

Lecture 12: Normal Means & Multiple Testing

Merlise Clyde

October 11



Normal Means Model

Suppose we have normal data with

$$Y_i \mid \mu_i, \sigma^2 \stackrel{iid}{\sim} (\mu_i, \sigma^2)$$

- Means Model $\mu_i \stackrel{iid}{\sim} g$, "random effects" distribution

Multiple Testing

- $H_{0i} : \mu_i = 0$ versus $H_{1i} : \mu_i \neq 0$
- n hypotheses that may potentially be closely related, e.g. H_{01} no difference in expression gene i between cases and controls, for n genes



Strategy Ia

- p-value, p_i , or testing H_{0i} versus H_{1i} for each i
- $p_i < \alpha$ implies reject H_{0i} in favor of H_{1i} , e.g. $\alpha = 0.05$

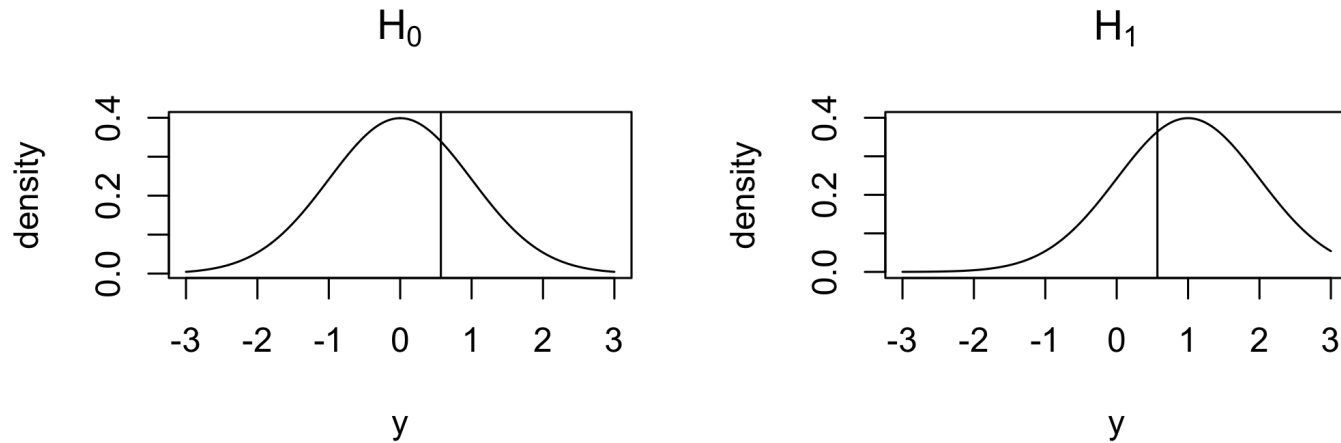
Limitations?

- overall lots of type I errors potentially in testing over and over again
- α is the probability of making a type I error in an individual test, but not the probability of the family-wise type 1 error, e.g. the probability of making at least one type 1 error in the n tests)



Power

- very low power (high type II error rate) because we have a single observation per hypothesis



- low power unless we have good separation between the two distributions (large difference relative to noise)
- low power may actually lead to very few type I errors even in multiple testing but often still lots of type I and type II errors

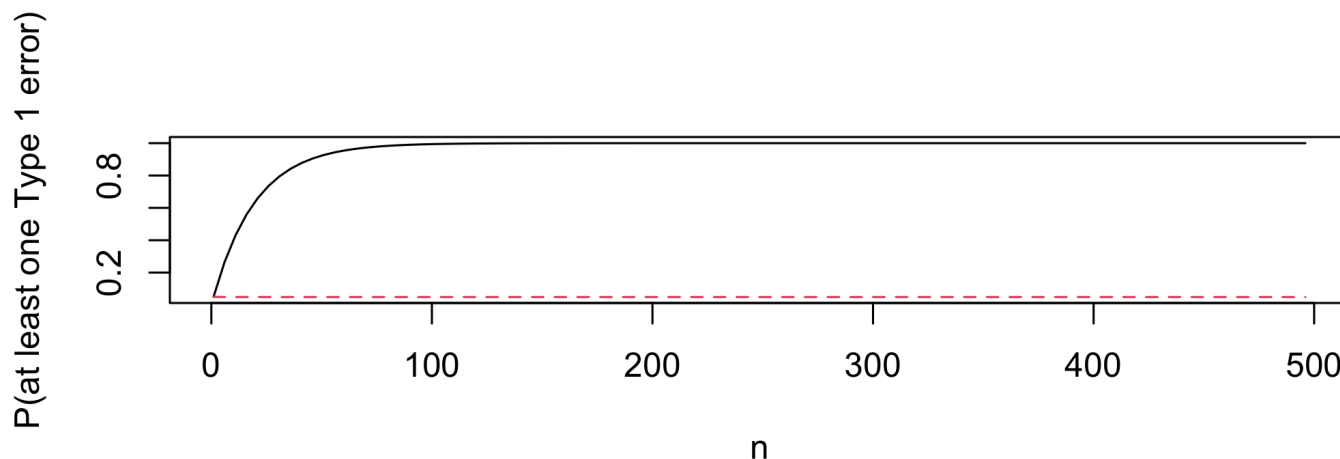


Strategy 1b

Adjust the level of each test to reflect how many tests you are conducting

- Probability of at least one Type I error if tests are independent

$$1 - \Pr(0 \text{ Type I errors in } n \text{ tests}) = 1 - (1 - \alpha)^n$$



- to control the increase in Type I errors with n we may need to decrease the α threshold with n



Classical Strategy

- control the family-wise error rate. Assuming independence across tests (reality ?) replace α with α/n

Bonferroni correction: keeps overall family wise error at α

- if we have 10,000 tests $\alpha_{\text{Bon}} = 0.05/10000$ very small
- in the extremely low power setting, probably very few tests exceed the new threshold (conservative)



False Discovery Rate (FDR)

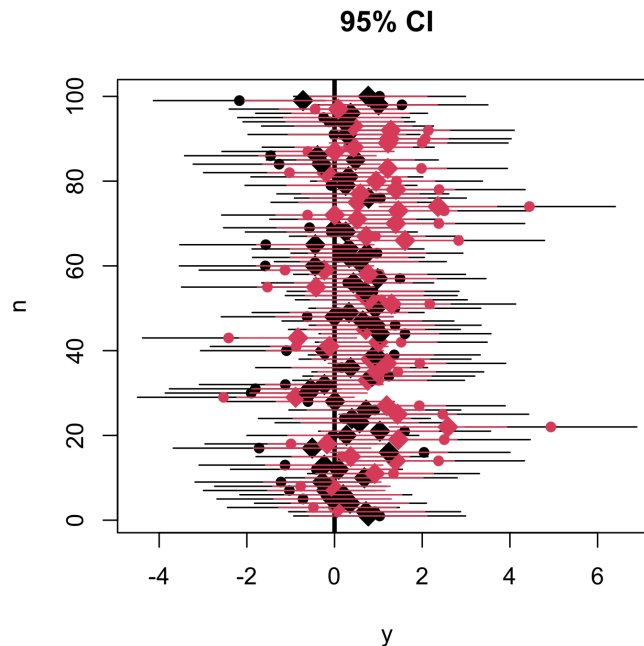
- FDR threshold α_{FDR}
- if $p_i < \alpha_{\text{FDR}}$, call this is a "discovery"
- collect all of our discoveries, say 100 out of 10,000 genes
- we want that the proportion of discoveries that are false (i.e H_0 was actually true) to be small
- control the proportion of false discoveries at level α instead of individual p-values
- Benjamini & Hochberg (BH) (1995 JRSS-B) propose a simple choice for α_{FDR} based on n and assuming n independent tests
- Issue: we will still have lower power in this low data scenario!
- Borrow strength!



Strategy II: Hierarchical Model

$$Y_i \mid \mu_i, \sigma^2 \stackrel{iid}{\sim} (\mu_i, \sigma^2)$$
$$\mu_i \stackrel{iid}{\sim} g$$

- naive approach: choose g as $N(\mu, \sigma_\mu^2)$ & estimate μ and σ_μ^2 (Empirical Bayes) $\hat{\mu} = \bar{y}$ and $s_y^2 = 1 + \hat{\sigma}_\mu^2$, so $\hat{\sigma}_\mu^2 = \max(0, 1 - s_y^2)$



Informal approach to testing

- Conclude in favor of H_{1i} if $0 \notin (\mu_{Li}, \mu_{Ui})$
- otherwise fail to reject

Question: Do we expect this approach to have a huge Type I error rate exploding with n (# tests)? Why or why not?

- shrinkage and borrowing of information leads to narrower CI
- information from the other y_i s enters into the posterior for μ_i through the estimates of μ and σ_μ^2

$$\mu_i \mid y_1, \dots, y_n \sim N \left(\frac{y_i + \hat{\mu}/\hat{\sigma}_\mu^2}{1 + 1/\hat{\sigma}_\mu^2}, \frac{1}{1 + 1/\hat{\sigma}_\mu^2} \right)$$

- when σ_μ^2 is small credible intervals are much narrower than with MLE



Hypothetical Setting

- first $i = 1, 2, 3$ "signals" (H_{1i} is true)
- add $n - 3$ nulls (H_{0i} is true)

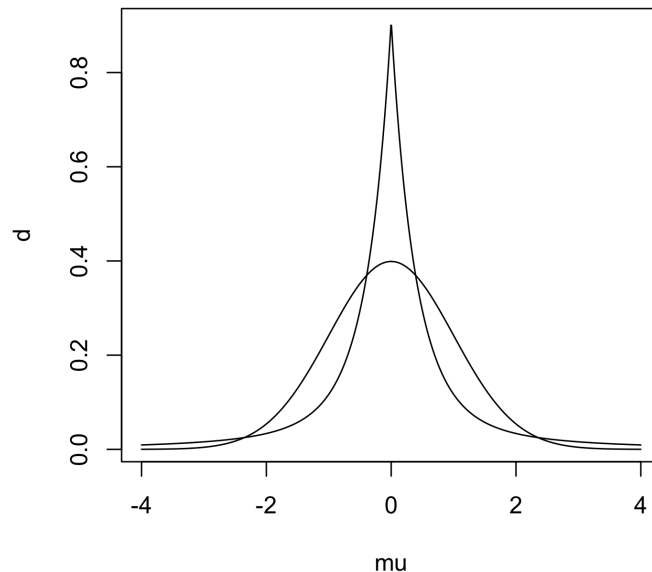
Does throwing in more nulls lead to more Type I errors?

- what happens to $\hat{\mu}$ and $\hat{\sigma}_{\mu}^2$?
- what happens to the credible intervals?



Informal Approach B

- an issue with the $N(\mu, \sigma_\mu^2)$ for g in the hypothetical setting is that it can capture only noise and not the signals. (signals are outliers under normal model)
- choose a more flexible g to capture both noise and signal!



Local-Global Scale Mixtures of Normals

Local scale

$$\begin{aligned}\mu_i &| \lambda_i, \tau \sim N(0, \lambda_i \tau) \\ \lambda_i &\sim f \quad \text{local-scale} \\ \tau &\sim h \quad \text{global-scale}\end{aligned}$$

- density that is concentration around zero to shrink noise to zero
- heavy tails avoid over-shrinkage of signals (want heavier than normal)
- Includes:
 - horseshoe
 - generalized double pareto
 - Dirichlet Laplace



Note

- a single Gaussian or Double Exponential prior (Bayes Lasso) have exponential tails same as likelihood (in the normal means problem)
- single parameter controls tail behaviour and concentration at zero
- will overshrink the signal if there are many noise cases
- Good shrinkage prior allows separate control of the concentration around zero and tails
- tails need to exhibit bounded influence
- continuous versions/relaxations of a spike and slab prior

$$\mu_i \sim \pi_0 \delta_0 + (1 - \pi)g$$

- allows formal Bayes multiple testing $H_{0i} : \mu = 0$
- $\pi_0 = \Pr(H_{0i} \text{ is true})$ another unknown to learn from the data; provides automatic adjustment for multiple testing error!

