Technical Report: Fast Updating for Bayesian Joint Hierarchical Model for Prediction of Latent Health States

R Yates Coley, Aaron Fisher, Scott L Zeger

August 24, 2015

Abstract

This technical report presents an importance sampling algorithm for rapidly obtaining updated individualized predictions for the Bayesian joint model proposed in Coley at al. (2015). Algorithm details are given and performance is assessed. Data and code are available at `http://github.com/rycoley/XXX`.

1 Introduction

Coley et al. (2015) presents a Bayesian joint hierarchical model for predicting latent health states from longitudinal clinical measurements. Model development was motivated by an application to active surveillance of low risk prostate cancer. Previous joint modeling approaches (e.g., [6] are unsuitable for this context as they do not accommodate measurement error– specifically, cancer state determinations based on biopsied tissue are prone to misclassification– nor do they allow for observations to be missing not at random [7]. The proposed model addresses these limitations, enabling estimation of an individual's underlying cancer state. These individualization predictions can then be communicated to clinicians and patients to inform decision making.

For this prediction model to be most useful in a clinical setting, however, it is necessary to be able to update posterior estimates quickly to incorporate new biopsy or PSA results during a patient's visit. This precludes re-running the full MCMC to obtain updated posteriors of patient-specific variables. Instead, we use importance sampling [1] to obtain rapid prediction updates. Using posterior estimates obtained from current data, the proposed importance sampling algorithm updates these estimates for either a new patient or new data on an existing patient in a matter of seconds.

Many options exist for performing online updating of Bayesian models. We chose to use an importance sampling approach because...

In this this technical report, we first provide an overview of the latent class prediction model before detailing the importance sampling algorithm we have developed for enabling real-time

updates. Next, we will apply the proposed algorithm to simulated data and compare predictions to those obtained by full MCMC runs. Finally, we close with a discussion of future research.

2 Bayesian Joint Hierarchical Latent Class Model

In this section, we briefly summarize the joint model proposed in Coley et al. (2015) for predicting latent cancer state for men participating in Active Surveillance for low risk prostate disease. Predictions are made by incorporating information from repeated prostate specific antigen (PSA) and biopsy measurements for all individuals, as well as true cancer state information observed in a subset of the cohort. In this technical report, we will focus on the model that assumes biopsy and latent class observation are missing at random, that is, not associated with the latent state after conditioning on observed clinical variables. For more model details and application background, please see Coley et al. (2015).

Let η_i indicate the true cancer state for individual i, $i = 1, \ldots, n$, defined using the the Gleason score [3, 4] that would be assigned if his entire prostate were to be removed and pathologic analysis performed: $\eta_i = 0$ if Gleason ≤ 6 or *indolent* and $\eta_i = 1$ if Gleason ≥ 7 or *aggressive*. True cancer state then follows a Bernoulli distribution– $\eta_i \sim Bernoulli(\rho)$– where we assume here, for simplicity, a shared underlying probability of aggressive cancer, ρ. This true cancer state is observed in a subset of patients who choose to have their prostate surgically removed; as such, η_i is a partially latent variable.

Next, define the following mixed effects model [5] for an individual's PSA over time where mean effects for predictors are allowed to vary across groups defined by the partially latent true cancer state:

$$\left[Y_{im} | \eta_i = k, \mathbf{X}_{im}, \mathbf{Z}_{im} \right] = \mathbf{X}_{im}\boldsymbol{\beta} + \mathbf{Z}_{im}\mathbf{b}_i + \epsilon_{im}$$

where Y_{im} is the observed (log-transformed) PSA, $\mathbf{X}_{im}$ and $\mathbf{Z}_{im}$ are vectors of covariates for individual i's mth PSA measurement, $\boldsymbol{\beta}$ is a parameter vector for fixed effects, and $\mathbf{b}_i$ is the patient-specific vector of random effects. Following the specification of a Bayesian mixed effects model presented by Gelman and Hill (2006), unscaled random effects are centered at the mean effects for each latent class k, $\boldsymbol{\mu}_k$:

$$\left[\check{\mathbf{b}}_i | \eta_i = k \right] \sim MVN(\boldsymbol{\mu}_k, \Sigma_k), \quad k = 0, 1$$

where Σ_k is a covariance matrix that allows for correlation between random effects. Random

effects are then scaled with parameter vector $\boldsymbol{\xi}$: $\mathbf{b}_i = diag(\check{\mathbf{b}}_i \boldsymbol{\xi}^T)$. Lastly, residuals ϵ_{im} are assumed to follow a normal distribution with mean 0 and variance σ^2.

Finally, let (B_{ij}, R_{ij}) denote binary variables for individual i in annual interval j indicating, respectively, whether a biopsy was performed and, when it was, if *reclassification* was observed, i.e. a determination of Gleason ≥ 7 made. (B_{ij}, R_{ij}) is defined for $j = 1, \ldots, J_i$ where J_i is the time of reclassification or censoring for patient i. For each time interval with a biopsy, $B_{ij} = 1$, we use logistic regression to model its result conditional on true cancer state:

$$P(R_{ij} = 1 | B_{ij} = 1, \eta_i = k, \mathbf{V}_{ij}, \boldsymbol{\gamma}) = \text{logit}^{-1}\big(\mathbf{V}_{ij}(k)\boldsymbol{\gamma}\big) \tag{1}$$

where $\mathbf{V}_{ij}(k)$ is a matrix of time-varying predictors including η_i and $\boldsymbol{\gamma}$ is a parameter vector.

We then define the joint probability of the parameters given data and unobserved latent variables:

$$\begin{aligned} &L\Big(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\gamma} \mid \big(\eta_i, (\mathbf{Y}_i, \mathbf{b}_i, \underline{\mathbf{X}_i}, \underline{\mathbf{Z}_i}), (\mathbf{B}_i, \mathbf{R}_i, \underline{\mathbf{V}_i})\big), i = 1, ..., n\Big) \\ &= \prod_{i=1}^{n} \rho^{\eta_i} (1-\rho)^{1-\eta_i} f(\mathbf{Y}_i | \eta_i, \underline{\mathbf{X}_i}, \underline{\mathbf{Z}_i}, \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2)\, g(\mathbf{b}_i | \boldsymbol{\mu}_{\eta_i}, \Sigma_{\eta_i}) \\ &\qquad \prod_{j=1}^{J_i} \big(P(R_{ij} = 1 | \eta_i, \mathbf{V}_{ij}, \boldsymbol{\gamma})^{R_{ij}} P(R_{ij} = 0 | \eta_i, \mathbf{V}_{ij}, \boldsymbol{\gamma})^{1-R_{ij}}\big)^{B_{ij}} \end{aligned} \tag{2}$$

where f and g are multivariate normal densities for the vector of log-transformed PSAs $\mathbf{Y}_i$, and random effects $\mathbf{b}_i$, respectively, each with mean and covariance as defined above, given covariance matrices $\underline{\mathbf{X}_i} = [X_{i1}, \ldots, X_{iM_i}]$ and $\underline{\mathbf{Z}_i} = [Z_{i1}, \ldots, Z_{iM_i}]$. $\mathbf{B}_i, \mathbf{R}_i$ represent vectors for biopsy and reclassification for individual i with associated covariance matrix $\underline{\mathbf{V}_i}$.

Bayesian methods are used to identify posteriors for model parameters and latent variables. Discussion of prior specification can be found in Coley et al. (2015). After specifying priors for all model parameters, the join posterior distribution of the parameter and latent variables is:

$$\begin{aligned} &p\Big(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\gamma}, (\eta_i, \mathbf{b}_i) \mid \big((\mathbf{Y}_i, \underline{\mathbf{X}_i}, \underline{\mathbf{Z}_i}), (\mathbf{B}_i, \mathbf{R}_i, \underline{\mathbf{V}_i})\big); \boldsymbol{\Theta}\Big) \\ &\quad \propto L\Big(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\gamma} \mid \big(\eta_i, (\mathbf{Y}_i, \mathbf{b}_i, \underline{\mathbf{X}_i}, \underline{\mathbf{Z}_i}, (\mathbf{B}_i, \mathbf{R}_i, \underline{\mathbf{V}_i})\big)\Big) \times \boldsymbol{\pi}\big(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\gamma} | \boldsymbol{\Theta}\big) \end{aligned} \tag{3}$$

where $\boldsymbol{\pi}(\cdot|\boldsymbol{\Theta})$ denotes the joint prior density for model parameters with hyperparameters $\boldsymbol{\Theta}$ with indexing on i, j, and k suppressed for clarity in presentation.

3 Importance Sampling Algorithm for Fast Prediction Updates

Next, we detail an importance sampling algorithm that enables rapid updates of joint latent class model predictions. To simplify presentation, we abbreviate notation for the joint posterior given above in Equation 3:

$$p(\theta, b_{1:n}|y_{1:n}) \propto \prod_{i=1}^{n}[f(y_i|b_i,\theta)g(b_i|\theta)]\pi(\theta) \tag{4}$$

where y_i is the vector of clinical measurements (PSA and biopsy) for patient i, $y_{1:n}$ is the list of measurements for the first n patients, b_i is a vector of latent variables (latent class and random effects) for patient i, $b_{1:n}$ is a list of latent variables for the first n patients, θ contains the population-level parameters, π is the prior for θ, and f and g are multivariate distributions coming from the model likelihood in Equation 2.

After posterior samples from the joint model are obtained for current data, importance sampling to update these estimates given new data requires three steps: first, generating proposal values for the latent variables to be updated, second, calculating weights for proposed values, and, third, weighting proposed values to estimate an updated posterior. We first illustrate how this process can be used to quickly estimate latent variables for a new patient before showing how similar calculations can be done to incorporate newly measured data on existing patients in real-time.

For a new patient (indexed by $i = n + 1$), obtaining posterior predictions of latent variables requires calculating expectations with respect to the posterior distribution based on all $n + 1$ patients (i.e. $p(\theta, b_{1:(n+1)}|y_{1:(n+1)})$). While we cannot immediately draw from this distribution, we can evaluate a function that is proportional to its density (Equation 4). The posterior distribution based on the first n patients provides an appropriate proposal distribution (q) from which to generate candidate values of $(\theta, b_{1:(n+1)})$:

$$q(\theta, b_{1:(n+1)}) \quad := \quad g(b_{n+1}|\theta)p(\theta, b_{1:n}|y_{1:n})$$

This approach is analogous to a one-step particle filter [1] and, practically, consists of taking J draws of θ and $b_{1:n}$ from the previously fitted posterior in Equation 4. Then, conditional on θ, we draw b_{n+1} from the distribution g. We index each of the resulting draws as $(\theta^{(j)}, b^{(j)}_{1:(n+1)})$,

with $j = 1, \ldots, J$. The importance weights w_j are then proportional to

$$\begin{aligned} w^{(j)} &\propto \frac{p(\theta^{(j)}, b^{(j)}_{1:(n+1)}|y_{1:(n+1)})}{q(\theta^{(j)}, b^{(j)}_{1:(n+1)})} \\ &\propto \frac{\prod_{i=1}^{n+1}[f(y_i|b_i^{(j)}, \theta^{(j)})g(b_i^{(j)}|\theta^{(j)})]\pi(\theta^{(j)})}{g(b_{n+1}^{(j)}|\theta^{(j)})\prod_{i=1}^{n}[f(y_i|b_i^{(j)}, \theta^{(j)})g(b_i|\theta^{(j)})]\pi(\theta^{(j)})} \\ &= f(y_i|b_i^{(j)}, \theta^{(j)}) \end{aligned} \tag{5}$$

The final weights $w^{(j)}$ are standardized to sum to 1. The new posterior for $(\theta, b_{1:(n+1)})$ can then be represented as the mixture distribution satisfying $P(\theta = \theta^{(j)}, b_{1:(n+1)} = b^{(j)}_{1:(n+1)}) = w^{(j)}$. A posterior mean for $b_{(n+1)}$ can be calculated as $\sum_{j=1}^{J} w^{(j)} b^{(j)}_{(n+1)}$. The unstandardized weights can also be used in a rejection sampling procedure, although we found this approach to be less computationally efficient than importance sampling for our scenario.

For a patient k with existing data, where we already have a posterior sample for their latent variable values, we instead use this posterior as our proposal distribution $q(\theta^{(j)}, b^{(j)}_{1:n})$, with $i \leq n$. Let $y^{k+}_{1:n}$ refer to the data set after incorporating new data on patient k, where $y_i^+ = y_i$ if $k \neq i$. The importance weights in Equation 5 then simplify to

$$\begin{aligned} w^{(j)} &\propto \frac{p(\theta^{(j)}, b^{(j)}_{1:n}|y^+_{1:n})}{q(\theta^{(j)}, b^{(j)}_{1:n})} \\ &\propto \frac{\prod_{i=1}^{n}[f(y_i^+|b_i^{(j)}, \theta^{(j)})g(b_i^{(j)}|\theta^{(j)})]\pi(\theta^{(j)})}{\prod_{i=1}^{n}[f(y_i|b_i^{(j)}, \theta^{(j)})g(b_i^{(j)}|\theta^{(j)})]\pi(\theta^{(j)})} \\ &= \frac{f(y_k^+|b_k^{(j)}, \theta^{(j)})}{f(y_k|b_k^{(j)}, \theta^{(j)})} \end{aligned}$$

Let L_k denote that number of measurements for which we've previously fit a posterior for b_k, and let N_k denote the number of new measurements we wish to incorporate into this posterior. Then, y_k^+ can be expressed as the vector $y_k^+ = (y_{k[1]}, y_{k[2]}, \ldots y_{k[L_k]}, y^+_{k[L_k+1]}, \ldots y^+_{k[L_k+N_k]})$, where $y^+_{k[l]}$ is the l^{th} measurement from patient k. If the repeated measures for each patient are independent conditional on b_i, as is the case in the proposed model, then the above ratio reduces to

$$
\begin{aligned}
w^{(j)} &\propto \frac{\prod_{l=1}^{L_k+N_k} f(y^{+}_{k[l]}|b_k^{(j)}, \theta^{(j)})}{\prod_{l=1}^{L_i} f(y_{k[l]}|b_k^{(j)}, \theta^{(j)})} \\
&= \prod_{l=L_k+1}^{L_k+N_k} f(y^{+}_{k[l]}|b_k^{(j)}, \theta^{(j)})
\end{aligned}
$$

We then proceed as above to get a re-weighted posterior for the latent variables of patient k.

For implementation in clinical practice, proposals for new patients can be re-generated prior to actually observing new data, so that only weight calculating and re-weighting of the proposal distribution needs to be done in real-time. Then, predictions for each patient can be obtained in a matter of seconds. By random chance, some patients will have data such that very few of the pre-generated proposed latent values will receive high weights; this can cause instability in posterior means. However, such scenarios can be detected by monitoring the effective size of the posterior sample, also known as the effective number of particles. When this number drops below a pre-specified threshold (e.g., 500), the procedure can be repeated with a larger set of pre-generated proposals.

4 Application

We assessed performance of the proposed importance sampling approach in a simulated dataset. We compared the posterior probability of latent class membership (specifically, $P(\eta_i = 1)$) obtained from MCMC performed in `RJAGS` to the predictions obtained by the proposed importance sampling algorithm for each patient for whom true state was not observed. Agreement was examined for the new patient scenario, that is, proposal values for an individual's latent state were generated from population-level parameters and all his available data was used to calculate sampling weights. Code for simulating data and obtaining estimates is available at `http://github.com/rycoley/XXX`.

Results of this comparison are shown in Figure 1. Agreement between methods for newly introduced patients is strong; correlation between posterior probabilities from `JAGS` and the importance sampler is 0.XX. These findings indicate that the proposed importance sampling algorithm is an appropriate substitute for full MCMC runs in order to provide real-time updates in a clinical setting.

Effective sample size...

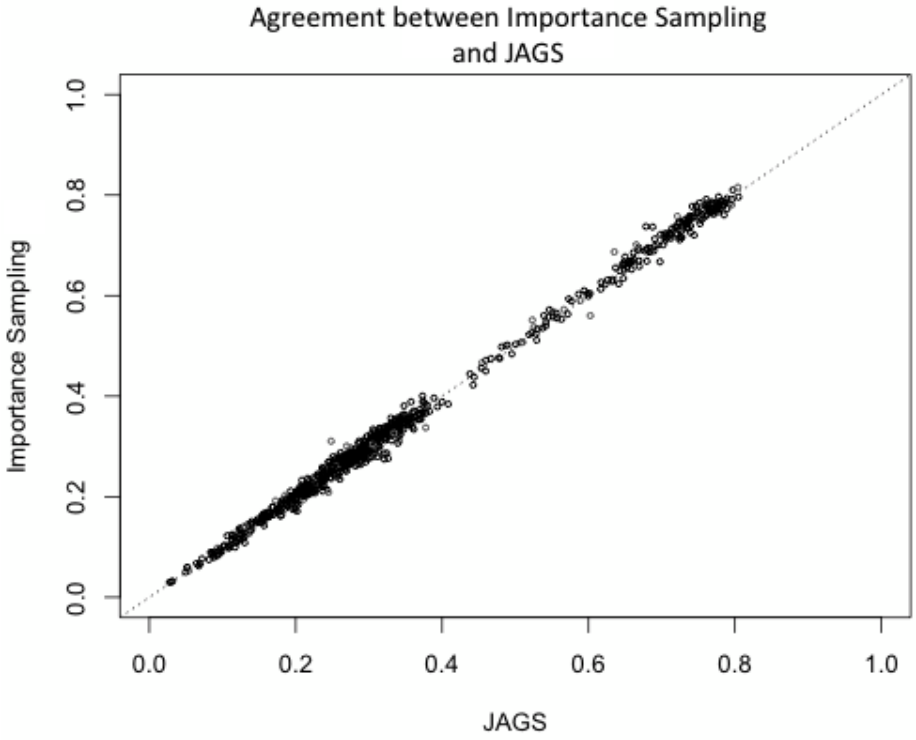

Figure 1: Agreement between importance sampling and JAGS posterior predictions of aggressive prostate cancer state for a new patient. (Dashed line indicates the axis of equality, i.e., perfect agreement.)

5 Conclusion

The joint model of Coley et al. (2015) is among a growing number of statistical models for making individualized health predictions and recommendations. Development of such *precision medicine* methods must occur within a framework for clinical implementation. Specifically, concerns about convenience, security, and effective communication must be addressed alongside statistical considerations. In this technical report, we have presented an importance sampling algorithm for obtaining fast predictions of latent health state based on the joint modeling framework of Coley et al. (2015). This approach informs decision-making by enabling doctors and patients to access updated predictions in real-time in a clinical setting.

References

[1] Christopher M Bishop et al. *Pattern recognition and machine learning*, volume 4. springer New York, 2006.

[2] Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, 2006.

[3] D.F. Gleason. The Veteran's Administration Cooperative Urologic Research Group: Histologic grading and clinical staging of prostatic carcinoma. In M. Tannenbaum, editor, *Urologic Pathology: The Prostate*, pages 171–198. Lea and Febiger, Philadelphia, 1977.

[4] Donald F Gleason. Histologic grading of prostate cancer: a perspective. *Human pathology*, 23(3):273–279, 1992.

[5] Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

[6] Haiqun Lin, Bruce W Turnbull, Charles E McCulloch, and Elizabeth H Slate. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association*, 97(457):53–65, 2002.

[7] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2014.