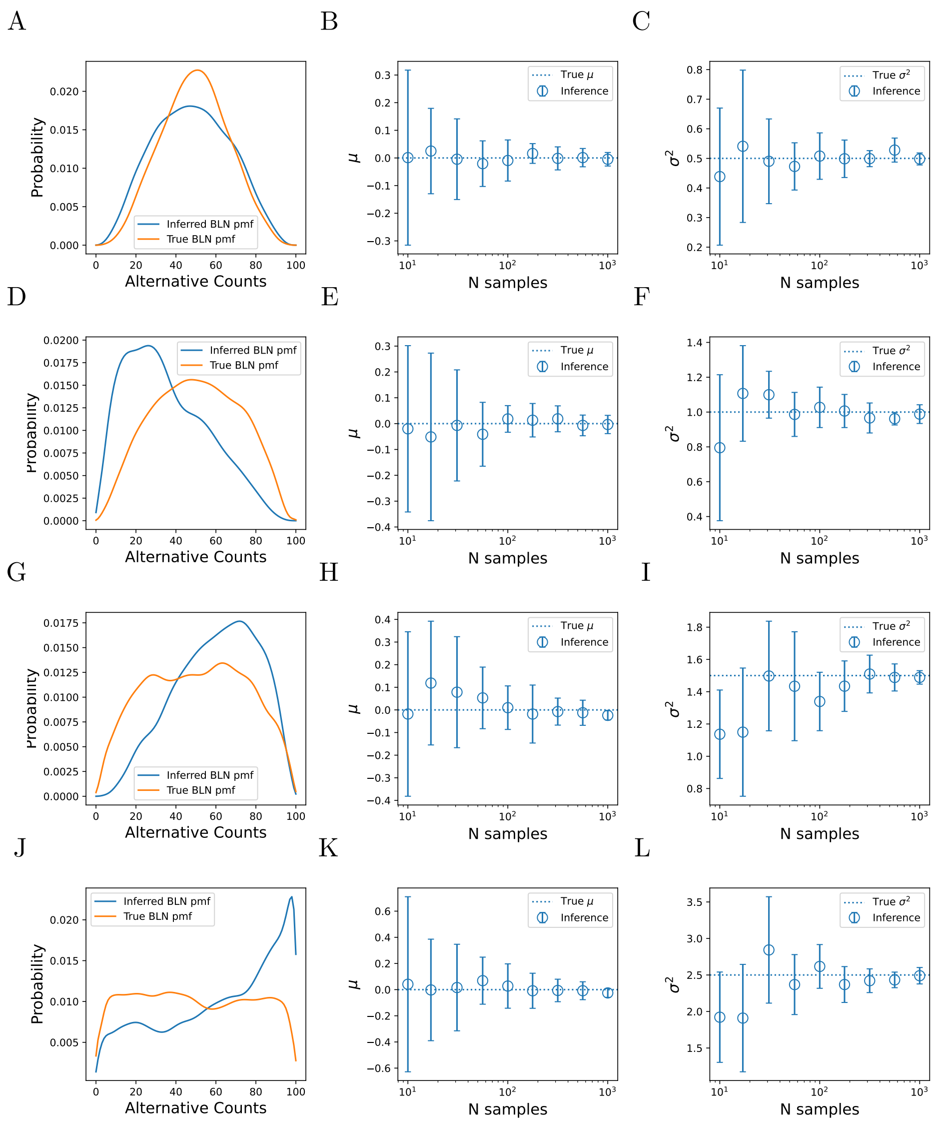

Figure 3: (A,D,G,J) The BLN pmf of alternative counts given the true ($\mu = 0$) and inferred parameters for $N_{\text{samples}} = 10$ of an arbtrary replicate simulation / inference experiment. The mean $\pm$ standard deviation of the BLN mean parameter μ (B,E,H,K) and variance parameter σ^2 (C,F,I,L) for different number of samples. Here, each row displays results for BLN variances = 0.5, 1, 1.5, and 2.5.

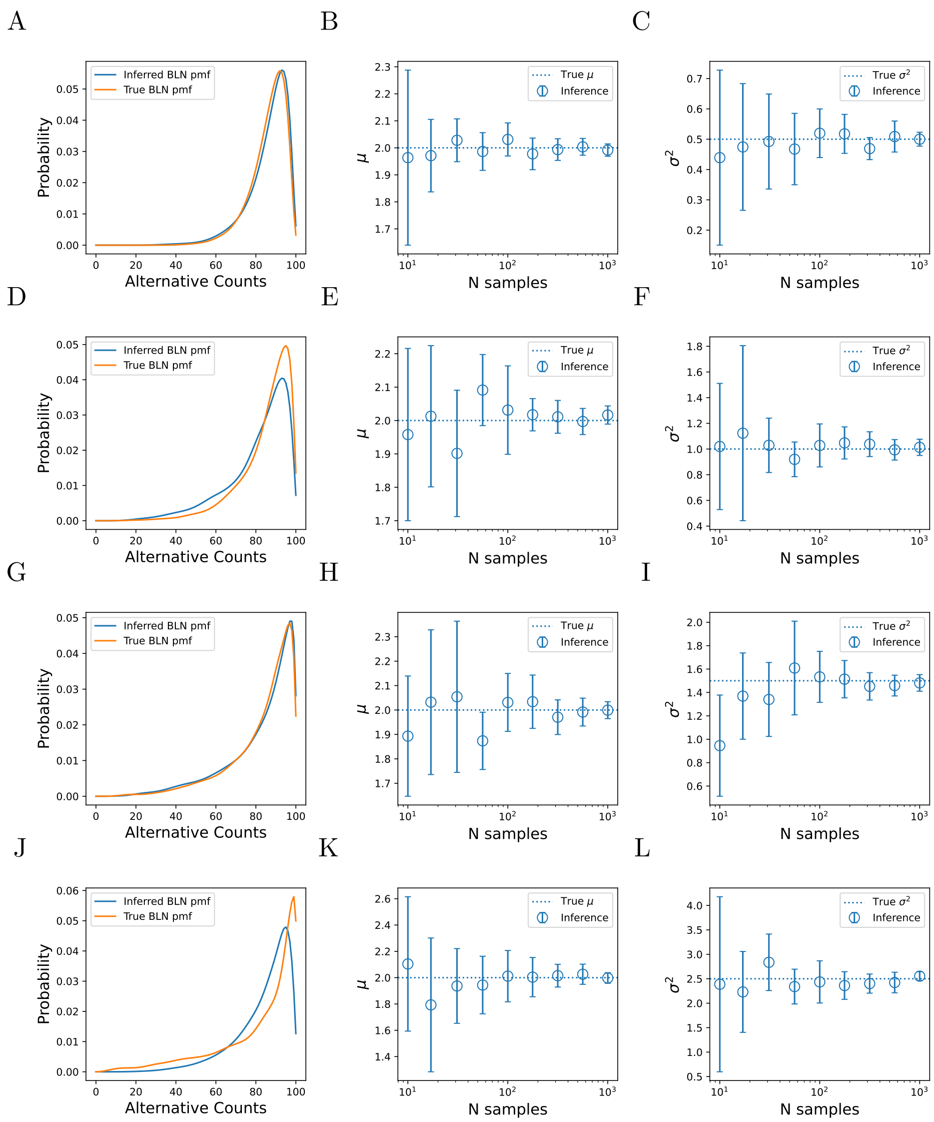

Figure 4: (A,D,G,J) The BLN pmf of alternative counts given the true ($\mu = 2$) and inferred parameters for $N_{\text{samples}} = 10$ of an arbtrary replicate simulation / inference experiment. The mean $\pm$ standard deviation of the BLN mean parameter μ (B,E,H,K) and variance parameter σ^2 (C,F,I,L) for different number of samples. Here, each row displays results for BLN variances = 0.5, 1, 1.5, and 2.5.