NOTES ON USING FISHER INFOMATION TO TUNE THE WORKING PARAMETERS

L. DUAN

Consider Bernoulli regression $y_i \sim Bern(g(x_i\theta))$:

Without augmented data z, Fisher information based on the marginal distribution is:

$$I(\theta|y) = X'diag\left\{\frac{\partial g(x_i\theta)/\partial x_i\theta}{\{g(x_i\theta)(1-g(x_i\theta))\}}\right\}X \tag{0.1}$$

Since we do not know θ, during adaptation, we plug in the MAP of θ (or simply the current value) as an approximate.

Conditionally on z, the information matrix with working parameters could take two possible forms:

(1) The conditional information is independent of z (e.g. probit),

$$I^*(\theta|y,z) = X'diag\{r_i\}X \tag{0.2}$$

In this case, simply setting $r_i = \frac{\partial g(x_i\theta)/\partial x_i\theta}{\{g(x_i\theta)(1-g(x_i\theta))}$ fully calibrates the difference between (0.1) and (0.2).

(2) The conditional information is dependent on z (e.g. logit),

$$I^*(\theta|y,z) = X'diag\{z_i\}X$$

In its corresponding CDA, we cannot multiply r_i directly on z_i since it would otherwise give intractable marginal. Instead, we inject r_i as a parameter into $\mathbb{E}z_i$. For example, in logit, this is $z_i \sim$ Polya-Gamma$(r_i, 0)$. Therefore, on the conditional Fisher information, we also obtain its expectation over the z.

$$\mathbb{E}_z I^*(\theta|y,z) = X'diag\{\mathbb{E}z_i(r_i)\}X \tag{0.3}$$

With r_i, we can also make $\mathbb{E}z_i(r_i) = \frac{\partial g(x_i\theta)/\partial x_i\theta}{\{g(x_i\theta)(1-g(x_i\theta))}$ and completely calibrate the difference between (0.1) and (0.3).