The homework needs to be returned to the TAs by May 24.

1. (a) Suppose we observe an $n$ dimensional Gaussian vector $y \sim N(\mu, I_n)$ and wish to estimate a linear transformation $L\mu$ of the mean vector $\mu$ by means of a statistic $T(y)$. Under appropriate conditions which you may want to discuss, give an unbiased estimate for the MSE, namely, $\mathbb{E} \|T(y) - L\mu\|^2$ of this estimate.

(b) Suppose now that the $n$ dimensional Gaussian vector $y$ is distributed as $y \sim N(\mu, \Sigma)$ with known covariance matrix $\Sigma$. We are interested in estimating the mean vector $\mu$. Give an unbiased estimate for the MSE, namely, $\mathbb{E} \|T(y) - \mu\|^2$ of any appropriate estimate. [Hint: You may use the result of the previous question or any other method.]

2. Show that false coverage rate (FCR) control can be adapted to show directional FDR control in the following sense: assume we have $T_i \overset{\text{ind}}{\sim} N(\mu_i, 1)$ for $1 \leq i \leq n$, and apply the BHq procedure for testing $H_i: \mu_i = 0$ vs. $\mu_i \neq 0$. If $H_i$ is rejected, provide an estimate $\hat{s}_i$ of the sign of $\mu_i$. Now consider the directional FDR

$$FDR_{\text{dir}} = \mathbb{E} \left[ \frac{V}{R \vee 1} \right],$$

where $V$ is the number of directional errors

$$V = \sum_{i: H_i \text{ rejected}} 1\{\text{sgn}(\mu_i) \neq \hat{s}_i\}$$

with the convention that $\text{sgn}(0) = 0$. That is, the error rate counts not only the nulls that are falsely rejected but also those non-nulls for which we got the signs wrong. Show that the BHq procedure controls the directional FDR at level $q$.

3. Try to replicate the baseball example from Lecture 19. That is, download a real data set from whatever source you like and about any topic you like and see whether the James-Stein estimate improves on the MLE. Explain where you obtained the data set, how you transformed the data (in case you did) and how you happen to (approximately) know the true vector of means. Comment on your findings.