

Likelihood Theory and Extensions (BIOS:7110)
Breheny

Assignment 4

Due: Monday, September 29

1. *Gaussian graphical model.* As we discussed in class, the precision matrix Θ is of interest as it describes conditional independence relationships. One way to estimate Θ is $\hat{\Theta} = \mathbf{S}^{-1}$, where $\mathbf{S}$ is the sample variance-covariance matrix (throughout this problem, you may assume that both $\mathbf{S}$ and Σ are full rank). Here, we consider a different approach.

- (a) Suppose we partition Θ so that the top left corner is isolated (i.e., the top left corner of the partition is 1×1 and the bottom right is $(d - 1) \times (d - 1)$, where d is the dimension of the multivariate distribution). Show that

$$-\Theta_{21}/\Theta_{11} = \Sigma_{22}^{-1}\Sigma_{21},$$

where the variance Σ is partitioned in the same way as Θ . Hint: definition of a matrix inverse.

- (b) Now consider the conditional distribution of $x_1|\mathbf{x}_2$. Show that if $\mathbf{x}$ is multivariate normal, then the conditional distribution of $x_1|\mathbf{x}_2$ can be written as

$$X_1 = \alpha + \mathbf{x}_2^\top \boldsymbol{\beta} + \varepsilon,$$

where $\varepsilon \sim N(0, \sigma^2)$. Express $\boldsymbol{\beta}$ and σ^2 in terms of the precision matrix Θ .

- (c) Part (b) suggests that we can estimate Θ using linear regression. Simulate some multivariate normal data using the following code:

```
set.seed(1)
n <- 100
A <- rnorm(n)
B <- A + rnorm(n)
C <- B + rnorm(n)
D <- B + rnorm(n)
X <- cbind(A, B, C, D)
S <- cov(X)
```

Then regress each element of $\mathbf{x}$ on the others. We are going to use these regression fits to estimate Θ ; however, let us carry out a simple model selection procedure first, in which we drop any covariates that are not significant at the $\alpha = 0.05$ level. Then refit the model with only the significant covariates, and use $\hat{\beta}$ and $\hat{\sigma}^2$ to fill in the appropriate elements of $\hat{\Theta}$; set $\beta_j = 0$ if the term was not included in the model. As an answer, you only need to provide $\hat{\Theta}$, not the full summary of all the regression fits.

- (d) Does your estimate of Θ from (c) reflect the correct conditional independence relationships among A, B, C, and D? Comment briefly.
- (e) Letting $\mathbf{x}^\top = [A \ B \ C \ D]$, show that the data generating mechanism of the above code results in $\mathbf{x}$ having a multivariate normal distribution, and calculate the true precision matrix Θ^* .
- (f) We now have two estimators for Θ : $\mathbf{S}^{-1}$ and the estimator from part (c). Which one is more accurate (for this particular data set)? Quantify the overall accuracy using $\|\hat{\Theta} - \Theta^*\|_F$.

- (g) One downside of (c) is that it produces an asymmetric estimate $\hat{\Theta}$. One simple remedy is to use $\tilde{\Theta} = \frac{1}{2}\hat{\Theta} + \frac{1}{2}\hat{\Theta}^\top$ instead. Does this symmetrized estimate improve accuracy?
2. *Power calculation using the noncentral χ^2 distribution.* Suppose there is a latent random variable of interest Z that is continuously distributed between 0 and 1, but we observe only which of 10 bins it falls into: $(0, 0.1), (0.1, 0.2), \dots, (0.9, 1.0)$. Thus, we observe $\mathbf{x}$, a 10-dimensional random vector of counts corresponding to the bins, with n denoting the total count. This problem involves attempting to test the null hypothesis that all bins are equally likely by assuming that $\mathbf{x}$ (approximately) follows a multivariate normal distribution.
- Using the mean and variance of a multinomial distribution under the null, provide a test statistic $T(\mathbf{x})$ that follows an approximate χ^2 distribution (i.e., that would follow a χ^2 distribution if $\mathbf{x}$ were multivariate normal with this mean and variance).
 - Now suppose that $Z \sim \text{Beta}(1, 2)$. Create a plot overlaying two beta distributions: this one and the one corresponding to the null hypothesis.
 - Under the alternative distribution specified in (b), $T(\mathbf{x})$ will no longer follow an ordinary χ^2 distribution, but instead a noncentral χ^2 distribution. Create a plot overlaying two χ^2 densities, one of the null hypothesis and the other with a noncentrality parameter of 10. Use the number of degrees of freedom appropriate to this problem.
 - Derive the noncentrality parameter for the distribution of $T(\mathbf{x})$ under the alternative hypothesis. Note that there is a problem with this calculation: the alternative hypothesis affects both the mean and the variance. For the purposes of this calculation, only account for its effect on the mean — assume that the variance is unchanged. Write an R function implementing this calculation. Turn in a .R file that defines a single function and nothing else. I should be able to run `ncp(n)` after sourcing this file, and it should return the correct noncentrality parameter.
 - Create a plot of n versus power (assuming an $\alpha = 0.05$ significance threshold), where n ranges from 10 to 100. Note that this calculation uses both the null distribution from (a) and the alternative distribution you derived in (d).
 - In the power calculation above, there are two potential issues: (a) the true distribution is multinomial, not multivariate normal, and (b) we ignored the impact of the alternative distribution on variance when calculating the noncentrality parameter. Carry out a simulation to compare the true power to our approximation. Draw samples of size $n = 50$ from the multinomial distribution and carry out the χ^2 test that you derived above (don't "correct" for continuity). Calculate the average power over $N = 10,000$ replications.
 - Briefly, how do these two approaches compare? In particular, suppose you were had to perform a power calculation like this for a real-world project: which approach would you use? Why?
3. *Bounded in probability vs convergence in distribution.* Prove that if a sequence of random variables X_n converges in distribution, then X_n is bounded in probability. Note: this is true for random vectors as well, of course, but main ideas of this proof come across more clearly in the scalar case.
4. *Convergence in quadratic mean and probability.* This problem involves proving the theorems from the slide titled "Convergence in mean vs convergence in probability" from the "Modes of convergence" lecture. Hint: Markov's inequality.
- Prove that if $\mathbf{x}_n \xrightarrow{r} \mathbf{x}$ for some $r > 0$, then $\mathbf{x}_n \xrightarrow{P} \mathbf{x}$.
 - Prove that if $\mathbf{a} \in \mathbb{R}^d$, then $\mathbf{x}_n \xrightarrow{\text{qm}} \mathbf{a}$ if and only if $\mathbb{E}\mathbf{x}_n \rightarrow \mathbf{a}$ and $\mathbb{V}\mathbf{x}_n \rightarrow \mathbf{0}_{d \times d}$.