STAT 30850 Project Report

Arjun Biddanda (abiddanda@uchicago.edu)
Joseph Marcus (jhmarcus@uchicago.edu)

February 29, 2016

Let

t - time index of a test statistic streaming in
X - a vector of t test statistics that have streamed in
X_t - the test statistic at the t^{th} time point
Z - vector of latent states of X_t being a signal or null
Z_t - latent state at time t of X_t being a signal or null
π_0 - proportion of nulls
μ_1 - mean of the signals
σ_1^2 - variance of the signals

We model X_t as a mixture of gaussians:

$$X_t \mid \pi_0, \mu_1, \sigma_1 \sim \pi_0 N(0,1) + (1-\pi_0)N(\mu_1, \sigma_1^2)$$
$$X_t \mid Z_t = 0 \sim N(0,1)$$
$$X_t \mid Z_t = 1, \mu_1, \sigma_1^2 \sim N(\mu_1, \sigma_1^2)$$

We can reparamertize this model in terms of the prescison ϕ_1 of the signals and write down the likelihood of the model conditioned on the latent indicators as:

$$L(\pi_0, \mu_1, \sigma_1^2 \mid X, Z) \propto (\pi_0)^{n_0} exp(-\frac{1}{2}\sum_{t:z_t=0} x_t^2) \cdot (1-\pi_0)^{n_1} exp(-\frac{\phi_1}{2}\sum_{t:z_t=1}(x_t-\mu_1)^2)$$

where n_0 and n_1 are the number observed of nulls and signals respectivley. We can then set priors on π_0, μ_1, ϕ_1 that all satisfy conjugacy:

$$\pi_0 \sim Beta(\alpha, \beta)$$
$$\phi_1 \sim Gamma(\frac{a}{2}, \frac{b}{2})$$
$$\mu_1 \mid \phi_1 \sim Normal(\mu^*, \frac{1}{\alpha^* \phi_1})$$

thus the posteriors can be written as: (**DOUBLE CHECK THIS**)

$$\pi_0 \mid X, Z = 0 \sim Beta(\alpha + n_0, \beta + n_1)$$
$$\phi_1 \mid X, Z \sim Gamma(\frac{a+n_1}{2}, b + \sum_{t:z_t=1}(x_t-\mu_1)^2)$$
$$\mu_1 \mid X, Z, \phi_1 \sim Normal(\frac{\alpha^*\mu^* + n_1 + \bar{x_1}}{\alpha^* + n_1}, \frac{1}{(\alpha^* + n_1)\phi_1})$$

We propose to implement a gibbs sampler to sample from the posterior distributions of the parameters:

1. Set $\pi_0^{(0)}$, $\mu_1^{(0)}$ and $\phi_1^{(0)}$