Bayesian Nonparametric Multiple Testing

Luis Gutiérrez *and others* "A Bayesian Nonparametric Multiple Testing Procedure for Comparing Several Treatments Against a Control." *Bayesian Analysis*, June 2019.

Soumyakanti Pan

Department of Biostatistics, UCLA

Problem statement

- Devise a strategy to test for differences between 1 control group and p treatment groups ($p > 1$).

- Setup:

$$\text{Control: } \mathbf{y}_c : (y_{c,1}, \ldots, y_{c,n_c})^\top \sim G_c$$
$$\text{Treatment: } \mathbf{y}_k : (y_{k,1}, \ldots, y_{k,n_k})^\top \sim G_k, k = 1, \ldots, p$$

- Most existing methods focuses on comparison of locations.

- Distributions can differ in features other than location, such as tails, symmetry and number of modes.

- We want to test which G_k's are different from the baseline distribution of the control population G_c.

Hypothesis

- Model space $\mathcal{M} = \{H_\gamma : \gamma = (\gamma_1, \ldots, \gamma_p) \in \{0,1\}^p\}$ where $H_{(\gamma_1, \ldots, \gamma_p)}$ represents the following model for the data $\mathbf{y}$.

$$y_{c,i} \overset{\text{iid}}{\sim} G_c, i = 1, \ldots, n_c$$

$$y_{k,i} \mid \gamma_k \overset{\text{iid}}{\sim} \begin{cases} G_k^*, & \text{if } \gamma_k = 1 \\ G_c, & \text{if } \gamma_k = 0 \end{cases}, i = 1, \ldots, n_k$$

where, $G_k^* \neq G_c$ for $k = 1, \ldots, p$.

- We are interested in finding the H_γ best supported by the data $\mathbf{y}$; hence, look for high posterior probability $\pi(H_\gamma \mid \mathbf{y})$.

Example

- For $p = 3$,

    $H_{(0,0,0)}$: all treatments equal to control

    $H_{(1,0,0)}$: only first treatment is different from control

- Intuition behind structure of model space, $\mathcal{M}$. Need a prior structure such that $H_{(0,0,0)} \longrightarrow H_{(1,0,0)}$.

- ANOVA analogy: (3 treatments: $\beta_1, \beta_2, \beta_3$)

    $$H_{(0,0,0)} : y_{ij} = \mu + \epsilon_{ij} \text{ vs } H_{(1,0,0)} : y_{ij} = \mu + \beta_1 + \epsilon_{ij}$$
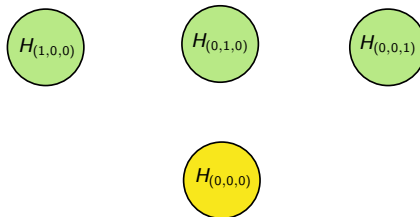
    If $\beta_1 = 0$, $H_{(1,0,0)} \equiv H_{(0,0,0)}$ implying nested structure of models

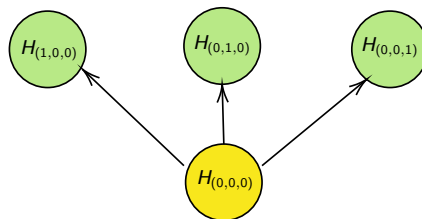 $$H_{(0,0,0)} \subseteq H_{(1,0,0)}.$$
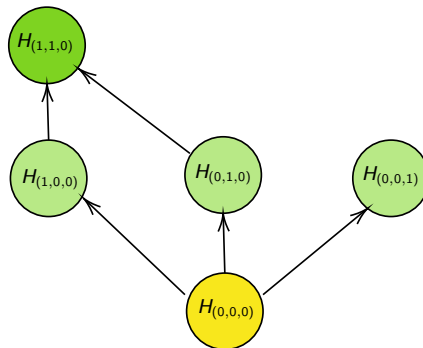
Structure of model space $\mathcal{M}$

$H_{(0,0,0)}$

Background
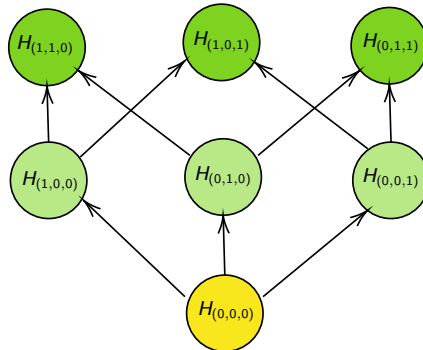○○○●○

Model and priors
○○○○○○

Inference
○

Applications
○○○

Conclusion

References

Problem

Structure of model space $\mathcal{M}$

Structure of model space $\mathcal{M}$

Structure of model space $\mathcal{M}$

Background
○○○●○

Model and priors
○○○○○○

Inference
○

Applications
○○○

Conclusion
○○○

References

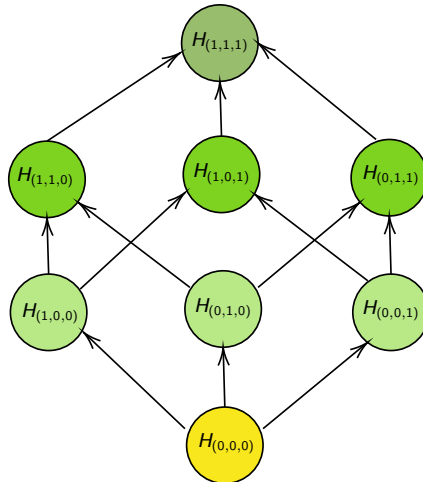Problem

Structure of model space $\mathcal{M}$

Structure of model space $\mathcal{M}$

Goals

This notion of nesting reflects the way in which model parameters are specified: the parameters of a hypothesis nested within others are also present in the hypotheses that nest it.

1. Develop a Bayesian testing procedure which will yield the posterior probabilities $\pi(H_\gamma \mid \mathbf{y}), \forall H_\gamma \in \mathcal{M}$.

2. Relax parametric assumptions and identify differences beyond location and scale.

3. If difference is found, gain insight about which aspects of the distributions differ.

Background
○○○○○

Model and priors
●○○○○○

Inference
○

Applications
○○○

Conclusion

References

Model definition

Suppose $y_{k,i}$ be the i-th observation in population k and $x_{k,i}$ indicates the population membership i.e, $x_{k,i} = k$ with $k \in \{c, 1, \ldots, p\}$.

$$y_{k,i} \mid x_{k,i}, \mathcal{P} \overset{\text{ind}}{\sim} \int \phi(y_{k,i} \mid \mu, \sigma^2) dP_{x_{k,i}}(\mu, \sigma^2)$$

$$\mathcal{P} \mid H_\gamma \sim \pi_{\text{DDP}}(\cdot \mid H_\gamma)$$

$$H_\gamma \sim \pi_{\mathcal{M}}$$

where $\mathcal{P} = \{P_x : x \in \{c, 1, \ldots, p\}\}$. $\pi_{\text{DDP}}(\cdot \mid H_\gamma)$ is a prior induced by Dependent Dirichlet Process (MacEachern 2000) under H_γ.

Background	Model and priors	Inference	Applications	Conclusion	References
ooooo	ooooooo	o	ooo		

DDP prior π_{DDP}

The prior $\pi_{\text{DDP}}(\cdot \mid H_\gamma)$ is induced by a process $\mathcal{P}$ whose elements are,

$$P_x = \sum_{k=1}^{\infty} \omega_j \delta_{\left(\mu_j(x), \sigma_j^2(x)\right)}$$

where weights are defined with stick-breaking construction specified by $\text{Beta}(1, \kappa)$ and hyperprior $\kappa \sim \text{Gamma}(a_1, a_2)$ for extra flexibility.
The atoms are defined as

$$\mu_j(x) = \mu_{c,j} + \eta_{x,j} \text{ and } \sigma_j^2(x) = \sigma_{c,j}^2 \tau_{x,j}$$

for $x \in \{c, 1, \ldots, p\}$. Set $\eta_{c,j} = 0$ and $\tau_{c,j} = 1$.

Mixing distributions

The DDP prior specified results the following mixing distributions of the populations.

$$\text{Control: } G_c(\cdot) = \sum_{j=1}^{\infty} \omega_j \Phi(\cdot \mid \mu_{c,j}, \sigma_{c,j}^2)$$

$$\text{Treatment: } G_k(\cdot) = \sum_{j=1}^{\infty} \omega_j \Phi(\cdot \mid \mu_{c,j} + \eta_{k,j}, \sigma_{c,j}^2 \tau_{k,j})$$

Thus, the possible differences in the k-th population are captured by the changes in the random sequences $\{\eta_{k,j}, \tau_{k,j}\}_{j \geq 1}$. Note how atoms from two populations can have the same value - this similarity among mixing distributions enables us to make comparison between groups.

Background
○○○○○

Model and priors
○○○●○○

Inference
○

Applications
○○○

Conclusion

References

Priors on atoms $\pi_k(\cdot \mid \gamma_k, s, \epsilon, b)$

How to make $(\eta_{k,j}, \tau_{k,j})$ concentrate around $(0, 1)$ whenever $G_c = G_k$?
Simple - spike-and-slab prior (George and McCulloch 1997).

$$\text{Control atoms: } \mu_{c,j} \sim \mathcal{N}(0, \epsilon s)$$
$$\sigma_{c,j}^2 \sim \mathcal{IG}(b/s, b/s)$$
$$\text{Treatment atoms: } \eta_{k,j} \mid H_{\boldsymbol{\gamma}} \sim \gamma_k \mathcal{N}(0, \epsilon s) + (1 - \gamma_k)\mathcal{N}(0, \epsilon)$$
$$\tau_{k,j} \mid H_{\boldsymbol{\gamma}} \sim \gamma_k \mathcal{IG}(b/s, b/s) + (1 - \gamma_k)\mathcal{IG}(b, b)$$

Note ϵ and s control the variance of $\eta_{k,j}$ and $\tau_{k,j}$ (ϵ small near 0 and s large). Hence, when $\gamma_k = 0$, the atoms of the k-th population will concentrate tightly about $\{\mu_{c,j}, \sigma_{c,j}^2\}_{j \geq 1}$ imposing the nested structure.

Background
○○○○○

Model and priors
○○○○●○

Inference
○

Applications
○○○

Conclusion

References

Priors on $\mathcal{M}$

$\pi(H_0) = \rho/(\rho + 1)$ and the prior for the alternative hypotheses H_γ is obtained from the recursion

$$\pi(H_\gamma) = \pi_p(\ell)\binom{p}{\ell}^{-1} = \left\{ \rho \sum_{j=1}^{p-\ell} \pi_p(\ell + j)\binom{\ell + j}{\ell} \right\} \binom{p}{\ell}^{-1}$$
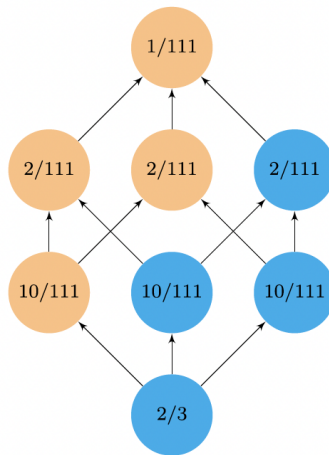
where $\ell = \sum_{k=1}^{p} \gamma_k$ and $\rho > 0$ is a hyperparameter which fixes the relative odds of belief in a set of local alternatives versus a local null hypothesis.

Background
○○○○○

Model and priors
○○○○○●

Inference
○

Applications
○○○

Conclusion

References

Intuition behind $\pi(H_\gamma)$

- Suppose, $p = 3$.
- Note, $H_{(1,0,0)}$ is nested in $H_{(1,1,0)}, H_{(1,0,1)}, H_{(1,1,1)}$.
- If $\rho = 2$,

$$\pi(H_{(1,0,0)}) = \frac{10}{111}$$

$$= 2 \times \left(\frac{2}{111} + \frac{2}{111} + \frac{1}{111} \right)$$

- $\pi(H_\gamma) = \rho \sum_{\gamma' \in \mathsf{Nest}(\gamma)} \pi(H'_\gamma)$

Prior probabilities with $\rho = 2$

Posterior inference

- Overcome infinite-dimensionality by considering an augmented model and then proceed with Gibbs sampler (just sampling location, scale and weights of a finite mixture model) assisted by conjugacy.

- Approximate posterior probability of model $H_{\gamma'} \in \mathcal{M}$ by

$$\pi(H_{\gamma'} \mid \mathbf{y}) = \frac{1}{B} \sum_{\ell=1}^{B} \mathbf{1}_{\{\gamma^{(\ell)} = \gamma'\}}$$
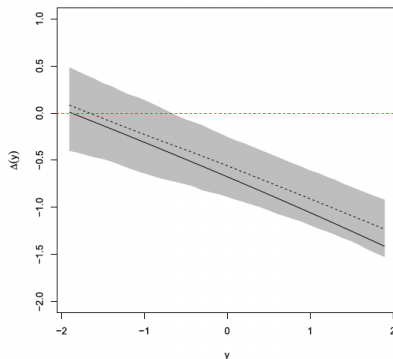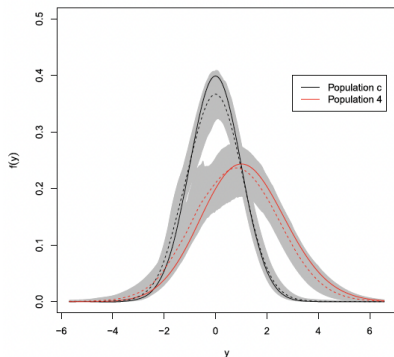
- Use shift function to visualize the aspects in which $G_k \neq G_c$.

$$\Delta_k(y) = G_k^{-1}(G_c(y)) - y$$

Facilitates to quantify uncertainty for the set $\{y : \Delta_k(y) \neq 0\}$.

Background
ooooo

Model and priors
oooooo

Inference
o

Applications
●oo

Conclusion

References

Synthetic example 1

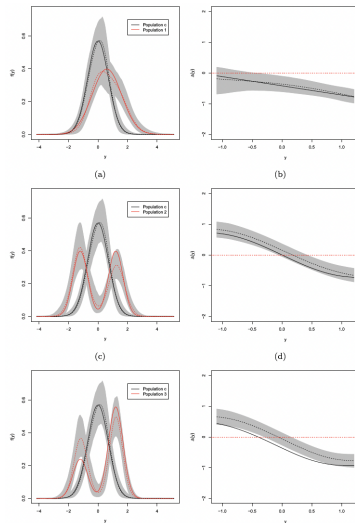| True model $H_{(0,0,1)}$ | Control $\mathcal{SN}(0,1,0)$ | Trt 1 $\mathcal{SN}(0,1,0)$ | Trt 2 $\mathcal{SN}(0,1,0)$ | Trt 3 $\mathcal{SN}(0,1,1)$ |

Synthetic example 2

Under the true model $H_{(1,1,1)}$,

Control	$\mathcal{N}(0, 0.49)$
Trt 1	$\mathcal{N}(0.6, 1)$
Trt 2	$0.5\mathcal{N}(-1.2, 0.25)+$
	$0.5\mathcal{N}(1.2, 0.25)$
Trt 3	$0.3\mathcal{N}(-1.2, 0.25)+$
	$0.7\mathcal{N}(1.2, 0.25)$

Background
○○○○○

Model and priors
○○○○○○

Inference
○

Applications
○○●

Conclusion

References

Educational achievement by school type

School-type proxy for socio-economic status; educational achievement strongly correlated with socio-economic status?
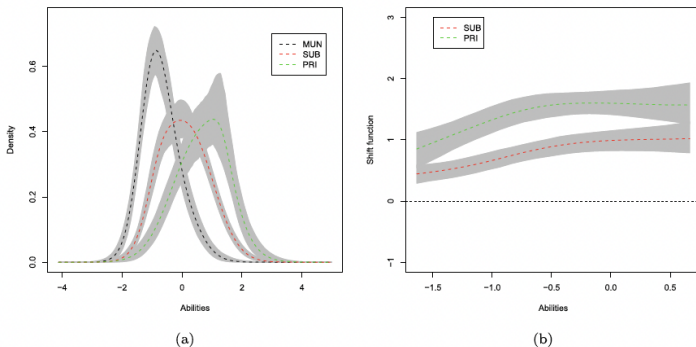

Figure 6: (a) Posterior mean and 95% point-wise credibility intervals of the densities for the educational achievement by school type of Section 5.1. (b) Shift functions for the comparison of subsidized and private schools versus municipal (control).

Remarks

- The nesting structure is facilitated by considering common weights.

- Advantage is that it yields density estimate for all populations relaxing parametric assumptions

- Shift function is calculated from the estimated densities and potentially provides new insight into the aspect of differences.

- This approach does not scale with large p owing to high computational cost.

📄 George, Edward I. and Robert E. McCulloch (1997). "Approaches for Bayesian Variable Selection". In: *Statistica Sinica* 7.2, pp. 339–373.

📄 MacEachern, S. N. (2000). *Dependent Dirichlet processes*.