Advanced Methods in Biostatistics I

Lecture 12

Martin Lindquist

October 5, 2017

Multivariate normality

- In order to perform inference on the linear model, we typically assume the response variable follows a multivariate normal distribution.
- Today we will continue our discussion of the multivariate normal distribution, and link it to linear models.

Multivariate normality

- Suppose that $\mathbf{y}_1$ and $\mathbf{y}_2$ are jointly multivariate normal with $\Sigma_{12} \neq 0$, i.e.

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right).$$

- Further assume that Σ_{11} is nonsingular.

Marginal distributions

- The marginal distributions for $\mathbf{y}_1$ and $\mathbf{y}_2$ are

$$\mathbf{y}_1 \sim N_p(\boldsymbol{\mu}_1, \Sigma_{11})$$

and

$$\mathbf{y}_2 \sim N_q(\boldsymbol{\mu}_2, \Sigma_{22}),$$

respectively.

Conditional distributions

- The conditional distribution of $\mathbf{y}_2 \mid \mathbf{y}_1$ is $N(\boldsymbol{\mu}_{2|1}, \Sigma_{2|1})$, where

$$\boldsymbol{\mu}_{2|1} = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$$

and

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}'.$$

Block matrix inversion

Theorem

If $\mathbf{A}$ and $\mathbf{D}$ are symmetric and all inverses exist,

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{E}^{-1} & -\mathbf{E}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{E}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{E}^{-1}\mathbf{B}\mathbf{D}^{-1} \end{pmatrix},$$

where $\mathbf{E} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$.

Conditional distributions

- Now consider the partitioned variance matrix.

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

- The upper diagonal element of Σ^{-1}, which we will denote K_{11} is given by $(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}')^{-1}$.
- Note $K_{11}^{-1} = Var(\mathbf{X}_1 \mid \mathbf{X}_2)$.

Conditional distributions

- Suppose that $\mathbf{y}_1 = (y_{11}\ y_{12})'$.
- Recall that independence and absence of correlation are equivalent for the multivariate normal.
- Hence, the random variables y_{11} and y_{12} are independent if the $(1,2)^{th}$ element of Σ is 0.

Conditional distributions

- Claim: The random variables y_{11} and y_{12} are independent conditional on $\mathbf{y}_2$ if the $(1,2)^{th}$ element of Σ^{-1} is 0.
- Note if y_{11} and y_{12} are independent conditional on $\mathbf{y}_2$ then the 2×2 matrix K_{11}^{-1} will be a diagonal matrix.
- Hence, it must hold that K_{11} is a diagonal matrix.

Conditional distributions

- The fact that we chose to analyze the first two elements conditional on the others was arbitrary.
- Hence, whether or not a particular off diagonal element of Σ^{-1} is zero determines the conditional independence of those random variables given the rest of the elements in the random vector.

Gaussian graphical models

- This forms the basis of so-called Gaussian graphical models.
- For undirected Gaussian graphical models a missing edge (i, j) in the underlying graph G corresponds to the conditional independence of y_i and y_j are independent conditional on all other elements y_k.
- The graph defined by ascertaining which elements of Σ^{-1} are zero is called a conditional independence graph.

Normal model

- Let's continue working with the linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

- However, we will now add an additional assumption to the model.
- Let us begin by assuming that $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$.
- Now our model can be written: $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$

Normal likelihood

- Under the normality assumption, it is now possible to obtain maximum likelihood estimates for the model parameters.
- Recall the likelihood is the joint density of **y**, written $L(\beta, \sigma^2)$.
- We seek values of β and σ^2 that maximize the likelihood for the observed data.

Normal likelihood

- It is often beneficial to work with the log-likelihood,

$$\log(L(\beta, \sigma^2))$$

  which achieves a maximum for the same values of β and σ^2 as the likelihood.

- Alternatively, we can seek to minimize the deviance, which is $-2\log(L(\beta, \sigma^2))$.

Normal likelihood

- Let $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$, then we can write the log-likelihood as follows:

$$\begin{aligned} & \log(L(\beta, \sigma^2) \\ = & -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) \\ = & -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta) \end{aligned}$$

- Taking the derivatives and setting equal to zero we can derive the MLEs.

Normal likelihood

- Taking the derivatives gives the following results:

$$\begin{aligned}
\frac{\partial}{\partial \beta} \log(L(\beta, \sigma^2) &= \frac{2}{2\sigma^2}(\mathbf{X}\mathbf{y}' - \mathbf{X}'\mathbf{X}\beta) \\
\frac{\partial}{\partial \sigma^2} \log(L(\beta, \sigma^2) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)
\end{aligned}$$

- Setting these equations equal to 0 and solving for β and σ^2 gives us the MLEs.

Normal likelihood

- The MLEs of β and σ^2 are given by:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

and

$$\hat{\sigma}^2 = ||\mathbf{y} - \mathbf{X}\beta||^2/n.$$

- Note $\hat{\beta}$ is the same as the least-squares estimator.
- The estimator $\hat{\sigma}^2$ is the average of the squared residuals.
- We often use the unbiased estimate given by

$$s^2 = ||\mathbf{y} - \mathbf{X}\beta||^2/(n-p).$$

Sufficient statistics

Definition

A statistic $T(\mathbf{y})$ is a sufficient statistic for θ if the conditional distribution of the sample $\mathbf{y}$ given the value of $T(\mathbf{y})$ does not depend on θ.

Neyman-Fisher Factorization Theorem

Theorem

The statistic T is sufficient for θ if and only if functions g and h can be found such that

$$f(\mathbf{y}|\theta) = h(\mathbf{y})g(\theta, T(\mathbf{y})).$$

Properties

Theorem

The estimators $\hat{\beta}$ and $\hat{\sigma}^2$ are sufficient statistics for β and σ^2.

Properties

Theorem

$$\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

Normal likelihood

- Now consider the case where $\mathbf{Y} \sim N(\mathbf{X}\beta, \Sigma)$ with known Σ.
- Then the log-likelihood is proportional to

$$\log(L(\beta, \sigma^2) \quad \propto \quad -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)'\Sigma^{-1}(\mathbf{y} - \mathbf{X}\beta)$$

Normal likelihood

- Taking the derivatives and setting equal to zero we can derive the MLE:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}.$$

- This is the so-called generalized least squares estimate.

Bayesian statistics

- In the next example we assume some familiarity with Bayesian calculations and inference.

- In Bayesian analysis, one chooses a density $f(\theta)$, the prior, that expresses our beliefs about θ before we see any data.
- Then one chooses a statistical model $f(\mathbf{y}|\theta)$, the likelihood, that reflects our belief about $\mathbf{y}$ given θ.
- After observing $\mathbf{y}$, we update our beliefs and calculate the posterior distribution $f(\theta|\mathbf{y})$.

Bayes calculations

- Updating is is done as follows:

$$\begin{aligned} f(\text{Param}|\text{Data}) &= \frac{f(\text{Param}, \text{Data})}{f(\text{Data})} \\ &\propto f(\text{Data}|\text{Param}) f(\text{Param}) \\ &= \text{Likelihood} \times \text{Prior}. \end{aligned}$$

- The posterior distribution is used for subsequent inference.

Example

- Suppose that $\mathbf{y} \mid \mu \sim N(\mu \mathbf{J}_n, \sigma^2 \mathbf{I})$ and $\mu \sim N(\mu_0, \tau^2)$ where $\mathbf{y}$ is $n \times 1$ and μ is a scalar.
- The normal distribution placed on μ is the "prior" and the terms μ_0 and τ^2 are assumed to be known.
- For this example, let's assume that σ^2 is also known.

Example

- The goal is to calculate $f(\mu \mid \mathbf{y})$, the posterior distribution.
- This is done by multiplying the prior times likelihood.

Example

- Retaining only terms involving μ we have that the log of $f(\mu \mid \mathbf{y})$ is given by:

$$\begin{aligned}
\log(f(\mathbf{y} \mid \mu)) \quad &+ \quad \log(f(\mu)) \\
&\propto \quad -\frac{1}{2\sigma^2}||\mathbf{y} - \mu \mathbf{J}_n||^2 - \frac{1}{2\tau^2}(\mu - \mu_0)^2 \\
&\propto \quad \mu n \bar{y}/\sigma^2 - \mu^2 n/(2\sigma^2) - \mu^2/(2\tau^2) + \mu\mu_0/\tau^2 \\
&= \quad \mu\left(\frac{\bar{y}}{\sigma^2/n} + \frac{\mu_0}{\tau^2}\right) - \frac{\mu^2}{2}\left(\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}\right)
\end{aligned}$$

Example

- This can be recognized as the log density of a normally distributed random variable with variance

$$Var(\mu \mid \mathbf{y}) = \left(\frac{1}{\sigma^2/n} + \frac{1}{\tau^2} \right)^{-1} = \frac{\tau^2 \sigma^2/n}{\sigma^2/n + \tau^2}$$

  and mean

$$E[\mu \mid \mathbf{y}] = \left(\frac{1}{\sigma^2/n} + \frac{1}{\tau^2} \right)^{-1} \left(\frac{\bar{y}}{\sigma^2/n} + \frac{\mu_0}{\tau^2} \right).$$

Example

- Note, we can express the expected value as follows:

$$E[\mu \mid \mathbf{y}] = p\bar{y} + (1 - p)\mu_0$$

where

$$p = \frac{\tau^2}{\tau^2 + \sigma^2/n}.$$

- Thus $E[\mu \mid \mathbf{y}]$ is a mixture of the empirical mean and the prior mean.

Example

- How much the means are weighted depends on the ratio of the variance of the mean (σ^2/n) and the prior variance (τ^2).
- As we collect more data ($n \to \infty$), or if the data is not noisy ($\sigma \to 0$) or we have a lot of prior uncertainty ($\tau \to \infty$) the empirical mean dominates.
- In contrast as we are more certain a priori ($\tau \to 0$) the prior mean dominates.