Thanks to D11221003

Exponential Family

- Suppose $Y_1, \ldots, Y_n$ are independent random variables.
- Let $f(y_i; \theta_i, \phi)$ be the Probability Mass Function (PMF) or Probability Density Function (PDF) of Y_i, where ϕ is a scale parameter.
- If we can write
$$f(y_i; \theta_i, \phi) = \exp\left(y_i\theta_i - b(\theta_i)\frac{1}{a(\phi)} + c(y_i, \phi)\right),$$
then we call the PMF or the PDF $f(y_i; \theta_i, \phi)$ an exponential family.

Problem 1.

Find the form of GLM for the following distributions, and show the resonable link function:

1. Normal distributions
2. Inverse Gaussian
3. Binomial distribution
4. Poisson distribution
5. Gamma distribution
6. Beta

Normal Distribution

Assume $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$. Then, $E(Y_i) = \mu_i$ and σ is a scale parameter. The Probability Density Function (PDF) is given by

$$\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y_i-\mu_i)^2}{2\sigma^2}} = \exp\left\{\frac{y_i\mu_i - \mu_i^2/2}{\sigma^2} - \left(\frac{1}{2}\log(2\pi\sigma^2) - \frac{y_i^2}{2\sigma^2}\right)\right\}.$$

Thus, use

- $\theta_i = \mu_i$,
- $b(\theta_i) = \frac{\theta_i^2}{2}$,
- $\phi = \sigma^2$,
- $a(\phi) = \phi$,
- $c(y_i, \phi) = -\frac{1}{2}\log(2\pi\phi) - \frac{y_i^2}{2\phi}$.

$$E[Y_{ij} \mid x_{ij}] = h(x_{ij}^T\beta) = x_{ij}^T\beta$$

Inverse Gaussian Distribution

Let us rewrite the probability density function (pdf) of the Inverse Gaussian distribution with parameters μ_i and λ:

$$f(y_i; \mu_i, \lambda) = \left(\frac{\lambda}{2\pi y_i^3}\right)^{1/2} \exp\left\{-\frac{\lambda(y_i - \mu_i)^2}{2\mu_i^2 y_i}\right\}, \quad y > 0$$

in the following form:

$$\begin{aligned} f(y_i; \mu_i, \lambda) &= \exp\left\{-\frac{\lambda(y_i - \mu_i)^2}{2\mu_i^2 y_i} + \frac{1}{2}\ln\left(\frac{\lambda}{2\pi y_i^3}\right)\right\} \\ &= \exp\left\{\frac{-\frac{1}{2\mu_i^2} y_i + \frac{1}{\mu_i}}{\frac{1}{\lambda}} + \left(\frac{1}{2}\ln\frac{\lambda}{2\pi y_i^3} - \frac{\lambda}{2y_i}\right)\right\} \end{aligned}$$

Now, let's identify the exponential family components:

- Canonical parameter: $\theta_i = -\frac{1}{2\mu_i^2}$
- $b(\theta_i) = -\frac{1}{\mu_i} = -(-2\theta_i)^{\frac{1}{2}}$
- $\phi = \lambda$
- $a(\phi) = \frac{1}{\phi}$
- $c(y_i, \phi) = \frac{1}{2}\ln\frac{\phi}{2\pi y_i^3} - \frac{\phi}{2y_i}$

Thus, the Inverse Gaussian distribution can be shown to be a member of the exponential family.

$$E[Y_{ij} \,|\, x_{ij}] = h(x_{ij}^T\beta) = \exp(x_{ij}^T\beta) > 0$$

Binomial Distribution

Assume $Y_i \sim \text{Bin}(n_i, p_i)$. Then, $E(Y_i) = n_i p_i$. The Probability Mass Function (PMF) is given by

$$\binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i - y_i} = \exp\left\{y_i \log\left(\frac{p_i}{1-p_i}\right) + n_i\log(1-p_i) - \log\binom{n_i}{y_i}\right\}.$$

Thus,

- $\theta_i = \log\left(\frac{p_i}{1-p_i}\right)$,
- $b(\theta_i) = n_i\log(1 + e^{\theta_i})$,
- $\phi = 1$, $a(\phi) = 1$,
- $c(y, \phi) = -\log\binom{n_i}{y_i}$.

$$E[Y_{ij} \,|\, x_{ij}] = h(x_{ij}^T\beta) = n\frac{\exp(x_{ij}^T\beta)}{1 + \exp(x_{ij}^T\beta)} \in (0, n)$$

Poisson Distribution

Assume $Y_i \sim \text{Poisson}(\lambda_i)$. Then, $E(Y_i) = \lambda_i$. The Probability Mass Function (PMF) is given by

$$\frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} = \exp\left\{y_i \log(\lambda_i) - \lambda_i - \log(y_i!)\right\}.$$

Thus,

- $\theta_i = \log(\lambda_i)$,
- $b(\theta_i) = e^{\theta_i}$,
- $\phi = 1$, $a(\phi) = 1$,
- $c(y_i, \phi) = -\log(y_i!)$.

$$E[Y_{ij} \,|\, x_{ij}] = h(x_{ij}^T \beta) = \exp(x_{ij}^T \beta) > 0$$

Gamma Distribution

Assume $x_i \sim \Gamma(\alpha, \beta_i)$, where β_i is unknown. Then, $E(x_i) = \frac{\alpha}{\beta_i}$. The Probability Mass Function (PMF) is given by

$$\frac{\beta_i^{\alpha} x_i^{\alpha-1} e^{-\beta_i x_i}}{\Gamma(\alpha)} = \exp\left\{\alpha \log x_i + \alpha \log(\beta_i) - \log(\Gamma(\alpha)) - \log(x_i) - \beta_i x_i\right\}.$$

Assuming α is known, if we choose $y_i = x_i$, then

- $\theta_i = -\beta_i$ ($\theta_i < 0$),
- $b(\theta_i) = -\alpha \log(-\theta_i)$,
- $\phi = 1$, and $a(\phi) = 1$.

Remark: We can also choose $y_i = -x_i$ and $\theta_i = \beta_i$. In this case, $b(\theta_i) = -\alpha \log(\theta_i)$.

$$E[Y_{ij} \,|\, x_{ij}] = h(x_{ij}^T \beta) = \exp(x_{ij}^T \beta) > 0$$

Problem 2.

Paper Summarization for "Sliced Inverse Regression for Dimension Reduction"[a]

[a]Li, Ker-Chau. "Sliced Inverse Regression for Dimension Reduction." Journal of the American Statistical Association 86, no. 414 (1991): 316–27. https://doi.org/10.2307/2290563.

1. The paper proposes a data-analytic tool, sliced inverse regression (SIR), for reducing the dimension of the input variable x without going through any parametric or nonparametric model-fitting process.
2. The paper explores the inverse view of regression, that is, instead of regressing the output variable y against the input variable x, it regresses x against y. The paper shows that under a suitable condition, the inverse regression curve $E(x \mid y)$ will fall into the effective dimension reduction (e.d.r.) space, which is the linear space generated by the unknown row vectors β in the model $y = f(\beta_1 x, \cdots, \beta_k x, \epsilon)$.
3. The paper develops a simple algorithm for SIR, which consists of standardizing x, slicing the range of y into several intervals, computing the slice means of x within each interval, and conducting a principal component analysis on the covariance matrix of the slice means to estimate the e.d.r. directions.
4. The paper investigates the theoretical properties of SIR under a design condition that is satisfied by elliptically symmetric distributions. The paper shows that SIR yields $\sqrt{n}-$ consistent estimates for the e.d.r. directions and provides a chi-squared statistic to test whether an estimated direction is spurious.

—

Reference:[1]

[1]Wikipedia contributors, "Sliced inverse regression," Wikipedia, The Free Encyclopedia, `https://en.wikipedia.org/wiki/Sliced_inverse_regression` (accessed October 23, 2023).

Problem 3.

Binary response ($Y_{ij} = 0/1$)

Logistic regression model:

1. $\mathbb{E}[Y_{ij} \,|\, x_{ij}] = h(x_{ij}^T\beta) = \frac{\exp(x_{ij}^T\beta)}{1+\exp(x_{ij}^T\beta)}$
2. $\mathrm{Var}[Y_{ij} \,|\, x_{ij}] = \mathbb{E}[Y_{ij} \,|\, x_{ij}](1 - \mathbb{E}[Y_{ij} \,|\, x_{ij}])$
3. $\mathbf{Cor}[Y_{ij}, Y_{ik} \,|\, x_{ij}, x_{ik}] = \alpha$
4. odd ratio: $OR(Y_{ij}, Y_{ik}) = \frac{P(Y_{ij}=1,Y_{ik}=1)P(Y_{ij}=0,Y_{ik}=0)}{P(Y_{ij}=1,Y_{ik}=0)P(Y_{ij}=0,Y_{ik}=1)}$

$$
\begin{aligned}
\alpha = \mathbf{Cor}[Y_{ij}, Y_{ik} \,|\, x_{ij}, x_{ij}] &= \frac{\mathbb{E}[(Y_{ij} - \mathbb{E}[Y_{ij} \,|\, x_{ij}])(Y_{ik} - \mathbb{E}[Y_{ik} \,|\, x_{ik}])]}{\sigma_{Y_{ij} \,|\, x_{ij}} \sigma_{Y_{ik} \,|\, x_{ik}}} \\
&= \frac{\mathbb{E}[(Y_{ij} - \mathbb{E}[Y_{ij} \,|\, x_{ij}])(Y_{ik} - \mathbb{E}[Y_{ik} \,|\, x_{ik}])]}{\sqrt{\mathrm{Var}[Y_{ij} \,|\, x_{ij}]}\sqrt{\mathrm{Var}[Y_{ik} \,|\, x_{ik}]}} \\
&= \frac{\mathbb{E}[(Y_{ij} - \mathbb{E}[Y_{ij} \,|\, x_{ij}])(Y_{ik} - \mathbb{E}[Y_{ik} \,|\, x_{ik}])]}{\sqrt{\mathbb{E}[Y_{ij} \,|\, x_{ij}](1 - \mathbb{E}[Y_{ij} \,|\, x_{ij}])}\sqrt{\mathbb{E}[Y_{ik} \,|\, x_{ik}](1 - \mathbb{E}[Y_{ik} \,|\, x_{ik}])}} \\
&= \frac{\mathbb{E}[Y_{ij}Y_{ik} - Y_{ij}h(x_{ik}^T\beta) - Y_{ik}h(x_{ik}^T\beta) + h(x_{ij}^T\beta)h(x_{ik}^T\beta)]}{\sqrt{h(x_{ij}^T\beta)(1 - h(x_{ij}^T\beta))}\sqrt{h(x_{ik}^T\beta)(1 - h(x_{ik}^T\beta))}} \\
&= \frac{\mathbb{E}[Y_{ij}Y_{ik}] - h(x_{ij}^T\beta)h(x_{ik}^T\beta)}{\sqrt{h(x_{ij}^T\beta)(1 - h(x_{ij}^T\beta))}\sqrt{h(x_{ik}^T\beta)(1 - h(x_{ik}^T\beta))}} \\
&= \frac{P(Y_{ij} = 1, Y_{ik} = 1) - h(x_{ij}^T\beta)h(x_{ik}^T\beta)}{\sqrt{h(x_{ij}^T\beta)(1 - h(x_{ij}^T\beta))}\sqrt{h(x_{ik}^T\beta)(1 - h(x_{ik}^T\beta))}} \in [0, 1]
\end{aligned}
$$

....

Denoting $P_{ij,ik} = P(Y_{ij} = 1, Y_{ik} = 1)$.

$$
\begin{aligned}
\gamma = OR(Y_{ij}, Y_{ik}) &= \frac{P(Y_{ij} = 1, Y_{ik} = 1)P(Y_{ij} = 0, Y_{ik} = 0)}{P(Y_{ij} = 1, Y_{ik} = 0)P(Y_{ij} = 0, Y_{ik} = 1)} \\
&= \frac{P_{ij,ik}[1 - (h(x_{ij}^T\beta) - P_{ij,ik}) - (h(x_{ik}^T\beta) - P_{ij,ik}) - P_{ij,ik}]}{[h(x_{ij}^T\beta) - P_{ij,ik}][h(x_{ik}^T\beta) - P_{ij,ik}]} \\
&= \frac{P_{ij,ik}[1 - h(x_{ij}^T\beta) - h(x_{ik}^T\beta) + P_{ij,ik}]}{[h(x_{ij}^T\beta) - P_{ij,ik}][h(x_{ik}^T\beta) - P_{ij,ik}]}
\end{aligned}
$$

Solve for $P_{ij,ik}$ using γ and $h(\cdot)$.

Problem 4.

Describe

1. how to conduct the EM (Expectation-Maximization) algorithm
2. how to conduct MCMC

EM

Denoting

$$Q(q|q_0, x) = \mathbb{E}_{q_0}\left[\log L_c(q|x, Z)\right],$$

the EM algorithm indeed proceeds "iteratively" by maximizing $Q(q|q_0, x)$ at each iteration, and, if $q^{(\hat{1})}$ is the value of q maximizing $Q(q|q_0, x)$, by replacing q_0 by the updated value $q^{(\hat{1})}$. In this manner, a sequence of estimators $\{q^{(\hat{j})}\}_j$ is obtained, where

$$Q(q^{(\hat{j})}|q^{(j\hat{-}1)})$$

Pick a starting value $q^{(\hat{0})}$ and set $m = 0$. Repeat

1. Compute (the E-step)
$$Q(q|q^{(\hat{m}),x}) = \mathbb{E}_{q^{(\hat{m})}}\left[\log L_c(q|x, Z)\right],$$
where the expectation is with respect to $k(z|q^{(\hat{m}),x})$.
2. Maximize $Q(q|q^{(\hat{m}),x})$ in q and take (the M-step)
$$q^{(m\hat{+}1)} = \arg\max_q Q(q|q^{(\hat{m}),x})$$
and set $m = m + 1$.

until a fixed point is reached; i.e., $q^{(m\hat{+}1)} = q^{(\hat{m})}$.

....

Gibbs Sampling (Markov Chain Monte Carlo)

Consider $f\left(y_{ij} \mid u_i, \beta\right) = e^{\left(\frac{y_{ij}\theta_{ij} - \psi\left(\theta_{ij}\right)}{a(\phi)} - c\left(y_{ij}; \phi\right)\right)}$

with $g\left(u_i \mid G\right) = (2\pi)^{\frac{-q}{2}}|G|^{\frac{-1}{2}}e^{\frac{-u_i^T G^{-1} u_i}{2}}$, and $h\left(\mu_{ij}\right) = x_{ij}^T\beta + z_{ij}^T u_i$.

The likelihood function of (β, G) is

$$L(\beta, G \mid \underset{\sim}{y}) \propto \prod_{i=1}^{n}\prod_{j=1}^{m_i} f(y_{ij} \mid u_i, \beta)|G|^{-1/2}\exp\{\frac{u_i^T G^{-1} u_i}{2}\}du_i.$$

In a Bayesian approach to analyzing the random effects GLM, the parameters (β, G) are random variables and are treated symmetrically with the longitudinal measurements and unobserved latent variables. Thus, the random effects GLM is an example of a hierarchical Bayes model.

Assumptions: $[\beta \mid G, U, \underset{\sim}{y}] = [\beta \mid U, \underset{\sim}{y}]$, $[G \mid \beta, U, \underset{\sim}{y}] = [G \mid U]$ and $[U \mid \beta, G, \underset{\sim}{y}]$.

1. Assume that β has a flat prior function. Then,

$\left[\beta \mid U^{(k)}, \underset{\sim}{y}\right] \propto \prod_{i=1}^{n} \prod_{j=1}^{m_i} f\left(y_{ij} \mid U_i^{(k)}, \beta\right) \approx N\left(\beta^{(k)}, V_\beta^{(k)}\right)$, as $n \rightarrow \infty$, where $\beta^{(k)}$ is the maximum likelihood estimator and $V_\beta^{(k)}$ is the inverse of the Fisher information.

Adjustment for smaller samples - "Rejection sampling" (Ripley, 1987)

Let $f\left(\beta \mid U^{(k)}, \underset{\sim}{y}\right)$ and $\phi\left(\beta \mid \beta^{(k)}, V_\beta^{(k)}\right)$ denote separately the true density and Gaussian density. Choose a constant $c \geq 1$ such that $c\phi\left(\beta \mid \beta^{(k)}, V_\beta^{(k)}\right) \geq f\left(\beta \mid U^{(k)}, \underset{\sim}{y}\right)$.

Step1: Generate $\beta^* \sim \phi\left(\beta \mid \beta^{(k)}, V_\beta^{(k)}\right)$ and $u \sim U(0,1)$.

Step2: If $\frac{f\left(\beta^* \mid b^{(k)}, \underset{\sim}{y}\right)}{c\phi\left(\beta^* \mid \beta^{(k)}, V_\beta^{(k)}\right)} < u, \beta^{(k+1)} = \beta^*$. Otherwise, the process returns to Step1.

2. $\left[G \mid U^{(k)}\right]$

Assume that $\pi(G) \propto |G|^{-1}$: non-informative prior (see Box and Tiao, 1973).

Then, $\left[G \mid U^{(k)}\right] \sim$ Inverted Wishart $\left(S^{(k)}, n-q+1\right)$, where $S^{(k)} = \sum_{i=1}^{n} U_i^{(k)} U_i^{(k)^T}$

Remark

If A $\sim$ Wishart $\left(\Sigma_{p \times p}, n\right)$, the p.d.f of A is $f_{\mathrm{A}}(\mathrm{A}) \propto |\mathrm{A}|^{\frac{-1}{2}(n-p-1)} e^{\frac{-1}{2} tr \Sigma^{-1} \mathrm{A}}$.

It implies that $\mathrm{B} = \mathrm{A}^{-1} \sim$ Inverted Wishart $\left(\Sigma^{-1}, n\right)$ with p.d.f. $f_{\mathrm{B}}(\mathrm{B}) \propto |\mathrm{B}|^{\frac{-1}{2}(n+p+1)} e^{\frac{-1}{2} tr \Sigma^{-1} \mathrm{B}^{-1}}$.

Thus,

$$\pi\left(G \mid U^{(k)}\right) \propto |G|^{\frac{-1}{2}(n+2)} e^{\frac{-1}{2} \operatorname{tr}\left(S^{(k)} G^{-1}\right)},$$

i.e. $\left[G \mid U^{(k)}\right] \sim$ Inverted Wishart $\left(S^{(k)}, n-q+1\right)$.

3. $\left[U \mid \beta^{(k)}, G^{(k)}, y\right]$

Using

$$f\left(U_i \mid \hat{\beta}^{(k)}, G^{(k)}, \underset{\sim}{y_i}\right) \propto f\left(\underset{\sim}{y_i} \mid U_i, \hat{\beta}^{(k)}\right) g\left(U_i \mid G^{(k)}\right) \triangleq f_n\left(U_i \mid \hat{\beta}^{(k)}, G^{(k)}, \underset{\sim}{y_i}\right),$$

we can find the mode and curvature of $f_n\left(U_i \mid \hat{\beta}^{(k)}, G^{(k)}, \underset{\sim}{y_i}\right)$, which matches a Gaussian density.

Using the surrogate response

$$Z_i^* = X_i\beta + D_i U_i + \operatorname{Diag}\left(h'\left(\mu_i\right)\right)\left(y_i - \mu_i\right),$$

the maximum value of $f_n\left(U_i \mid \hat{\beta}^{(k)}, G^{(k)}, \underset{\sim}{y_i}\right)$ occurs at

$$U_i = \left(D_i^T Q_i^{-1} D_i + G^{(k)^{-1}}\right)^{-1} D_i^T Q_i^{-1}\left(Z_i^* - X_i \beta^{(k)}\right) = G^{(k)} D_i \left(D_i G^{(k)} D_i^T + Q_i\right)^{-1}\left(Z_i^* - X_i \hat{\beta}^{(k)}\right)$$

and its curvature is $V_i = \left(D_i^T Q_i^{-1} D_i + G^{(k)^{-1}}\right)^{-1}$. Similar to the method in (3), $U_i^{(k)}$ can be obtained.

Generalized Estimating Equations (GEE), which is a multivariate analogue of quasi-likelihood.

$S_\beta(\beta,\alpha)=\sum_{i=1}^n\left(\frac{\partial\mu_i}{\partial\beta}\right)^T[\mathrm{Var}\,(Y_i)]^{-1}(Y_i-\mu_i)$, where $\mu_i=h\left(x_{ij}^T\beta\right),\mathrm{Var}\,(Y_i)=\mathrm{Var}\,(Y_i;\beta,\alpha)$

$S_\alpha(\beta,\alpha)=\sum_{i=1}^n\left(\frac{\partial\eta_i}{\partial\alpha}\right)^T H_i^{-1}(\omega_i-\eta_i)$, where $\omega_i=\left(R_{i1}R_{i2},\cdots,R_{i1}R_{im_i},\cdots,R_{i1}^2,\cdots,R_{im_i}^2\right)$

$\eta_i=\mathbb{E}[\omega_i\,|\,(\beta,\alpha)]$, and $H_i=\mathrm{Var}(\omega_i)$, with $R_{ij}=\frac{Y_{ij}-\mu_{ij}}{\sqrt{\mathrm{Var}(Y_{ij})}}$

The estimator, say $(\hat\beta,\hat\alpha)$ of (β,α) is defined to be the solution of the above equations,

i.e. $S_\beta(\hat\beta,\hat\alpha)=0$ and $S_\alpha(\hat\beta,\hat\alpha)=0$.

Problem 5.

Theorem 3.1.

Under the regularity conditions, $n^{\frac12}\left[\begin{pmatrix}\hat\beta\\ \hat\alpha\end{pmatrix}-\begin{pmatrix}\beta\\ \alpha\end{pmatrix}\right]\xrightarrow{d}N(0,\Sigma)$, where Σ can be estimated by

$$\left(\frac1n\sum_{i=1}^n C_i^TB_i^{-1}D_i\right)^{-1}\left(\frac1n\sum_{i=1}^n C_i^TB_i^{-1}V_{0i}B_i^{-1}C_i\right)\left(\frac1n\sum_{i=1}^n D_i^TB_i^{-1}C_i\right)^{-1},$$

where $C_i=\begin{pmatrix}\frac{\partial\mu_i}{\partial\beta}&0\\0&\frac{\partial\eta_i}{\partial\alpha}\end{pmatrix}$, $B_i=\begin{pmatrix}\mathrm{Var}\,(Y_i)&0\\0&H_i\end{pmatrix}$, $D_i=\begin{pmatrix}\frac{\partial\mu_i}{\partial\beta}&\frac{\partial\mu_i}{\partial\alpha}\\ \frac{\partial\eta_i}{\partial\beta}&\frac{\partial\eta_i}{\partial\alpha}\end{pmatrix}$,

and $V_{0i}=\begin{pmatrix}y_i-\mu_i\\ \omega_i-\eta_i\end{pmatrix}^{\otimes2}$.

—

Hint

$$\begin{aligned}\begin{pmatrix}0\\0\end{pmatrix}&=\begin{pmatrix}S_\beta(\hat\beta,\hat\alpha)\\S_\alpha(\hat\beta,\hat\alpha)\end{pmatrix}\\&=\begin{pmatrix}S_\beta(\beta,\alpha)\\S_\alpha(\beta,\alpha)\end{pmatrix}+\left.\begin{pmatrix}\frac{\partial S_\beta(\beta,\alpha)}{\partial\beta}&\frac{\partial S_\beta(\beta,\alpha)}{\partial\alpha}\\ \frac{\partial S_\alpha(\beta,\alpha)}{\partial\beta}&\frac{\partial S_\alpha(\beta,\alpha)}{\partial\alpha}\end{pmatrix}\right|_{\begin{pmatrix}\hat\beta^*\\ \hat\alpha^*\end{pmatrix}}\left[\begin{pmatrix}\hat\beta\\ \hat\alpha\end{pmatrix}-\begin{pmatrix}\beta\\ \alpha\end{pmatrix}\right],\end{aligned}$$

where $\begin{pmatrix}\hat\beta^*\\ \hat\alpha^*\end{pmatrix}$ lies on the line segment between $\begin{pmatrix}\beta\\ \alpha\end{pmatrix}$ and $\begin{pmatrix}\hat\beta\\ \hat\alpha\end{pmatrix}$.

From the first order Taylor Expansion, we can obtain that

$$\begin{aligned}n^{1/2}\left[\begin{pmatrix}\hat\beta\\ \hat\alpha\end{pmatrix}-\begin{pmatrix}\beta\\ \alpha\end{pmatrix}\right]&=-n\left.\begin{pmatrix}\frac{\partial S_\beta(\beta,\alpha)}{\partial\beta}&\frac{\partial S_\beta(\beta,\alpha)}{\partial\alpha}\\ \frac{\partial S_\alpha(\beta,\alpha)}{\partial\beta}&\frac{\partial S_\alpha(\beta,\alpha)}{\partial\alpha}\end{pmatrix}^{-1}\right|_{\begin{pmatrix}\hat\beta^*\\ \hat\alpha^*\end{pmatrix}}n^{-1/2}\begin{pmatrix}S_\beta(\beta,\alpha)\\S_\alpha(\beta,\alpha)\end{pmatrix}\\&=-nV^{*-1}\,n^{-1/2}\begin{pmatrix}S_\beta(\beta,\alpha)\\S_\alpha(\beta,\alpha)\end{pmatrix}.\end{aligned}$$

We have

$$
\begin{aligned}
V^* \xrightarrow{p} V &= \mathbb{E}\left[\begin{pmatrix} \dfrac{\partial S_\beta(\beta,\alpha)}{\partial\beta} & \dfrac{\partial S_\beta(\beta,\alpha)}{\partial\alpha} \\ \dfrac{\partial S_\alpha(\beta,\alpha)}{\partial\beta} & \dfrac{\partial S_\alpha(\beta,\alpha)}{\partial\alpha} \end{pmatrix}\right] \\
&= \begin{pmatrix} \left(\dfrac{\partial\mu_i}{\partial\beta}\right)^T [\mathrm{Var}(Y_i)]^{-1}\left(\dfrac{\partial\mu_i}{\partial\beta}\right) & \left(\dfrac{\partial\mu_i}{\partial\beta}\right)^T [\mathrm{Var}(Y_i)]^{-1}\left(\dfrac{\partial\mu_i}{\partial\alpha}\right) \\ \left(\dfrac{\partial\eta_i}{\partial\alpha}\right)^T H_i^{-1}\left(\dfrac{\partial\eta_i}{\partial\beta}\right) & \left(\dfrac{\partial\eta_i}{\partial\alpha}\right)^T H_i^{-1}\left(\dfrac{\partial\eta_i}{\partial\alpha}\right) \end{pmatrix} \\
&= \mathbb{E}[C_i^T B_i^{-1} D_i],
\end{aligned}
$$

which is estimated by $\left(\frac{1}{n}\sum_{i=1}^n C_i^T B_i^{-1} D_i\right)$.

—

$$
\begin{aligned}
S_\beta(\beta,\alpha) &= \sum_{i=1}^n \left(\frac{\partial\mu_i}{\partial\beta}\right)^T [\mathrm{Var}\,(Y_i)]^{-1}(Y_i-\mu_i) \\
&= \sum_{i=1}^n U_i.
\end{aligned}
$$

$$
\begin{aligned}
S_\alpha(\beta,\alpha) &= \sum_{i=1}^n \left(\frac{\partial\eta_i}{\partial\alpha}\right)^T H_i^{-1}(\omega_i-\eta_i) \\
&= \sum_{i=1}^n Z_i.
\end{aligned}
$$

By CLT,

$$
n^{-1/2}S_\beta(\beta,\alpha) = \sqrt{n}\frac{1}{n}\sum_{i=1}^n U_i = \sqrt{n}(\bar{U}_i - \mathbb{E}[U_i]) \xrightarrow{d} N(0,\sigma_U^2),
$$

where

$$
\begin{aligned}
\sigma_U^2 &= \mathrm{Var}[U_i] \\
&= \mathrm{Var}[\left(\frac{\partial\mu_i}{\partial\beta}\right)^T [\mathrm{Var}\,(Y_i)]^{-1}(Y_i-\mu_i)] \\
&= \left(\frac{\partial\mu_i}{\partial\beta}\right)^T [\mathrm{Var}\,(Y_i)]^{-2}\left(\frac{\partial\mu_i}{\partial\beta}\right)\mathrm{Var}(Y_i) \\
&= \left(\frac{\partial\mu_i}{\partial\beta}\right)^T [\mathrm{Var}\,(Y_i)]^{-1}\left(\frac{\partial\mu_i}{\partial\beta}\right)
\end{aligned}
$$

Similarly,

$$
n^{-1/2}S_\alpha(\beta,\alpha) = \sqrt{n}\frac{1}{n}\sum_{i=1}^n Z_i = \sqrt{n}(\bar{Z}_i - \mathbb{E}[Z_i]) \xrightarrow{d} N(0,\sigma_Z^2),
$$

where

$$
\begin{aligned}
\sigma_Z^2 &= \mathrm{Var}[Z_i] \\
&= \mathrm{Var}[\left(\frac{\partial\eta_i}{\partial\alpha}\right)^T H_i^{-1}(\omega_i-\eta_i)] \\
&= \left(\frac{\partial\eta_i}{\partial\alpha}\right)^T H_i^{-2}\,\mathrm{Var}[\omega_i]\left(\frac{\partial\eta_i}{\partial\alpha}\right) \\
&= \left(\frac{\partial\eta_i}{\partial\alpha}\right)^T H_i^{-1}\left(\frac{\partial\eta_i}{\partial\alpha}\right).
\end{aligned}
$$

Thus,

$$n^{-1/2}\begin{pmatrix} S_\beta(\beta,\alpha) \\ S_\alpha(\beta,\alpha) \end{pmatrix} \xrightarrow{d} N(0,\Sigma_S),$$

where

$$\Sigma_S = \begin{pmatrix} \sigma_U^2 & \sigma_{UZ}^2 \\ \sigma_{UZ}^2 & \sigma_Z^2 \end{pmatrix},$$

with

$$\sigma_{UZ}^2 = \mathbf{Cov}(U_i, Z_i) = \mathbb{E}[U_i Z_i]$$

$$= \mathbb{E}\left[\left(\frac{\partial \mu_i}{\partial \beta}\right)^T [\mathrm{Var}\,(Y_i)]^{-1}\,(Y_i-\mu_i)\cdot\left(\frac{\partial \eta_i}{\partial \alpha}\right)^T H_i^{-1}\,(\omega_i-\eta_i)\right].$$

Since

$$\begin{aligned}&\mathbb{E}(C_i^T B_i^{-1} V_{0i} B_i^{-1} C_i)\\ &= \mathbb{E}\left[C_i^T B_i^{-1}\begin{pmatrix} (y_i-\mu_i)^2 & (y_i-\mu_i)(\omega_i-\eta_i) \\ (y_i-\mu_i)(\omega_i-\eta_i) & (\omega_i-\eta_i)^2 \end{pmatrix} B_i^{-1} C_i\right],\end{aligned}$$

where

$$C_i^T B_i^{-1} = \begin{pmatrix} \left(\frac{\partial \mu_i}{\partial \beta}\right)^T [\mathrm{Var}(Y_i)]^{-1} & 0 \\ 0 & \left(\frac{\partial \eta_i}{\partial \alpha}\right)^T H_i^{-1} \end{pmatrix},$$

we can estimate Σ_S with $\left(\frac{1}{n}\sum_{i=1}^n C_i^T B_i^{-1} V_{0i} B_i^{-1} C_i\right)$.

By Slutsky Theorem,

$$n^{1/2}\left[\begin{pmatrix} \hat\beta \\ \hat\alpha \end{pmatrix} - \begin{pmatrix} \beta \\ \alpha \end{pmatrix}\right] \xrightarrow{d} N(0,\Sigma),$$

where $\Sigma = V^{-T}\Sigma_S V^{-1}$.

Therefore, Σ can be estimated by

$$\left(\frac{1}{n}\sum_{i=1}^n C_i^T B_i^{-1} D_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n C_i^T B_i^{-1} V_{0i} B_i^{-1} C_i\right)\left(\frac{1}{n}\sum_{i=1}^n D_i^T B_i^{-1} C_i\right)^{-1},$$