Exponential Family Responses

Observing $Y|X \sim \text{Bin}(m, p(x))$ with $f(y) \propto p^y (1 - p)^{m-y}$, one may estimate $\eta = \log \{p/(1 - p)\}$ via the minimization of

$$-\frac{1}{n} \sum_{i=1}^{n} \{Y_i \eta(x_i) - m_i \log(1 + e^{\eta(x_i)})\} + \frac{1}{2} J(\eta).$$

Observing $Y|X \sim \text{Poisson}(\lambda(x))$ with $f(y) \propto \lambda^y e^{-\lambda}$, one may estimate $\eta = \log \lambda$ via the minimization of

$$-\frac{1}{n} \sum_{i=1}^{n} \{Y_i \eta(x_i) - e^{\eta(x_i)}\} + \frac{1}{2} J(\eta).$$

Observing $Y|X \sim \text{Gamma}(\alpha, \beta(x))$ with $f(y) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}$, one may estimate $\eta = \log(\alpha \beta)$ via the minimization of

$$\frac{1}{n} \sum_{i=1}^{n} \{Y_i e^{-\eta(x_i)} + \eta(x_i)\} + \frac{1}{2} J(\eta).$$

$L(\eta) = \frac{1}{n} \sum_i l_i(\eta(x_i))$, and the link $\eta$ is free of constraint.

Iterated Weighted Least Squares

Write $u = d1/d\eta$ and $w = d^2l/d\eta^2$. The quadratic approximation of $l_i(\eta)$ at $\tilde{\eta}_i = \tilde{\eta}(x_i)$ is seen to be $(1/2)\tilde{\eta}_i (\tilde{Y}_i - \eta)^2 + C_i$, where $\tilde{Y}_i = \tilde{\eta}_i - \tilde{u}_i / \tilde{w}_i$ and $C_i$ is a constant. One may iterate on

$$\frac{1}{n} \sum_{i=1}^{n} \tilde{w}_i (\tilde{Y}_i - \eta(x_i))^2 + \lambda J(\eta).$$

For exponential family with $f(y) \propto \exp\{ (y\theta - b(\theta))/a(\phi) \}$, one may assess the discrepancy between $(\eta, \phi)$ and $(\eta_\lambda, \phi)$ via

$$\text{KL}(\eta, \eta_\lambda) = \{ \mu (\theta - \theta_\lambda) - (b(\theta) - b(\theta_\lambda)) \},$$

where $\mu = E[Y] = \hat{b}(\theta)$, or $\text{SKL}(\eta, \eta_\lambda) = (\mu - \mu_\lambda)(\theta - \theta_\lambda)$.

Averaging over the sampling points, one has the loss

$L(\eta, \eta_\lambda) = \frac{1}{n} \sum_{i=1}^{n} (\mu - \mu_\lambda)(\theta - \theta_\lambda)(x_i) \approx \frac{1}{n} \sum_{i=1}^{n} w(\eta_i)(\eta - \eta_\lambda)^2(x_i)$.  

Note that $w = d^2l/d\eta^2 = \bar{b}(\theta)(d\theta/d\eta)^2 - (y - \mu)(d^2\theta/d\eta^2)$.  

C. Gu Fall 2015
Indirect Cross-Validation

\[ \tilde{Y}_i = \eta(x_i) + \tilde{\varepsilon}_i + o_p(1), \]
where \( E[\tilde{\varepsilon}_i] = 0 \) and \( E[\varepsilon_i^2] = a(\phi)/w_i \). Fixing \( \hat{\eta} \), \( U_w(\lambda) \) or \( V_w(\lambda) \) aims to minimize \( \frac{1}{n} \sum_{i=1}^n w_i \left( \eta_{\lambda, \hat{\eta}}(x_i) - \eta(x_i) \right)^2 \),
where \( \eta_{\lambda, \hat{\eta}} \) is the minimizer of the PWLS functional.

Iterating on the PWLS problem, one may jointly update \((\lambda, \hat{\eta})\), with the \( U_w(\lambda)/V_w(\lambda) \)-selected \( \eta_{\lambda, \hat{\eta}} \) as the new \( \hat{\eta} \). The iteration usually converges in 5-10 steps, but with no guarantee.

For \( a(\phi) \) known, \( U_w(\lambda) \) is preferred to \( V_w(\lambda) \).

Direct Cross-Validation

Consider \( \text{RKL}(\eta, \eta_{\lambda}) = \frac{1}{n} \sum_{i=1}^n \left\{ - \mu(x_i) \theta_{\lambda}(x_i) + b(\theta_{\lambda}(x_i)) \right\} \).
Replacing \( \mu(x_i) \theta_{\lambda}(x_i) \) by \( Y_i \theta_{\lambda}^{[i]}(x_i) \), one has

\[ V_0(\lambda) = -\frac{1}{n} \sum_{i=1}^n \left\{ Y_i \theta_{\lambda}(x_i) - b(\theta_{\lambda}(x_i)) \right\} + \frac{1}{n} \sum_{i=1}^n Y_i \left( \theta_{\lambda}(x_i) - \theta_{\lambda}^{[i]}(x_i) \right), \]

where \( \theta_{\lambda} - \theta_{\lambda}^{[i]} \approx (d\theta/d\eta)(\eta_{\lambda} - \eta_{\lambda}^{[i]}) = \zeta(\eta_{\lambda} - \eta_{\lambda}^{[i]}) \). Substituting \( \eta_{\lambda, \eta_{\lambda}}^{[i]}(x_i) \) for \( \eta_{\lambda}^{[i]}(x_i) \), and using the fact that

\[ \eta_{\lambda}(x_i) - \eta_{\lambda, \eta_{\lambda}}^{[i]}(x_i) = \frac{a_{i,i} - \tilde{u}_i}{1 - a_{i,i} \tilde{w}_i}, \]

some hand-waving yields, for \( \alpha = 1 \),

\[ V_g(\lambda) = \frac{1}{n} \sum_{i=1}^n l_i(\eta_{\lambda}(x_i)) + \alpha \frac{\text{tr}(A_w W^{-1})}{n - \text{tr}A_w} \frac{1}{n} \sum_{i=1}^n Y_i \zeta_i(-\tilde{u}_i). \]
Direct Cross-Validation

For logistic regression with $Y_i \sim \text{Bin}(m_i, p(x_i))$, the general procedure amounts to “delete-$m_i$.” Using $Y_i = \sum_{j=1}^{m_i} Y_{ij}$ for $Y_{ij}$ binary, one may derive the invariant version of CV.

Poisson regression is isomorphic to density estimation with binned data. The general procedure amounts to “delete-1-bin,” but one may employ the “delete-1” CV from density estimation.

A comparison of direct and indirect CV in logistic regression:

Inferential Tools

Based on the penalized weighted least squares problem at the converged $\eta_\lambda$, approximate posterior means and standard deviations can be computed, and in turn approximate Bayesian confidence intervals can be constructed.

The cosine diagnostics can also be calculated based on the penalized weighted least squares problem at the converged $\eta_\lambda$.

Define $\text{KL}(\hat{\eta}, \eta) = \frac{1}{n} \sum_i \{ \hat{\mu}_i(\hat{\theta}_i - \theta_i) - (b(\hat{\theta}_i) - b(\theta_i)) \}$. The KL projection $\tilde{\eta} \in \mathcal{H}_0$ of $\hat{\eta} \in \mathcal{H}_0 \oplus \mathcal{H}_1$ satisfies

$$\text{KL}(\hat{\eta}, \eta_c) = \text{KL}(\hat{\eta}, \tilde{\eta}) + \text{KL}(\tilde{\eta}, \eta_c) + \frac{1}{n} \sum_i (\tilde{\mu}_i - \hat{\mu}_i)(\tilde{\theta} - \theta_c)(x_i)$$

where $\sum_i (\tilde{\mu}_i - \hat{\mu}_i)(\tilde{\theta} - \theta_c)(x_i) \approx 0$ vanishes for $\eta = \theta$. 

C. Gu  Fall 2015
Software: gssanova in gss

Three gss suites are available for non-Gaussian regression, with gssanova implementing direct CV, and gssanova1 and gssanova0 implementing indirect CV. Eight families are implemented, which include binomial, Poisson, gamma, inverse Gaussian, and negative binomial, plus three more for survival data.

Only one link is implemented for each family: logit for binomial and negative binomial, and log for Poisson, gamma, and inverse Gaussian. The binomial and Poisson links are canonical.

The dispersion $a(\phi)$ is known for binomial, Poisson, and negative binomial, unknown for gamma and inverse Gaussian.

For the negative binomial family with $f(y) = \frac{\Gamma(\nu+y)}{\Gamma(\nu) y!} p^\nu (1-p)^y$, one may observe $(y_i, \nu_i)$ for $\nu_i$ known, or estimate a common $\nu$.

Spectral Density Estimation

For a stationary time series $\{X_t\}$ with $\gamma_k = \text{Cov}(X_t, X_{t+k})$, the spectral density is $f(\omega) = \frac{1}{\gamma_0} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i 2\pi k \omega}$, $\omega \in (-0.5, 0.5)$, which satisfies $\gamma_k = \gamma_0 \int_{-0.5}^{0.5} f(\omega) e^{i 2\pi k \omega} d\omega$; $f(\omega)$ is even.

Observing $x_t$, $t = 1, \ldots, T$, one may calculate the DFT

$$\hat{x}_\nu = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t e^{-i 2\pi t \nu / T},$$

yielding the periodogram $I(\omega_\nu) = |\hat{x}_\nu|^2$ on the Fourier frequencies $\omega_\nu = \nu / T$; $I(\omega_\nu) = I(\omega_{T-\nu})$.

For $T$ large, $I(\omega_\nu)$, $\omega_\nu \in (0, 0.5)$, are asymptotically independent exponential r.v.’s with $E[I(\omega_\nu)] \propto f(\omega_\nu)$.

```r
n <- length(x); ind <- 1:(ceiling(n/2)-1)
y <- (abs(fft(x))^2/n)[-1][ind]; xx <- ind/n
gssanova(y~xx,family="Gamma")
```