7.2 Independent observations:

1. Covariance functions

Let

$$Y=\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, E(Y)=\mu=\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, Cov(Y)=\sigma^2 V(\mu),$$

where σ^2 may be unknown and $V(\mu)$ is a matrix of known functions. Suppose $Y_1, Y_2, \cdots, Y_n$ are independent and $Var(Y_i)$ depends only on μ_i. Then,

$$V(\mu)=\begin{bmatrix} V_1(\mu_1) & 0 & \cdots & 0 \\ 0 & V_2(\mu_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_n(\mu_n) \end{bmatrix}$$

Note:

$V_1(\cdot), V_2(\cdot), \cdots, V_n(\cdot)$ may be taken to be identical.

2. Quasi-likelihood functions

Motivating example:

Let $Y \sim N(\mu, \sigma^2)$. Then, the log-likelihood function is

$$l(\mu) \propto \frac{-(y-\mu)^2}{2\sigma^2}.$$

Thus, the score function for μ is

$$U(\mu)=\frac{dl(\mu)}{d\mu}=\frac{y-\mu}{\sigma^2}.$$

Note that

$$\int_y^{\mu} U(t)dt=\int_y^{\mu} \frac{y-t}{\sigma^2}dt=\frac{1}{\sigma^2}\cdot\left(yt-\frac{t^2}{2}\right)\Bigg|_y^{\mu}=\frac{1}{\sigma^2}\cdot\frac{-(y-\mu)^2}{2} \propto l(\mu).$$

Denote

$$U=\frac{Y-\mu}{\sigma^2 V(\mu)}.$$

U has the following properties in common with the score function:

$$E(U)=0, Var(U)=-E\left(\frac{\partial U(\mu)}{\partial \mu}\right)=\frac{1}{\sigma^2 V(\mu)}.$$

Then, since the score function is the derivative of the log-likelihood function, the integral

$$Q(\mu) = \int_y^{\mu} \frac{y-t}{\sigma^2 V(t)} dt,$$

if exists, should behave like a log-likelihood function for μ under the very mild assumptions. $Q(\mu)$ is referred to as the quasi-likelihood, or as the log quasi-likelihood for μ based on the data y.

Example:

Let $V(\mu) = \mu$. Then, $Var(Y) = \sigma^2\mu$ and

$$U = \frac{Y-\mu}{\sigma^2\mu}.$$

Then

$$\begin{aligned}
Q(\mu) &= \int_y^{\mu} \frac{y-t}{\sigma^2 t} dt = \frac{1}{\sigma^2}\left[\int_y^{\mu} \frac{y}{t} dt - \int_y^{\mu} 1 dt\right] \\
&= \frac{1}{\sigma^2}[y \cdot log(\mu) - \mu + y - y \cdot log(y)] \\
&\propto y \cdot log(\mu) - \mu \\
&\equiv log - likelihood\ of\ Poisson\ radom\ variable
\end{aligned}$$

Example:

Let $V(\mu) = \mu^2$. Then, $Var(Y) = \sigma^2\mu^2$ and

$$U = \frac{Y-\mu}{\sigma^2\mu^2}.$$

Then

$$\begin{aligned}
Q(\mu) &= \int_y^{\mu} \frac{y-t}{\sigma^2 t^2} dt = \frac{1}{\sigma^2}\left[\int_y^{\mu} \frac{y}{t^2} dt - \int_y^{\mu} \frac{1}{t} dt\right] \\
&= \frac{1}{\sigma^2}\left[1 - \frac{y}{\mu} - log(\mu) + log(y)\right] \\
&= \frac{1}{\sigma^2}\left[-\frac{y}{\mu} - log(\mu)\right] + \frac{1}{\sigma^2}[1 + log(y)] \\
&\propto -\frac{y}{\mu} - log(\mu) \\
&\equiv log - likelihood\ of\ gamma\ radom\ variable
\end{aligned}$$

Note that for a gamma random variable Y with $E(Y) = \mu$, then $Var(Y) = \mu^2$.

The quasi-likelihood function for the complete data is the sum of the individual contributions

$$Q(\mu) = \sum_{i=1}^{n} Q_i(\mu_i) = \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 V_i(t)} dt.$$

The quasi-deviance function corresponding to a single observation is

$$D(y,\mu) = -2\sigma^2 Q(\mu) = 2\int_{\mu}^{y} \frac{y-t}{V(t)} dt,$$

which does not depend on σ^2.

3. Parameter estimation

Let the vector of parameters β related to the dependence of μ on the covariate x. Therefore, we can write $\mu_i = \mu_i(\beta)$. Thus

$$\frac{\partial Q_i(\mu_i)}{\partial \beta_r} = \frac{\partial Q_i(\mu_i)}{\partial \mu_i}\frac{\partial \mu_i}{\partial \beta_r} = \frac{y_i - \mu_i}{\sigma^2 V_i(t)} \cdot D_{ir},$$

where

$$D_{ir} = \frac{\partial \mu_i}{\partial \beta_r}.$$

Therefore,

$$\frac{\partial Q(\mu)}{\partial \beta_r} = \sum_{i=1}^{n} \frac{\partial Q_i(\mu_i)}{\partial \beta_r} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\sigma^2 V_i(t)} \cdot D_{ir}$$

$$= \frac{1}{\sigma^2}[D_{ir} \quad D_{ir} \quad \cdots \quad D_{ir}] \begin{bmatrix} V_1^{-1}(\mu_1) & 0 & \cdots & 0 \\ 0 & V_2^{-1}(\mu_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_n^{-1}(\mu_n) \end{bmatrix} \begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \\ \vdots \\ y_n - \mu_n \end{bmatrix}$$

$$= \frac{1}{\sigma^2} D_r^t V^{-1}(y-\mu)$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, D_r = \begin{bmatrix} D_{1r} \\ D_{2r} \\ \vdots \\ D_{nr} \end{bmatrix}.$$

Further,

$$U(\beta) = \frac{\partial Q(\mu)}{\partial \beta} = \frac{1}{\sigma^2} D^t V^{-1}(y-\mu),$$

where

$$D_{n\times p} = [D_1 \quad D_2 \quad \cdots \quad D_p] = \begin{bmatrix} D_{11} & D_{12} & \cdots & D_{1p} \\ D_{21} & D_{22} & \cdots & D_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1} & D_{n2} & \cdots & D_{np} \end{bmatrix}.$$

The covariance matrix of $U(\beta)$ is

$$I(\beta) = Cov[U(\beta)] = -E\left[\frac{\partial U(\beta)}{\partial \beta}\right] = \frac{1}{\sigma^2} D^t V^{-1} D.$$

Note:

Under the usual limiting conditions on the eigenvalues of $I(\beta)$, the asymptotic variance-covariance matrix of $\widehat{\beta}$ is

$$Cov(\widehat{\beta}) = I^{-1}(\beta) = \sigma^2 (D^t V^{-1} D)^{-1}.$$

That is, $I(\beta)$ plays the same role as the Fisher's information for ordinary likelihood functions.

To obtain the parameter estimate $\widehat{\beta}$, the Fisher's scoring method is

$$\widehat{\beta}_{n+1} = \widehat{\beta}_n + \left(\widehat{D}_n^t \widehat{V}_n^{-1} \widehat{D}_n\right)^{-1} \widehat{D}_n^t \widehat{V}_n^{-1} (y - \widehat{\mu}_n), n = 0, 1, \cdots$$

where

$$\widehat{D}_n = [D]_{\beta=\widehat{\beta}_n}, \widehat{V}_n = [V]_{\beta=\widehat{\beta}_n}, \widehat{\mu}_n = [\mu]_{\beta=\widehat{\beta}_n}.$$

To estimate σ^2, we can use the following statistic

$$\tilde{\sigma}^2 = \frac{1}{n-p} \cdot \sum_{i=1}^{n} \frac{(y_i - \widehat{\mu}_i)^2}{V_i(\widehat{\mu}_i)} = \frac{X^2}{n-p},$$

where $\widehat{\mu}_i$ are the estimates of μ_i based on $\widehat{\beta}$ and X^2 is the generalized Pearson statistic.

Note:

$$Var(Y_i) = \sigma^2 V_i(\mu_i)$$

$$\Rightarrow \sigma^2 = \frac{Var(Y_i)}{V_i(\mu_i)} = Var\left[\frac{Y_i - \mu_i}{\sqrt{V_i(\mu_i)}}\right]$$

Thus, the estimator similar to the sample variance is

$$\frac{\sum_{I=1}^{n}\left[\frac{Y_i - \mu_i}{\sqrt{V_i(\mu_i)}}\right]^2}{n-p} = \frac{1}{n-p} \cdot \sum_{i=1}^{n} \frac{(Y_i - \mu_i)^2}{V_i(\mu_i)}.$$

The statistic $\tilde{\sigma}^2$ can be obtained by replacing μ_i by $\widehat{\mu}_i$.