广义线性模型

S.T.L

2021 年 10 月 31 日

目录

第一章　广义线性模型

概述

现有一随机变量（也可以是向量）Y 与 p 维变量 $X=(X_1,\cdots,X_p)$，两者之间存在这某种统计关系，其中 Y 是来自由某一列参数 $\gamma=(\theta,\phi)$ 确定的分布族（此处我们讨论指数分布族），X 对 Y 的影响体现在 $\gamma=(\theta,\phi)$ 中的 θ 上，具体的有映射 $K:\mathcal{X}\to\Theta,\theta=K(X)$，而 ϕ 的取值与 X 无关，为讨厌参数. 对给定的 X，我们可以确定 Θ，同时如果对 ϕ 存在某些假设（如线性回归中的 σ^2 为常数），就可以确定 Y 的分布.

但是映射 K 的具体形式往往是不知道的，其未知性包含两点：1.θ 与 X 之间是一种什么关系（线性、二次、logit⋯⋯）；2.θ 与 X 之间满足这种关系后某些系数的取值.

由于 $K(x_1,\cdots,x_p)$ 是一个 p 元函数,对多元函数的处理是非常麻烦的,我们希望对 $x=(x_1,\cdots,x_p)$ 作某种变换，得到更好处理的单一变元 η 再与 θ 建立联系. 很直观的想法是对 X 作线性假设，令 $\eta=X\beta$.（这种做法实用性很广泛，在线性回归中我们发现很多关系都可以转化成线性关系（如多项式、乘积式作对数变换等，参见非线性回归））这里我们研究一类基于线性关系建立起的模型，即 $K(x)=h(\eta)=h(X\beta)$.

于是我们只需要选取某个特定的映射对 θ 与 η 建立联系（注意，这个联系是人为选取的，不同的联系会产生不同的模型，甚至某些选取完全是错误的，比如对指数模型作线性回归，无论作什么处理都不会得到很好的拟合结果），就可以建立 Y 与 X 之间的联系，而整个模型中需要估计的参数只有 β.

理解广义线性模型的核心在于理解各个变量的含义及其相互的联系，其他内容基本都是多元情形下的数理统计的概念. 因此在这里先列出各个记号及映射关系，其中可能有部分变量和函数在概述中没有提及，在后续会展开解释.

$$
\begin{tikzcd}
\underset{q\times p}{\boldsymbol{X}} \arrow[r, "\eta=X\beta"] & \underset{q\times 1}{\eta} \arrow[r, shift left, "h(\eta)"] \arrow[rr, bend left, "\xi=\mu^{-1}(h(\eta))"] & \underset{q\times 1}{\mu} \arrow[l, shift left, "g(\mu)"] \arrow[r, shift left, "\mu^{-1}"] & \underset{q\times 1}{\theta} \arrow[l, shift left, "b'(\theta)"] \arrow[r] & l(Y,\theta)
\end{tikzcd}
$$

1.1　引入

在一般线性回归中，响应变量 Y 与设计矩阵 $\boldsymbol{X}$ 之间存在着统计关系

$$Y=\boldsymbol{X}\beta+\epsilon$$

其中 $\mathbb{E}[\epsilon_i]=0,\mathrm{Cov}[\epsilon]=\sigma^2I$，特别的可取 $\epsilon\sim N_p(\mathbf{0},\sigma^2I)$.

对上述模型，可以有另一种理解：

$$Y_i\sim N(\mu_i,\sigma^2)\quad \mathbb{E}[Y_i]=\mu_i\quad \mu_i=\eta_i=X_i\beta \tag{1.1}$$

这里我们引入了变量 $\eta_i = X_i\beta$，来表示自变量 X_i 在 β 下的线性组合.

从而我们希望当 Y 不再服从正态分布，而是其他某种已知的分布时，建立 Y 与 $\eta = X\beta$ 之间的一种联系？考虑到 Y 是一个随机变量，因此需要寻找一个与 Y 相关的定量来描述 Y 的特性，因此模仿线性回归，我们希望建立 $\mathbb{E}[Y]$ 与 η 之间的关系。因此我们将1.1改写称如下形式

$$Y_i \sim f(\theta, \phi) \quad \mathbb{E}[Y_i] = \mu_i \quad g(\mu_i) = \eta_i = X_i\beta \tag{1.2}$$

上述三个元素就组成了广义线性模型（后续会对这三个元组给出具体说明）.

1.2　指数分布族

为便于建立各个变量与参数之间的关系（求导、求期望等），广义线性模型要求随机变量 Y 的分布来自指数分布族. 在介绍广义线性模型前，我们先来考察一些指数分布族的一些性质

定义 1.1. *称随机向量 Y 来自指数分布族，若其概率密度函数可以写成*

$$f_Y(y, \theta) = a(\theta)b(y)\exp\{y'Q(\theta)\}$$

一个更一般的参数化密度函数允许包含一些讨厌参数或扩散参数 ϕ. 此时我们将其密度函数写称如下形式

$$f(\boldsymbol{y}; \boldsymbol{\theta}, \phi) = \exp\left\{\frac{\boldsymbol{y}'\boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\phi)} + c(\phi, \boldsymbol{y})\right\} \tag{1.3}$$

其中 $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$ 称为自然参数；ϕ 未知，称为扩散参数或讨厌参数. 对给定的 ϕ 称 Θ 为自然参数空间，即所有满足下式的集合

$$0 < \int \exp\{\boldsymbol{y}'\boldsymbol{\theta}/a(\phi) + c(\phi, \boldsymbol{y})\}\mathrm{d}\boldsymbol{y} < \infty$$

此时 Θ 是凸的.

定理 1.1. *若 Y 来自上述指数分布族，则*

$$\mathbb{E}_{\boldsymbol{\theta}}[Y] = \frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \qquad \mathrm{Cov}_{\boldsymbol{\theta}}[Y] = a(\phi)\frac{\partial^2 b(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \triangleq \Sigma(\boldsymbol{\theta}) \tag{1.4}$$

1.3　广义线性模型的结构

一个 GLM 由三个部分组成，在介绍这三个部分时，我们先对我们拥有的观测结数据做一些解释：$X = \{X_1, \cdots, X_N\}$ 是我们收集到的 N 个 $q \times p$ 维的协变量（*有一些书中将 X 定义为不含常数 1 列的变量，而用 Z 表示 $(1, X)$，而此处，我们的 X 与 Z 表示同一个量*），与之对应的有 N 个 $q \times 1$ 维的观测结果 $Y = \{Y_1, \cdots, Y_N\}$.

这其实是我们常见的数据的推广，在学习回归的过程中，我们接触到的往往是 N 个 1 维的响应变量 Y，以及 N 个 p 维的协变量 X，即上述数据中取 $q = 1$ 的情形. 实际上，我们接触到的大部分模型都是 $p = 1$ 的，但为了让我们的模型更具泛用性，我们直接研究 $p \geq 1$ 时的情形.

下面我们来看 GLM 的三个部分：

1. 随机部分：*建立起了 $l(\theta, Y_i)$ 与 θ 以及 μ 之间的联系.*

   随机部分的 Y 由 N 个观测值 $Y = \{Y_1, \cdots, Y_N\}$ 组成. 每个 Y_i 相互独立且来自由相同 $q \times 1$ 维未知参数 $\boldsymbol{\theta}$ 决定的指数分布族. 考虑单个随机向量 Y_i 的指数分布族，并结合1.1则有

$$\mu_i \triangleq \mathbb{E}[Y_i] = \frac{\partial b_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left(\frac{\partial b_i}{\partial \theta_1}, \cdots, \frac{\partial b_i}{\partial \theta_N}\right)'$$

2. 系统部分: 这一部分建立起了 X 与 η 之间的联系.

   系统部分是一组解释变量构成的线性模型

$$\eta_i = X_i\beta$$

   其中 η_i 叫做线性预报；X_i 是 $q \times p$ 维矩阵 $(i = 1, \cdots, N)$，是对解释变量的一组观测值；β 是一组 p 维未知常向量.

3. 连接函数：这一部分建立起了 μ 与 η 之间的联系

   记 $\mathbb{E}[Y_i] = \mu_i$ 则连接函数可写作

$$g(\mu_i) = \eta_i = X_i\beta \quad i = 1, \cdots, N$$

   其中 $g : \mathcal{H} \to \mathbb{R}^q, \mathcal{H} \subset \mathbb{R}^q$ 是一个可逆且可微的函数（在一元情形下往往取作单调函数）.

   又由 $g(\mu)$ 可逆可微知其存在反函数 $h : \mathbb{R}^q \to \mathcal{H}, h(\eta) = g^{-1}(\theta)$，称为响应函数.

   当 $g(\mu)$ 与 $\mu = b'(\theta)$ 都给定时，我们还能直接写出 $\eta \to \theta$ 的映射 $\xi(\eta)$.

   $\boldsymbol{g(\mu)}$ 的几个特殊情形：

   (1) 自连接（identity link）:$g(\mu) = \mu$

   (2) 自然连接（canonical link）:$g(\mu(\theta)) = \theta$. 由1.1知，自然连接满足 $g(\mu_i) = b'^{-1}(\mu)$.

   下面给出一些自然函数习惯上的选取方式：

   (1) 如果响应变量的观察值是连续型数据，同时又比较对称，这时可以假定随机变量来自于正态总体.

   (2) 如果选择正态总体而连结函数是自然连结函数，这就成了普通的线性回归（读者可自行验证以加强对自然连接函数的理解）.

   (3) 如果响应变量的观察值是非负的连续型数据，则可应用 Γ 分布来构造广义线性回归模型.

   (4) 如果数据是非对称的生命数据，就可以应用逆高斯分布来构造广义线性回归模型.

1.4　得分函数与信息矩阵

首先我们需要对模型作一些必要的假设：

- $\boldsymbol{\beta}$ 的可容许参数空间 B 是开的
- 对所有参数 $\beta \in B, \boldsymbol{h}(Z_i\beta) \in y$ 其中 $i = 1, 2, \cdots . Z_i$ 为 $q \times p$ 的设计矩阵（这里我们不再用 X_i，而是使用 Z_i!!!），y 为 Y 的均值取值空间.
- $\boldsymbol{h}, \boldsymbol{g}$ 与 $\boldsymbol{\xi}$ 是二阶可微的，且 $\det(\partial \boldsymbol{g}/\partial \boldsymbol{\eta}) \neq 0$.
- 对足够大的 n，$\sum\limits_{i=1}^{n} Z_i Z_i'$ 满秩.

这一块由于主要就是计算，所以基本就是抄周勇老师的广义估计方差估计方法 P244-246 的内容，但由于这部分内容比较简略，会在文末会对 $S_i(\beta)$ 与 $F_i(\beta)$ 的推导作出一些解释.

考察 Y 的似然函数（likelihood function）

$$L(\theta,\phi;Y)=\prod_{i=1}^{N}exp\left\{\frac{\mathbf{Y}_i'\boldsymbol{\theta}_i-b(\boldsymbol{\theta}_i)}{a(\phi)}+c(\mathbf{Y}_i,\phi)\right\} \tag{1.5}$$

其对数似然函数可以写成

$$l(\theta,\phi;Y)=\sum_{i=1}^{N}\left\{\frac{\mathbf{Y}_i'\boldsymbol{\theta}_i-b(\boldsymbol{\theta}_i)}{a(\phi)}+c(\mathbf{Y}_i,\phi)\right\} \tag{1.6}$$

对数似然函数对 θ 的一阶导函数组成的向量函数称为得分函数（score function）.

$$\boldsymbol{S}(\boldsymbol{\theta})=\frac{\partial l}{\partial\boldsymbol{\theta}}=\sum_{i=1}^{N}\boldsymbol{S}_i(\boldsymbol{\theta}) \tag{1.7}$$

将 $a(\phi)$ 视作讨厌参数，于是得到相应的估计方程：

$$\frac{\partial l}{\partial\boldsymbol{\theta}}=\sum_{i=1}^{N}\frac{Y_i-\boldsymbol{\mu}_i}{a_i(\phi)}=0 \tag{1.8}$$

而在实际的广义线性模型中，我们感兴趣的是未知参数 $\boldsymbol{\beta}$，每个观测值 Y_i 对 $\boldsymbol{\beta}$ 的对数似然作出的贡献可记为

$$l_i(\boldsymbol{\beta})=[Y_i'\boldsymbol{\theta}_i-b(\boldsymbol{\theta}_i)]/a_i(\phi) \tag{1.9}$$

每个个体关于 β 的的分函数为 $\boldsymbol{S}_i(\boldsymbol{\beta})=\partial l_i/\partial\boldsymbol{\beta}$ 由此可得

$$\begin{aligned}\boldsymbol{S}_i(\boldsymbol{\beta})&=\frac{\partial\boldsymbol{h}(Z_i\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Sigma_i^{-1}(\boldsymbol{\beta})(Y_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}))\\&=Z_i'D_i(\boldsymbol{\beta})\Sigma_i^{-1}(\boldsymbol{\beta})(Y_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}))\end{aligned} \tag{1.10}$$

其中 $D_i(\boldsymbol{\beta})=\dfrac{\partial\boldsymbol{h}(\boldsymbol{\eta})}{\partial\eta}$ 是 $\boldsymbol{h}(\boldsymbol{\eta})$ 的雅可比矩阵在 $\eta_i=Z_i\boldsymbol{\beta}$ 的值，$\Sigma_i(\boldsymbol{\beta})=v(\boldsymbol{h}(Z_i\boldsymbol{\beta}))a_i(\phi)$. 上式还可以等价地写为

$$\boldsymbol{S}_i(\boldsymbol{\beta})=Z_i'W_i(\boldsymbol{\beta})\frac{\partial\boldsymbol{g}(\boldsymbol{\mu}_i)}{\partial\boldsymbol{\mu}'}(Y_i-\boldsymbol{\mu}_i(\boldsymbol{\beta})) \tag{1.11}$$

其中 W_i 是一个权矩阵

$$W_i(\boldsymbol{\beta})=\left[\frac{\partial\boldsymbol{g}(\boldsymbol{\mu}_i)}{\partial\boldsymbol{\mu}'}\Sigma_i(\boldsymbol{\beta})\frac{\partial\boldsymbol{g}(\boldsymbol{\mu}_i)}{\partial\boldsymbol{\mu}}\right]^{-1}=D_i\Sigma_i^{-1}D_i' \tag{1.12}$$

由此只需求解方程

$$\boldsymbol{S}(\boldsymbol{\beta})=\frac{\partial l}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{N}\frac{\partial l_i}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{N}\boldsymbol{S}_i(\boldsymbol{\beta})=0 \tag{1.13}$$

即可得到 $\boldsymbol{\beta}$ 的极大似然估计 $\hat{\boldsymbol{\beta}}_n$.

得分函数及 Fisher 信息阵的性质

下面考察得分函数及其 Fisher 信息阵的一些性质首先对 $\boldsymbol{S}_i$，其期望为 0:

$$\mathbb{E}_{\boldsymbol{\beta}}\boldsymbol{S}_i(\boldsymbol{\beta})=Z_i'D_i(\boldsymbol{\beta})\Sigma_i^{-1}(\boldsymbol{\beta})(\mathbb{E}[Y_i]-\boldsymbol{\mu}_i(\boldsymbol{\beta}))=0 \tag{1.14}$$

该统计结构的 Fisher 信息阵或期望信息阵为

$$F_i(\boldsymbol{\beta})=\mathrm{Cov}_{\beta}(\boldsymbol{S}_i(\boldsymbol{\beta}))=\mathbb{E}_{\beta}(\boldsymbol{S}_i(\boldsymbol{\beta})\boldsymbol{S}_i'(\boldsymbol{\beta}))=Z_i'W_iZ_i \tag{1.15}$$

观察信息阵为

$$F_{i,obs}(\boldsymbol{\beta}) = -\frac{\partial^2 l_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = F_i(\boldsymbol{\beta}) - \frac{1}{a(\phi)} R_i(\boldsymbol{\beta}) \tag{1.16}$$

其中

$$R_i(\boldsymbol{\beta}) = \sum_{r=1}^{q} Z_i' U_{ir} Z_i (y_{ir} - \mu_{ir}(\boldsymbol{\beta})) \tag{1.17}$$

其中 $U_{ir} = \dfrac{\partial^2 \xi_i(Z_i \boldsymbol{\beta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}$ 为 $q \times q$ 矩阵.（注意，1.16中 F 的表达式与很多介绍 GLM 的书都不同，多了一项 $a(\phi)^{-1}$，这一点将会在附注中说明.）

易证

$$\mathbb{E}_{\beta}[R_i(\boldsymbol{\beta})] = 0 \quad \mathbb{E}_{\beta}[F_{i,obs}(\boldsymbol{\beta})] = 0$$

由此可以证明在前述假设条件成立的情况下，通过极大似然估计得到的 $\hat{\boldsymbol{\beta}}_{ML}$ 是 $\boldsymbol{\beta}$ 的相合估计，且具有渐近正态分布.

定理 1.2. 假设前述条件成立，则

$$\hat{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}$$

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\mathbf{D}} N(0, \Sigma^{-1}(\boldsymbol{\beta}))$$

其中 $\Sigma(\boldsymbol{\beta}) = \lim\limits_{n \to \infty} n^{-1} \sum_{i=1}^{n} Z_i' W_i Z_i$.

这一部分本质上就是多元里的 MLE 的性质，因此不展开具体证明.

可见 MLE 是对 GLM 中参数 $\boldsymbol{\beta}$ 的一个较好的估计.

#：先介绍到这里吧，后续随缘写一些假设检验和其他方面的拓展，可能还会对二项分布展开介绍一下（毕竟全篇没有例子）. 属实是不想再敲笔记了 23333

另外，别跑，后面还有附注的证明. 敲了半天的公式 ==

1.5 附注（一些推导）

证明. 定理 *1.1*

$$\frac{\partial \ln f}{\partial \boldsymbol{\theta}}=\frac{1}{f}\frac{\partial f}{\partial \boldsymbol{\theta}}=\frac{1}{a(\phi)}\left\{\boldsymbol{y}-\frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}$$

$$\frac{\partial^2 \ln f}{\partial \boldsymbol{\theta}^2}=\frac{1}{f^2}\left[\frac{\partial^2 f}{\partial \theta^2}-\left(\frac{\partial f}{\partial \boldsymbol{\theta}}\right)^2\right]=-\frac{1}{a(\phi)}\frac{\partial^2 b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}$$

从而由正则化假设有

$$\mathbb{E}[Y]=\int \boldsymbol{y} f \mathrm{d}\boldsymbol{y}=\int a(\phi)\frac{\partial f}{\partial \boldsymbol{\theta}}\mathrm{d}\boldsymbol{y}+\frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}=\frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

$$\begin{aligned}\mathrm{Cov}[Y]&=\mathbb{E}(Y-\frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^2=\int\left(\boldsymbol{y}-\frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) f\mathrm{d}\boldsymbol{y}\\&=\int \frac{a^2}{f^2}\left(\frac{\partial f}{\partial \boldsymbol{\theta}}\right)^2 f\mathrm{d}y=\int a^2\left(\frac{\partial^2 f}{\partial \boldsymbol{\theta}^2}+\frac{f}{a(\phi)}\frac{\partial^2 b}{\partial \boldsymbol{\theta}^2}\right)\mathrm{d}\boldsymbol{y}=a(\phi)=\frac{\partial^2 b}{\partial \boldsymbol{\theta}^2}\end{aligned}$$

证明. 得分函数与 *Fisher* 信息阵相关的等式推导. 主要为*1.10*,*1.16*.

直接给出详细的证明往往是不讨喜的，因此这部分建议自己推导.

首先我们列出几个的一阶偏导结果，这些结果会在推导中反复用到 *(*注，为方便书写，将所有下标略去，且不再对向量可以加粗（因为实在太费时间了！）*)*

$$\frac{\partial l}{\partial \theta'}=\left(y'-\frac{\partial b}{\partial \theta'}\right)\Big/ a(\phi) \tag{1.18}$$

$$\frac{\partial \theta}{\partial \mu'}=\left(\frac{\partial \mu}{\partial \theta'}\right)^{-1}=\left(\frac{\partial^2 b}{\partial \theta\partial\theta'}\right)^{-1}=a(\phi)\Sigma^{-1} \tag{1.19}$$

$$\frac{\partial \mu}{\partial \eta'}=D'(\beta) \tag{1.20}$$

$$\frac{\partial \eta}{\partial \beta'}=Z \tag{1.21}$$

其中*1.19*中用到了逆映射定理. 于是由链式法则，有

$$\begin{aligned}S(\beta)=\frac{\partial l}{\partial \beta}&=\left(\frac{\partial l}{\partial \beta'}\right)'=\left(\frac{\partial l}{\partial \theta'}\frac{\partial \theta}{\partial \mu'}\frac{\partial \mu}{\partial \eta'}\frac{\partial \eta}{\partial \beta'}\right)'\\&=Z'D\Sigma^{-1}(y-\mu(\beta))\end{aligned} \tag{1.22}$$

同时上面的式子还可以写成

$$\frac{\partial l}{\partial \beta}=\left(\frac{\partial l}{\partial \theta'}\frac{\partial \theta}{\partial \eta'}\frac{\partial \eta}{\partial \beta'}\right)'=Z'\left(\frac{\partial \xi}{\partial \eta'}\right)'(y-\frac{\partial b}{\partial \theta})\Big/ a(\phi) \tag{1.23}$$

其中可记

$$\left(\frac{\partial \xi}{\partial \eta'}\right)'=\begin{bmatrix}\frac{\partial \xi_1}{\partial \eta_1} & \cdots & \frac{\partial \xi_n}{\partial \eta_1}\\ \vdots & & \vdots\\ \frac{\partial \xi_1}{\partial \eta_n} & \cdots & \frac{\partial \xi_n}{\partial \eta_n}\end{bmatrix}=\left(\frac{\partial \xi_1}{\partial \eta},\cdots,\frac{\partial \xi_q}{\partial \eta}\right)$$

又记 $y-\mu=\tau=(\tau_1,\cdots,\tau_q)'$，则有

$$
\begin{aligned}
F_{i,obs}(\beta) &= -\frac{\partial^2 l}{\partial\beta\partial\beta'} = -\frac{\partial}{\partial\beta'}\left(\frac{\partial l}{\partial\beta}\right) \\
&= -Z'\sum_{i=1}^{q}\left(\tau_i\frac{\partial^2\xi_i}{\partial\eta\partial\beta'} - \frac{\partial\xi_i}{\partial\eta}\frac{\partial\tau_i}{\partial\beta'}\right)\Big/a(\phi) \\
&= -\left(Z'\sum_{i=1}^{q}\left(\tau_i\frac{\partial^2\xi_i}{\partial\eta\partial\eta'}\frac{\partial\eta}{\partial\beta'}\right) + \left(\frac{\partial\xi}{\partial\eta'}\right)'\frac{\partial\tau}{\partial\beta'}\right)\Big/a(\phi) \qquad (1.24)
\end{aligned}
$$

其中

$$
\frac{\partial\xi}{\partial\eta'} = \frac{\partial\theta}{\partial\mu'}\frac{\partial\mu}{\partial\eta'} = a(\phi)\Sigma^{-1}D' \qquad (1.25)
$$

$$
\frac{\partial\tau}{\partial\beta'} = -\frac{\partial\mu}{\partial\beta} = -\frac{\partial\mu}{\partial\theta'}\frac{\partial\theta}{\partial\beta'} = -D'Z \qquad (1.26)
$$

代回*1.24*，即有

$$
\begin{aligned}
F_{i,bos}(\beta) &= Z'D\Sigma^{-1}DZ - \frac{1}{a(\phi)}\sum_{r=1}^{q}Z'\frac{\partial^2\xi_r}{\partial\eta\partial\eta'}Z(y_r-\mu_r) \qquad (1.27) \\
&= F_i(\beta) - \frac{1}{a(\phi)}R_i(\beta)
\end{aligned}
$$

没错这里与很多介绍 *GLM* 的书不同，最终推导得到的公式中 R 项会多出一个系数 $a(\phi)^{-1}$. 最早可以追溯到 *[Fahrmeir-Tutz,1994]* 的著作中的附录 A，那里就有遗漏. 当然可以去查看更原始的文献 *[Nelder and Wedderburn,1972]*，在另一种指数分布族的表达中建立了信息量的表达式，目测那里的推导没有遗漏任何系数.

1.6 参考资料 & 一些题外话

老师上课用的是周勇老师的教材，上来就考虑多元情形，属实把我上傻了，如果想系统地学习建议从 C. Radhakrishna Rao 入手，作者从 Y 为一元随机变量入手开始介绍，并举了很多常用的例子，如果觉得有学习多元情形的需求，再考虑阅读 Fahrmeir and Tutz 的书，毕竟那本题目就有 multivariate. 当然我只是学了个大概，然后到处抄书.

参考资料在这里列出，因为实在懒，就没搞严格的引用格式.

[Fahrmeir and Tutz,1994] Book Multivariate Statistical Model (Append A)

[Nelder and Wedderburn,1972] Generalized Linear Models

[C. Radhakrishna Rao, 1999] Generalized Linear Models (Chap 10)

[周勇,2013] 广义估计方程估计方法 (Chap 11)