Classical Inference for Gaussian Linear Models

Surya Tokdar

Gaussian linear model

- Data (z_i, Y_i), $i = 1, \cdots, n$, $\dim(z_i) = p$.
- Model
$$Y_i = z_i^T \beta + \epsilon_i, \quad \epsilon_i \overset{\text{IID}}{\sim} N(0, \sigma^2).$$
- Parameters: $\beta \in \mathbb{R}^p$, $\sigma^2 > 0$.
- Inference needed on $\eta = a^T \beta$
- Useful model for many types of analyses.

Matrix-vector notation

- Response vector, design matrix and error vector
$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad Z = \begin{pmatrix} z_1^T \\ z_2^T \\ \vdots \\ z_n^T \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$
- Model: $Y = Z\beta + \epsilon$, $\epsilon \sim N_n(0, \sigma^2 I_n)$.

Example 1

- Food expenditures $Y_1, \cdots, Y_n$.
- "Population" average μ and variability σ^2
- Model $Y_i \overset{\text{IID}}{\sim} N(\mu, \sigma^2)$, $\beta \in \mathbb{R}$, $\sigma^2 > 0$.
- This is a Gaussian linear model with $p = 1$, $z_i = 1$ and $\beta = \mu$

Example 2

- Body weight gains of n_1 rats on high protein: $H_1, \cdots, H_{n_1}$
- Same for n_2 rats on low protein: $L_1, \cdots, L_{n_2}$.
- Model
  - $H_i \overset{\text{IID}}{\sim} N(\mu_1, \sigma^2)$, $L_j \overset{\text{IID}}{\sim} N(\mu_2, \sigma^2)$, H_i's and L_j's independent
  - $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma^2 > 0$
- Gaussian linear model with $p = 2$, $n = n_1 + n_2$, $\beta = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$

$$Y = \begin{pmatrix} H_1 \\ \vdots \\ H_{n_1} \\ \hline L_1 \\ \vdots \\ L_{n_2} \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ \hline 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{matrix} \updownarrow n_1 \\ \updownarrow n_2 \end{matrix}$$

Example 3

- n subjects, males and females
- randomly assigned to treatment (drug) or control (placebo)
- Y_i = improvement in condition (sleep hours) of subject i

Example 3 as Gaussian linear model

- Use model $Y_i = z_i'\beta + \epsilon_i$, $\epsilon_i \overset{\text{IID}}{\sim} N(0, \sigma^2)$,
- where $p = 4$ and $z_i = (z_{i1}, z_{i2}, z_{i3}, z_{i4})'$ with

$$
\begin{aligned}
z_{i1} &= I(i\text{-th subject is F and gets T})\\
z_{i2} &= I(i\text{-th subject is F and gets C})\\
z_{i3} &= I(i\text{-th subject is M and gets T})\\
z_{i4} &= I(i\text{-th subject is M and gets C})
\end{aligned}
$$

- Let n_{FT} be the number of subjects who are F and get T. Similarly define n_{FC}, n_{MT} and n_{MC}.

Example 3 design matrix

$$
Z = \begin{pmatrix}
1 & 0 & 0 & 0\\
\vdots & \vdots & \vdots & \vdots\\
1 & 0 & 0 & 0\\
\hline
0 & 1 & 0 & 0\\
\vdots & \vdots & \vdots & \vdots\\
0 & 1 & 0 & 0\\
\hline
0 & 0 & 1 & 0\\
\vdots & \vdots & \vdots & \vdots\\
0 & 0 & 1 & 0\\
\hline
0 & 0 & 0 & 1\\
\vdots & \vdots & \vdots & \vdots\\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{matrix}
\updownarrow n_{\text{FT}}\\
\updownarrow n_{\text{FC}}\\
\updownarrow n_{\text{MT}}\\
\updownarrow n_{\text{MC}}
\end{matrix}
$$

Example 3 treatment effects

- Treatment effect for females: $\eta_F = \beta_1 - \beta_2$
- Treatment effect for males: $\eta_M = \beta_3 - \beta_4$
- Treatment effect difference: $\eta = \eta_F - \eta_M = \beta_1 - \beta_2 - \beta_3 + \beta_4$

Example 4

- n subjects
- Randomly assigned to treatment or control
- Y_i = improvement in condition for subject i
- Likely to depend on subject's age

Example 4 and Gaussian linear model

- $\text{treat}_i = 1$ for treatment and 0 for control
- Model:

$$Y_i = \beta_1 + \beta_2\,\text{treat}_i + \beta_3\,\text{age}_i + \beta_4\,\text{treat}_i \times \text{age}_i + \epsilon_i,$$

  $\epsilon_i \overset{\text{IID}}{\sim} N(0, \sigma^2)$
- i.e., $p = 4$, $z_i = (1, \text{treat}_i, \text{age}_i, \text{treat}_i \times \text{age}_i)$

Example 4 and quantities of interest

1. Expected improvement at age 30, receiving treatment:

$$a = (1, 1, 30, 30)^T$$

2. Treatment effect at age 30, i.e., expected additional improvement due to treatment at age 30

$$a = (0, 1, 0, 30)^T$$

3. Difference in treatment effects between age 20 and 30

$$a = (0, 0, 0, -10)^T$$

ML theory: the likelihood function

- Model: $Y \sim N_n(Z\beta, \sigma^2 I)$
- Observe $Y = y$ with $y = (y_1, \cdots, y_n)^T$.
- Log-likelihood:

$$\ell_y(\beta, \sigma^2) = \text{const} - \frac{n}{2}\log\sigma^2 - \frac{(y - Z\beta)^T(y - Z\beta)}{2\sigma^2}$$

MLE

- First order conditions:

$$0 = \frac{\partial}{\partial\beta}\ell_y(\beta, \sigma^2) = \frac{Z^T(y - Z\beta)}{\sigma^2}$$

$$0 = \frac{\partial}{\partial\sigma^2}\ell_y(\beta, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{(y - Z\beta)^T(y - Z\beta)}{2\sigma^4}$$

- $\hat\beta_{\text{MLE}} = (Z^TZ)^{-1}Z^Ty =: \hat\beta_{\text{LS}}$
- $\hat\sigma^2_{\text{MLE}} = \frac{(y - Z\hat\beta_{\text{LS}})^T(y - Z\hat\beta_{\text{LS}})}{n}$.
- Notation: $s^2_{y|z} = \frac{(y - Z\hat\beta_{\text{LS}})^T(y - Z\hat\beta_{\text{LS}})}{n-p}$, i.e., $\hat\sigma^2_{\text{MLE}} = \frac{n-p}{n}s^2_{y|z}$.

Least squares interpretation

- For any β

$$\begin{aligned}(y - Z\beta)^T(y - Z\beta) &= \|y - Z\beta\|^2 \\ &= \|y - Z\hat\beta_{\text{LS}}\|^2 + \|Z\hat\beta_{\text{LS}} - Z\beta\|^2 + 2(y - Z\hat\beta_{\text{LS}})^T(Z\hat\beta_{\text{LS}} - Z\beta) \\ &= \|y - Z\hat\beta_{\text{LS}}\|^2 + \|Z\hat\beta_{\text{LS}} - Z\beta\|^2 + 0\end{aligned}$$

- $\hat\beta_{\text{LS}}$ is the least-squares estimate of β

Profile log-likelihood of β

- For any β, $\ell_y(\beta, \sigma^2)$ is maximized in σ^2 at

$$\begin{aligned}\hat\sigma^2(\beta) &= \frac{(y - Z\beta)^T(y - Z\beta)}{n} \\ &= \frac{(n-p)s^2_{y|z} + (\beta - \hat\beta_{\text{LS}})^T(Z^TZ)(\beta - \hat\beta_{\text{LS}})}{n}\end{aligned}$$

- So the profile log-likelihood in β is

$$\begin{aligned}\ell^*_y(\beta) = \ell_y(\beta, \hat\sigma^2(\beta)) &= \text{const} - \frac{n}{2}\log\hat\sigma^2(\beta) - \frac{n}{2} \\ &= \text{const} - \frac{n}{2}\log\left\{1 + \frac{(\beta - \hat\beta_{\text{LS}})^T(Z^TZ)(\beta - \hat\beta_{\text{LS}})}{(n-p)s^2_{y|z}}\right\}\end{aligned}$$

ML intervals for $\eta = a^T\beta$

- Additional calculations show the profile log-likelihood in η is

$$\ell^*_y(\eta) = \text{const} - \frac{n}{2}\log\left\{1 + \frac{1}{(n-p)s^2_{y|z}} \times \frac{(\eta - a'\hat\beta_{\text{LS}})^2}{a^T(Z^TZ)^{-1}a}\right\}$$

- So ML intervals for η are of the form

$$a^T\hat\beta_{\text{LS}} \mp c_n\frac{s_{y|z}}{\sqrt{n_a}}$$

where $n_a = 1/\{a^T(Z^TZ)^{-1}a\}$, with thresholds $c_n > 0$

ML confidence interval for $\eta = a^T\beta$

- Let F_k denote the CDF of $t(k)$ distribution
- Notation: $z_k(\alpha) = F_k^{-1}(1 - \alpha/2)$
- $100(1-\alpha)\%$ ML confidence interval for $\eta = a^T\beta$ is

$$a^T\hat\beta_{\text{LS}} \mp z_{n-p}(\alpha)\frac{s_{y|z}}{\sqrt{n_a}}$$

- Due to the following fundamental theorem

A fundamental result

Theorem. *Let $Y \sim N_n(Z\beta, \sigma^2 I)$. Define $H = Z(Z'Z)^{-1}Z'$ and $\hat{\epsilon} = Y - Z\hat{\beta}_{LS} = (I_n - H)Y$. Then*

1. $\hat{\beta}_{LS} \sim N_p(\beta, \sigma^2(Z'Z)^{-1})$.
2. $\hat{\epsilon} \sim N_n(0, \sigma^2(I_n - H))$
3. $\hat{\beta}_{LS}$ *and* $\hat{\epsilon}$ *are independent.*
4. $\frac{1}{\sigma^2}\hat{\epsilon}'\hat{\epsilon} \sim \chi^2(n-p)$.

Coverage calculation

- Notation: $C_c(y) = a^T\hat{\beta}_{LS} \mp c\frac{s_{y|z}}{\sqrt{n_a}}$

$$\begin{aligned}\gamma((\beta,\sigma^2), C_c) &= P_{[Y|\beta,\sigma^2]}(a^T\beta \in C_c(Y)) \\ &= P_{[Y|\beta,\sigma^2]}\left(-c \le \frac{a^T\hat{\beta}_{LS} - a^T\beta}{s_{y|z}/\sqrt{n_a}} \le c\right) \\ &= P_{[Y|\beta,\sigma^2]}(-c \le T \le c)\end{aligned}$$

- By theorem $T \sim t(n-p)$ when $Y \sim N_n(Z\beta, \sigma^2 I_n)$
- And so $\gamma((\beta,\sigma^2), C_c) = 2F_{n-p}(c) - 1$
- For $c = z_{n-p}(\alpha)$ this number equals $1 - \alpha$

ML testing

- $H_0 : a^T\beta = \eta_0$ where η_0 is a fixed number.
- ML test $\delta_c(y)$: reject H_0 if $\eta_0 \notin C_c(y)$
- Null set: $\Theta_0 = \{(\beta,\sigma^2) : a^T\beta = \eta_0\}$ – note this is a set, not a single point.
- Size of δ_c is $1 - \gamma(C_c) = 2(1 - F_{n-p}(c))$. [Prove in HW]
- In particular $\delta_{z_{n-p}(\alpha)}$ has size α.

One sided hypothesis

- $H_0 : a^T\beta \le \eta_0$ where η_0 is a fixed number
- ML test $\delta_c(y)$: reject H_0 if $(-\infty, \eta_0] \cap C_c(y) = \emptyset$
- Same as rejecting H_0 when $\eta_0 < a^T\hat{\beta}_{LS} - cs_{y|z}/\sqrt{n_a}$
- Size of δ_c is $1 - F_{n-p}(c)$
- $\delta_{z_{n-p}(\alpha)}$ has size $\alpha/2$
- Can do the same for the other one-sided case: $H_0 : a^T\beta \ge \eta_0$.

Example: Chick Weight

- 50 chicks assigned to one of 4 protein diets
- One body weight measurement from each chick between 1 and 21 days after birth
- Data on (log) body weight, diet and time of measurement
- Model

$$\text{weight} = \beta_1 + \beta_2\text{Diet}_2 + \beta_3\text{Diet}_3 + \beta_4\text{Diet}_4 + \beta_5\text{Time} + \epsilon$$