Lecture 13: Bayesian Multiple Testing

Merlise Clyde

October 20

Normal Means Model

Recall normal model with

$$Y_i \mid \mu_i, \sigma^2 \overset{ind}{\sim} \mathsf{N}(\mu_i, \sigma^2)$$

- $H_{0i} : \mu_i = 0$ versus $H_{1i} : \mu_i \neq 0$
- Spike & Slab Prior:

$$\mu_i \overset{iid}{\sim} \pi_0 \delta_0 + (1 - \pi_0) g(\mu_i \mid 0, \tau, H_{i1})$$

- need to specify
  - π_0
  - g
- concern: is that # errors blows up with n (n = # tests = dimension of $\{\mu_i\}$)

Approach 1: Prespecify π_0

$\pi_0 = \Pr(H_{0i})$

- seemingly non-informative choice?

$$\pi_0 = 0.5$$

- What does imply about the "model size" prior? The number of times H_{1i} is true?

- Let

$$\gamma_i = \begin{cases} 1 \text{ if } H_{1i} \text{ is true} \\ 0 \text{ if } H_{0i} \text{ is true} \end{cases}$$

$$\gamma^{(n)} = (\gamma_1, \gamma_2, \ldots, \gamma_n)^2$$

- e.g. $\gamma^{(n)} = (0, 1, 0, 0, \ldots, 1)^T$

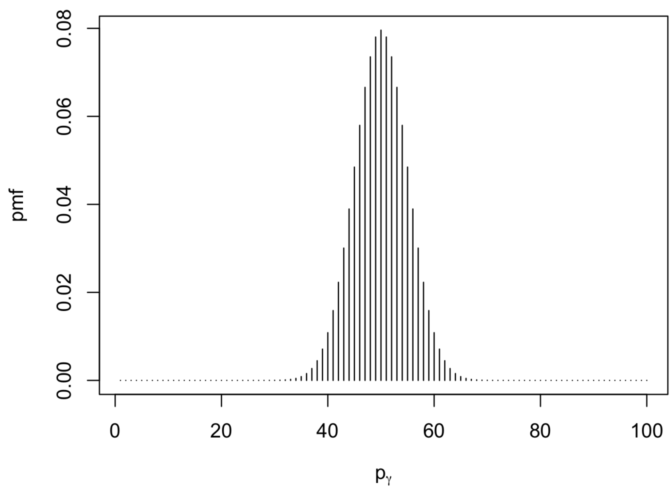
- model size $p_\gamma = \sum_{i=1}^n \gamma$ is the number of non-zero values

- Distribution of p_γ?

Induced Distribution

$$p_\gamma \sim \text{Binomial}(n, 1/2)$$

- Expected 1/2 of the hypotheses to be true a priori



Probabities of no or at least 1 signal

$$p_\gamma \sim \text{Binomial(n, 1/2)}$$

- probability of observing no signals $\gamma^{(n)} = (0, 0, 0, \ldots, 0)^T$ or $p_\gamma = 0$

$$\Pr(p_\gamma = 0) = \pi_0^n = 0.5^n$$

- approximately 0 for large n
- probability of at least one signal is $1 - 0.5^n \approx 1$

Control Type I Errors

- Suppose we want to fix π_0 that protexts against Type I errors blowing up as n increases

$$\Pr(p_\gamma = \mathbf{0}_n) = \frac{1}{2} = \pi_0^n$$

- "Bayesian Bonferroni Prior"

- so $\pi_0 = 0.5^{1/n}$ very close to 1 ! Need overwhelming evidence in the data for $\Pr(H_{1i} \mid y^{(n)})$ to not be ≈ 0!

- not a great idea to prespecify π_0!

Approach 2: Empirical Bayes

- We could try to maximize the marginal likelihood

$$Y_i \mid \mu_i, \sigma^2 \overset{ind}{\sim} \mathsf{N}(\mu_i, \sigma^2)$$

$$\mu_i \overset{iid}{\sim} \pi_0 \delta_0 + (1 - \pi_0)\mathsf{N}(\mu_i; 0, \tau)$$

$$\begin{aligned}
\mathcal{L}(\pi_0, \tau) &= \int_{\mathbb{R}^n} \prod_{i=1}^{n} \mathsf{N}(y_i; \mu_i, \sigma^2) \left\{\pi_0 \delta_0(\mu_i) + (1 - \pi_0)\mathsf{N}(\mu_i; 0, \tau)\right\} d\mu_1 \ldots d\mu_n \\
&= \prod_{i=1}^{n} \int_{\mathbb{R}} \mathsf{N}(y_i; \mu_i, \sigma^2) \left\{\pi_0 \delta_0(\mu_i) + (1 - \pi_0)\mathsf{N}(\mu_i; 0, \tau)\right\} d\mu_i
\end{aligned}$$

- Conjugate or nice setups we can integrate out μ_i and then maximize marginal likelihood for π_0 and τ
- Numerical integration or EM algorithms to get $\hat{\pi}_0^{\mathsf{EB}}$ and $\hat{\tau}^{\mathsf{EB}}$
- Clyde & George (2000) Silverman & Johnstone (2004) for wavelet regression

Expectation-Maximization

- introduce latent variables so that "complete" data likelihood is nice! e.g. γ:

$$y_i \mid \gamma_i, \tau \overset{ind}{\sim} \mathsf{N}(0,1)^{1-\gamma_i}\mathsf{N}(0, 1+\tau)^{\gamma_i}$$

$$\gamma_i \overset{iid}{\sim} \mathsf{Ber}(1-\pi_0)$$

- Iterate: For $t = 1, \ldots$
  - M-step: Solve for $(\hat{\pi}_0^{(t)}, \hat{\tau}^{(t)}) = \arg\max \mathcal{L}(\pi_o, \tau \mid \hat{\gamma}^{(t-1)})$
  - E-step: find the expected values of the latent sufficient statistics given the data, $\hat{\pi}_0^{(t)}$, $\hat{\tau}^{(t)}$

$$\hat{\gamma}^{(t)} = \mathsf{E}[\gamma_i \mid y, \hat{\pi}_0^{(t)}, \hat{\tau}^{(t)}]$$

Adding Noise

What happens to $\hat{\pi}_0^{\text{EB}}$ as we add more and more noise to a fixed number of signals?

- $\hat{\pi}_0^{\text{EB}}$ becomes closer to one as $n \to \infty$ for a fixed number of true signals
- This is good as it protects against Type I errors blowing up as n increases!
- However it becomes more and more difficult to find the few needles in a haystack!

Approach 3: Fully Bayes

Choose a prior for π_0: $\pi_0 \sim \text{Beta}(a, b)$

Consider the thought experiment ...

$$\gamma^{(n)} = (?, 0, , \ldots, 0)^T$$

where we don't know the first hypothesis but we know that the others are all null $\gamma_j = 0$ for $j = 2, \ldots, n$

- $\gamma_i \sim \text{Bernoulli}(1 - \pi_0)$
- Update the prior for π_0 to include the info $\gamma_j = 0$ for $j = 2, \ldots, n$

$$\pi(\pi_0 \mid \gamma_2, \ldots, \gamma_n) \propto \pi_0^{a-1}(1-\pi_0)^{b-1} \prod_{j=2}^{n} \pi_0^{1-\gamma_j}(1-\pi_0)^{\gamma_j}$$

$$\pi(\pi_0 \mid \gamma_2, \ldots, \gamma_n) \propto \pi_0^{a+n-1-1}(1-\pi_0)^{b-1}$$

Beta Posterior

$$\pi_0 \mid \gamma_2, \ldots, \gamma_n \sim \text{Beta(a + n - 1, b)}$$

with mean

$$\mathsf{E}[\pi_0 \mid \gamma_2, \ldots, \gamma_n] = \frac{a+n-1}{a+n-1+b}$$

- suppose $a = b = 1$ (Uniform prior)

$$\mathsf{E}[\pi_0 \mid \gamma_2, \ldots, \gamma_n] = \frac{n}{n+1}$$

- implies probability of $H_{01} \to 1$ and $H_{11} \to 0$ as $n \to \infty$ borrowing strength from other nulls
- Multiplicity adjustment as in the EB case
- Scott & Berger (2006 JSPI, 2010 AoS) show that above framework protects against increasing Type I errors with n; We also get FDR control automatically

Induced Prior on p_γ

Exercise: If $p_\gamma \mid \pi_0 \sim \text{Binomial}(n, 1-\pi_0)$ and $\pi_0 \sim \text{Beta}(1,1)$

--

- What is the probability that $p_\gamma = 0$
- What is the probability that $p_\gamma = n$
- What is the distribution of p_γ ?

This is a Beta-Binomial and in the special case $a = b = 1$ this is a discrete uniform on model size!

Bottomline: We need to "learn" key parameters in our hierarchical prior or the magic doesn't work! Borrowing comes through using all the data to inform about "global" parameters in the prior, in this case π_0

Posteriors, Inference and Decisions

- Joint posterir distribution of $\mu_1, \ldots, \mu_n$ averaged over hypotheses "Model averaging"
  - distribution is a spike at 0 and continous distribution
- select a hypothesis
- report posterior (summaries) conditional on a hypothesis
- issue **winner's curse** !
- need to have coherent conditional inference given that you selected a hypothesis.
- Don't report selected hypotheses but report results under model averaging!

Choice of g

$$\mu_i \overset{iid}{\sim} \pi_0 \delta_0 + (1 - \pi_0) g(\mu_i \mid 0, \tau, H_{i1})$$

- growing literature on posterior contraction in high dimensional settings as $n \to \infty$ with "sparse signals"
- posterior $\pi(\mu^{(n)}) \mid y^{(n)})$

Want

$$\Pr(\mu^{(n)} \in \mathcal{N}_{\epsilon_n}(\mu_0^{(n)}) \mid y^{(n)}) \to 1$$

- assume that there are s signals (fixed or growing slowly)
- signal values are bounded away from zero
- Want the posterior under the Spike and Slab prior to concentrate on this neighborhood (ie. probability 1)
- active area of research!