Modelling Stochastic Order in the Analysis of Receiver Operating Characteristic Data: Bayesian Non-parametric Approaches

Yu Zhu

Department of Statistical Science

University of California, Santa Cruz

Reference:

- Hanson, T.E., Kottas, A. and Branscum, A.J. (2008), Modelling stochastic order in the analysis of receiver operating characteristic data: Bayesian non-parametric approaches. Journal of the Royal Statistical Society: Series C (Applied Statistics), 57: 207-225

Outline:

- Background on ROC curves
- Bayesian Nonparametric Approaches
  - Motivations
  - Order Constraints
  - Dirichlet process mixtures (DPM)
  - Mixtures of finite Polya trees (MPT)
- Prior Elicitation
- Model Comparison
- Simulation Study
  - Implement DPM
- Case Study: Serology Score Analysis
  - Compare between DPM and MPT

ROC Curve

- For some continuous test, we define the true positive response probability (TPF) and the false positive response probability (FPF) under some cut-off value k as below:

$$TPF(k) = Pr(Y \geq k | D = 1) = Pr(Y_1 \geq k) = 1 - F_1(k)$$
$$FPF(k) = Pr(Y \geq k | D = 0) = Pr(Y_0 \geq k) = 1 - F_0(k)$$

- ROC curve represents the plot $\{1 - F_0(k), 1 - F_1(k)\}$ for all cut-off values k.
  - Denote $u = FPF(k) = 1 - F_0(k)$, ROC curve is given as:

$$ROC(u) = 1 - F_1(F_0^{-1}(1-u)), u \in [0,1]$$

- Area under the ROC curve (AUC):
  - One of the information summarization measures for ROC curve
  - The probability that the test outcome for a randomly chosen diseased subject exceeds the one exhibited by a randomly selected non-diseased individual

$$AUC = \int_0^1 ROC(u)du$$

  - AUC = 0.5 indicates a non-informative test

- A nature consequence of the stochastic order constraint:

$F_0 > F_1$ if and only if $ROC(u) > u$, $\implies$ $AUC > 0.5$

Bayesian Nonparametric Approaches

- Motivations for non-parametric models:
  - The distributions F_0 and F_1 often exhibit non-standard features such as multimodality and skewness that parametric models are not as flexible to capture
  - Non-parametric models can handle some other non-standard features that aren't known in advance

- Two non-parametric prior models are proposed incorporating the stochastic order constraint for F_0 and F_1 ,$F_1(t) \leq F_0(t)$ for all $t \in R$ (consider $F(x|\theta)$)
  - Bayesian inference avoids constrained optimization since a prior restriction $\{(\theta_1, \theta_2): \theta_1 < \theta_2\}$ would imply that stochastic order is retained a posteriori

- Dirichlet process mixtures:
  - DPMs generalize finite mixture models, offering practical advantages in modelling and inference for data that arise from non-standard distributions

$$F(\cdot; G) = \int K(\cdot; \boldsymbol{\theta})\, \mathrm{d}G(\boldsymbol{\theta}),\ G \sim \mathrm{DP}(\alpha, G_0).$$

- Mixtures of finite Polya trees:
  - It's straightforward to elicit prior information, because the parametric model is a centering special case

$$F \sim \int \mathrm{FPT}(c, F_{\boldsymbol{\theta}})\, p(\mathrm{d}c, \mathrm{d}\boldsymbol{\theta})$$

Dirichlet Process Mixtures (DPM)

- Location normal mixture models for the distributions for diseased and non-diseased populations:

$$F_l(t) \equiv F_l(t; H_l, \sigma^2) = \int N(t; \theta, \sigma^2) dH_l(\theta), l = 0, 1$$

- The $N(\theta, \sigma^2)$ distribution is stochastically ordered in θ for fixed σ^2
  i.e. if $\theta_1 \leq \theta_2$, $N(\theta_1, \sigma^2) \leq_{st} N(\theta_2, \sigma^2)$
  so we obtain the stochastic ordering for mixtures if the mixing distributions are stochastically ordered:

$$H_0 \leq_{st} H_1$$

$$F_0(t; H_0, \sigma^2) \leq_{st} F_1(t; H_1, \sigma^2)$$

- Introduce latent distribution functions H and G on $\mathbb{R}$:

$$H_0(t) = H(t) \qquad H_1(t) = H(t)G(t)$$

The stochastically ordered DPM model is defined as:

$$F_0(t; H, \sigma^2) = \int N(t; \theta, \sigma^2) dH(\theta)$$

$$F_1(t; H, G, \sigma^2) = \int N(t; max(\theta, \phi), \sigma^2) dH(\theta) dG(\phi)$$

$$H \sim DP(\alpha_H, N(\mu_H, \tau_H^2))$$

$$G \sim DP(\alpha_G, N(\mu_G, \tau_G^2))$$

Hyper-parameters: $\boldsymbol{\psi} = (\alpha_H, \mu_H, \tau_H^2, \alpha_G, \mu_G, \tau_G^2)$

Mixtures of Finite Polya Trees (MPT)

- Define the model directly as :

$$F_0(t) = H_0(t) = H(t)$$
$$F_1(t) = H_1(t) = H(t)G(t)$$

- The mixture of finite PT priors are assigned as:

$$H \sim \int FPT(c_H, H_{\boldsymbol{\theta}_H}) dP_H(c_H, \boldsymbol{\theta}_H)$$
$$G \sim \int FPT(c_G, G_{\boldsymbol{\theta}_G}) dP_G(c_G, \boldsymbol{\theta}_G)$$

- H is randomly centered at $H_{\boldsymbol{\theta}_H} = N(\mu_H, \tau_H^2)$ with $\boldsymbol{\theta}_H = (\mu_H, \tau_H^2)$

  G is randomly centered at $G_{\boldsymbol{\theta}_G} = N(\mu_G, \tau_G^2)$ with $\boldsymbol{\theta}_G = (\mu_G, \tau_G^2)$

- The levels of finite Polya trees:
  - Fix as $J_H = J_G \equiv J$
  - Bias-Variance tradeoff: increasing J to J +1 essentially doubles the number of conditional probabilities and decrease the bias, while also increases overall variability and can reduce the predictive ability

Prior Elicitation and Model Comparison

- In order to make models more comparable, specify the prior that ensures roughly the same amount of prior information is incorporated in each of the models
  - Control the prior variability based on the expected prior predictive densities: match $E\{f_0(\cdot)\}$ and $E\{f_1(\cdot)\}$ under two models by specifying same priors for μ_H , and τ_H^2 under the MPT model is given the prior that is induced by $\tau_H^2 + \sigma^2$ under DPM model. Same idea for G.

  - Use L1-distance as a measure of prior density variability to specify the priors for α_H, α_G and c_H, c_G

$$f_0(\cdot) = F_0'(\cdot)$$

$$\| f_0 - E\{f_0\} \|_1 = \int_{\mathbb{R}} | f_0(t) - E\{f_0(t)\} | \, \mathrm{d}t.$$

w.l.o.g., assume $E\{f_0(t)\} \approx N(t;0,1)$ $E\{F_0(A)\} \approx N(A;0,1)$

$$\tfrac{1}{2} \| f_0 - N(0,1) \|_1 = \tfrac{1}{2} \int_{\mathbb{R}} | f_0(t) - N(t;0,1) | \, \mathrm{d}t = \sup_{A \subset \mathbb{R}} | F_0(A) - N(A;0,1) |$$

Through simulation of prior densities under both models, can find out the specifications constructing similar prior median and corresponding 95% CI

- Compare models via log-pseudo-marginal likelihood(LPML) based on predictive utility
  - Calculate conditional predictive ordinate (CPO) for each observation under each model and calculate LPML

$$\mathrm{LPML}_0 = \Sigma_{i=1}^{n_0} \log(\mathrm{CPO}_{0i}) \qquad \mathrm{LPML}_1 = \Sigma_{j=1}^{n_1} \log(\mathrm{CPO}_{1j}) \qquad \mathrm{LPML} = \mathrm{LPML}_0 + \mathrm{LPML}_1$$

Simulation Study

- Random sample $n_0 = n_1 = 500$ observations for each population:

$$x_{0i} \sim 0.5N(0,1) + 0.1N(1,1) + 0.4N(-5,1), j = 1,2,...,n_1$$
$$x_{1j} \sim 0.5N(0,4) + 0.5N(1,1), i = 1,2,...,n_0$$

- We establish the DPM model and express in the hierarchical form with latent mixing parameters: $\boldsymbol{\theta} = \{\theta_i : i = 1,...,n_0, n_0{+}1, ..., n_0{+}n_1\}$ and $\boldsymbol{\phi} = \{\phi_i : i = 1,...,n_1\}$

$$x_{0i}|\theta_i, \sigma^2 \overset{\text{ind}}{\sim} N(\theta_i, \sigma^2), i = 1,2,...,n_0$$
$$x_{1j}|\theta_{n_0+j}, \phi_j, \sigma^2 \overset{\text{ind}}{\sim} N(max(\theta_{n_0+j}, \phi_j), \sigma^2), j = 1,2,...,n_1$$
$$\theta_i|H \overset{\text{IID}}{\sim} H, i = 1,...,n_0, n_0+1, ..., n_0+n_1$$
$$\theta_j|G \overset{\text{IID}}{\sim} G, j = 1,2,...,n_1$$
$$H, G|\alpha_H, \mu_H, \tau_H^2, \alpha_G, \mu_G, \tau_G^2 \sim DP(\alpha_H, N(\mu_H, \tau_H^2))DP(\alpha_G, N(\mu_G, \tau_G^2))$$

- Introduce the additional mixing parameters θ_{n_0+j} can retain the first-stage conditionally independent specification in the hierarchical model after marginalizing in model the random distribution functions H and G over their DP priors

Simulation Study

- Prior settings:

$$\mu_H \sim N(-2,4), \mu_G \sim N(0.5,4)$$
$$\tau_H^2 \sim IG(3,10), \tau_G^2 \sim IG(3,4)$$
$$\sigma^2 \sim IG(3,4)$$
$$\alpha_H, \alpha_G \sim G(2,0.9)$$

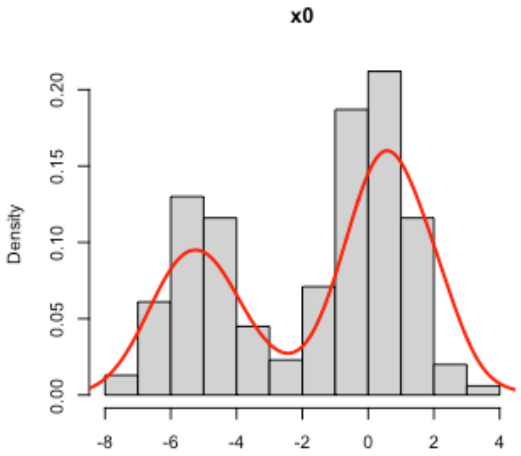
- Let $n_\theta^*(\leq n_0+n_1)$ with $\{\theta_l^* : l=1,...,n_\theta^*\}$ and $n_\phi^*(\leq n_1)$ with $\{\phi_l^* : l=1,...,n_\phi^*\}$ be the number of and values of the distinct components in $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$.

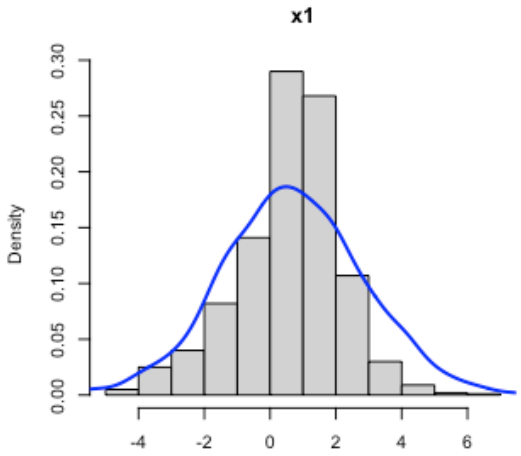
$$\alpha_H \sim G(a_{\alpha.H}, b_{\alpha.H}) \qquad \alpha_G \sim G(a_{\alpha.G}, b_{\alpha.G})$$
$$E(n_\theta^*) \approx a_{\alpha.H} b_{\alpha.H}^{-1} log(1+(n_0+n_1)a_{\alpha.H}^{-1} b_{\alpha.H}) \approx 13$$
$$E(n_\phi^*) \approx a_{\alpha.G} b_{\alpha.G}^{-1} log(1+n_1 a_{\alpha.G}^{-1} b_{\alpha.G}) \approx 12$$

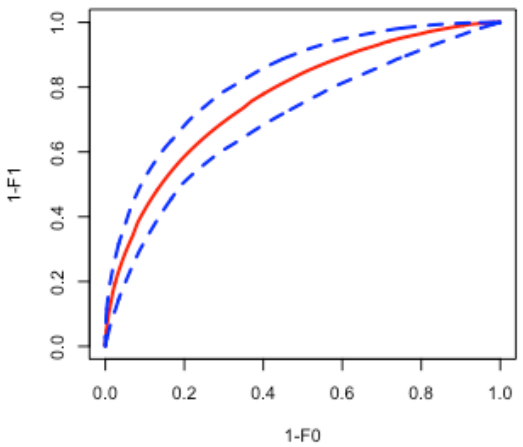
- Posterior simulation:
  The posterior simulation can be performed from $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2, \boldsymbol{\psi}|D)$ by integrating H and G over their DP priors:

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2, \boldsymbol{\psi}|D) \propto$$
$$\prod_{i=1}^{n_0} N(x_{0i}; \theta_i, \sigma^2) \prod_{j=1}^{n_1} N(x_{1j}; max(\theta_{n_0+j}, \phi_j), \sigma^2) p(\sigma^2) p(\boldsymbol{\theta}|\alpha_H, \mu_H, \tau_H^2) p(\boldsymbol{\phi}|\alpha_G, \mu_G, \tau_G^2) p(\boldsymbol{\psi})$$

Simulation Study







Serology Score Analysis

- Analyze the serology data from two different ELISA kits for detection of MAP in dairy cattle, the serology scores were log-transformed to facilitate the use of normal centering distributions of the non-parametric priors

- Prior specification:

DPM	MPT
$\mu_H, \mu_G \sim N(2,4)$	$\mu_H, \mu_G \sim N(2,1)$
$\tau_H^{-2}, \tau_G^{-2} \sim \Gamma(2,5)$	$\tau_H^{-2}, \tau_G^{-2} \sim \Gamma(3,1)$
$\sigma^{-2} \sim \Gamma(2,2.5)$	$J_H = J_G = 5$

Prior	DPM	MPT	Prior Median and 95% CI
A	$\alpha_H, \alpha_G \sim \Gamma(2,2)$	$c_H, c_G \sim \Gamma(2,1.5)$	0.5 (0.2, 1.3)
B	$\alpha_H, \alpha_G \sim \Gamma(5,1)$	$c_H, c_G \sim \Gamma(10,3)$	0.3 (0.1, 0.8)
C	$\alpha_H, \alpha_G \sim \Gamma(5,0.5)$	$c_H, c_G \sim \Gamma(15,2)$	0.2 (0.1, 0.6)

- Three priors for the DPM model precision parameters were considered, reflecting different levels of prior variability about the prior predictive density

Table 1. Synbiotic ELISA: LPML estimates and posterior summaries (medians and 95% CIs) for AUC under three prior choices for each of the DPM and MPT models†

Prior	*Results for DPMs*			*Results for MPTs*		
	$LPML_0$	$LPML_1$	*AUC*	$LPML_0$	$LPML_1$	*AUC*
A	−277.8	−408.9	0.720 (0.671,0.769)	−269.0	−398.6	0.730 (0.690,0.772)
B	−275.0	−408.1	0.716 (0.675,0.757)	−271.6	−400.3	0.738 (0.696,0.775)
C	−274.9	−408.6	0.720 (0.679,0.760)	−276.7	−407.1	0.743 (0.707,0.781)

Serology Score Analysis

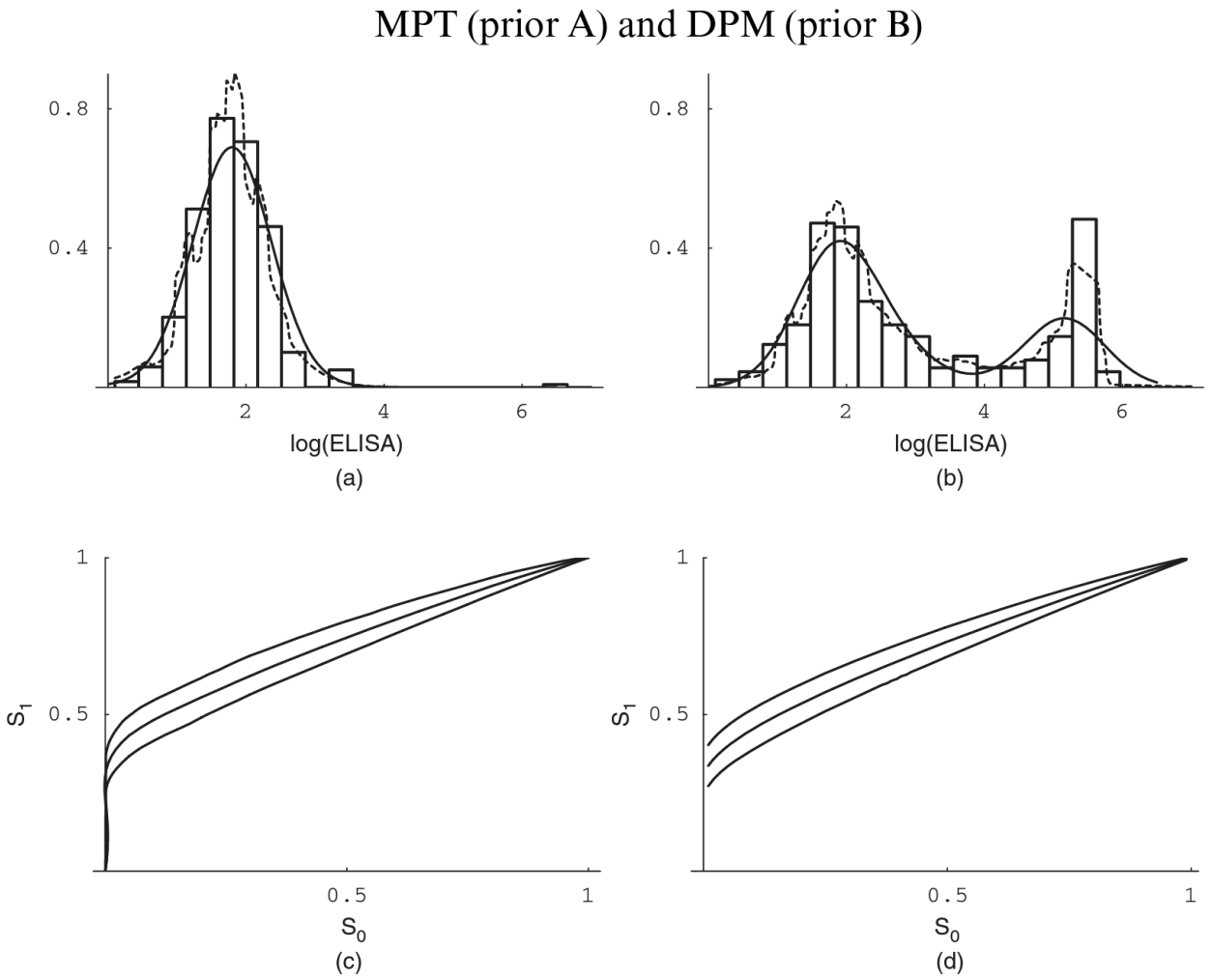


Fig. 1. (a), (b) Data histograms and estimated population densities of log-transformed serology scores (———, DPM results; -------, MPT results) and (c), (d) estimated ROC curves with 95% pointwise CIs for the Synbiotic ELISA test of Section 4.1: (a) non-infected group; (b) infected group; (c) MPT model; (d) DPM model

Thank you !

Appendix:

- We performed the posterior sampling based on Polya urn representation:

1. Sample $\theta_i|\{\theta_l : l \neq i\}, \alpha_H, \mu_H, \tau_H^2, D$ for $i : 1, ..., n_0$ with

$$p(\theta_i|\{\theta_l : l \neq i\}, \alpha_H, \mu_H, \tau_H^2, D) = \frac{\alpha_H q_0^\theta}{\alpha_H q_0^\theta + \sum_{j=1}^{n_0^{*-}} n_{0j}^- q_j^\theta} h(\theta_i|\alpha_H, \mu_H, \tau_H^2, \sigma^2, x_{0i}) + \sum_{j=1}^{n_0^{*-}} \frac{n_{0j}^- q_j}{\alpha_H q_0^\theta + n_{0j}^- q_j^\theta} \delta_{\theta_j^{*-}}(\theta_i)$$

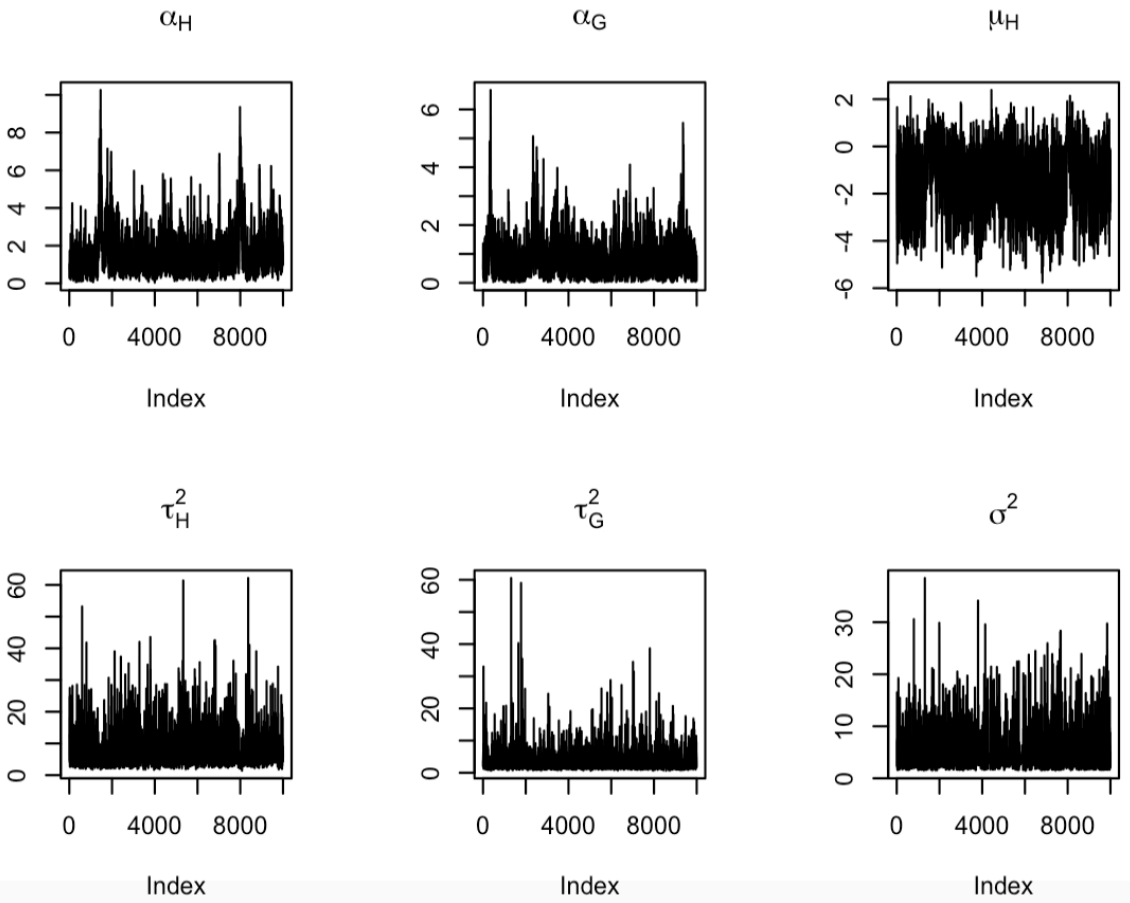
2. Sample $(\theta_{n_0+j}, \phi_j)|\{\theta_{n_0+l} : l \neq j\}, \{\phi_l : l \neq j\}, \alpha_G, \mu_G, \tau_G^2, \alpha_H, \mu_H, \tau_H^2, D$ for $j : 1, ..., n_1$ with Metropolis step that

$$p(\theta_{n_0+j}|\{\theta_{n_0+l} : l \neq j\}, \alpha_H, \mu_H, \tau_G^2, D) = \frac{\alpha_H q_0^\theta}{\alpha_H q_0^\theta + \sum_{j=1}^{n_1^{*-}} n_{1j}^- q_j^\theta} h(\theta_{n_0+j}|\alpha_H, \mu_G, \tau_G^2, \sigma^2, x_{1i}) + \sum_{j=1}^{n_1^{*-}} \frac{n_{1j}^- q_j}{\alpha_H q_0^\theta + n_{1j}^- q_j^\theta} \delta_{\theta_j^{*-}}(\theta_{n_0+j}$$

$$p(\phi_j|\{\phi_l : l \neq j\}, \alpha_G, \mu_G, \tau_G^2, D) = \frac{\alpha_G q_0^\phi}{\alpha_G q_0^\phi + \sum_{j=1}^{n_1^{*-}} n_{1j}^- q_j^\phi} h(\phi_j|\alpha_G, \mu_G, \tau_G^2, \sigma^2, x_{1i}) + \sum_{j=1}^{n_1^{*-}} \frac{n_{1j}^- q_j}{\alpha_G q_0^\phi + n_{1j}^- q_j^\phi} \delta_{\phi^{*-}}(\phi_j)$$

Accept the pair (θ_{n_0+j}, ϕ_j) with probability $min\{1, \frac{N(x_{1j}; max(\theta_{n_0+j}^{new}, \phi_j^{new}), \sigma^2)}{N(x_{1j}; max(\theta_{n_0+j}^{old}, \phi_j^{old}), \sigma^2)}\}$

3. Sample the rests following the structure in the slides


α_H
α_G
μ_H
τ^2_H
τ^2_G
σ^2
0 2 4 6 8
0 2 4 6
-6 -4 -2 0 2
0 20 40 60
0 10 20 30
0 4000 8000
Index