ARMY RESEARCH LABORATORY

Quantal Response: Practical Sensitivity Testing

by Joseph C. Collins

ARL-TR-6022 June 2012

NOTICES

Disclaimers

Army Research Laboratory

Aberdeen Proving Ground, MD 21005-5068

ARL-TR-6022 **June 2012**

Quantal Response: Practical Sensitivity Testing

Joseph C. Collins
Survivability/Lethality Analysis Directorate, ARL

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
June 2012	Final	1 October 2010 – 31 January 2012

4. TITLE AND SUBTITLE
Quantal Response: Practical Sensitivity Testing

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Joseph C. Collins

5d. PROJECT NUMBER
AH 80

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: RDRL-SLB-D
Aberdeen Proving Ground, MD 21005-5068

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-6022

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Sensitivity testing, or analysis of a binary quantal response to a continuous stimulus, has historically been modeled from first principles by assuming a single simple parameterization and imposing the form of a specific cumulative distribution function on the response. Application of the Generalized Linear Model approach admits arbitrary response functions and model complexity in a unified framework and subsumes the historical approach as a specific case. Estimation, hypothesis testing, and confidence interval computations are provided. Application to V50 ballistic limit armor acceptance testing is presented.

15. SUBJECT TERMS
sensitivity testing, quantal response, ballistic acceptance

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Joseph C. Collins
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (*Include area code*)
Unclassified	Unclassified	Unclassified	UU	50	410-278-6832

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

1. Background

The response (y) of an armor target to a ballistic threat can be characterized as penetration $(y = 1)$ or non-penetration $(y = 0)$. This is called a binary quantal response (QR), since there are exactly two possible outcomes. All other factors being constant, one may consider the effect of threat velocity (the stimulus) upon penetration (the response). The basic model for this interaction is that penetration is random, shots are independent, and that the probability of penetration is some function $G(v)$ of velocity,

$$\Pr[y = 1 \mid v] = G(v). \tag{1}$$

So y has a Bernoulli distribution conditional on velocity with Bernoulli parameter $G(v)$. Of course, G is bound between 0 and 1, and one expects that G is an increasing function of v. Thus, G has the functional form of a cumulative distribution function (cdf).

Of particular interest is the v_{50}, that velocity for which the probability of penetration equals 1/2.

$$G(v_{50}) = 0.5 \tag{2}$$

Analyses of the v_{50} and other quantities of interest are conducted by collecting samples of penetration responses and velocities, estimating the function G, and performing statistical inference on the results. This general paradigm is commonly known as "sensitivity analysis."

2. The Location-Scale Sensitivity Model

Quantal response (penetration) $y \in \{0,1\}$ to a continuous stimulus (velocity) v was originally modeled by using the standard normal cdf for a response function

$$G_o(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-u^2/2)\, du \tag{3}$$

and then estimating parameters m and s in the model

$$\mathrm{E}[y \mid v] = G_o\left(\frac{v - m}{s}\right) = \Pr[y = 1 \mid v], \tag{4}$$

where y has the indicated Bernoulli distribution, by the method of maximum likelihood (ML). This is the location-scale parameterization. In general, cdfs G_θ of location-scale distributions have the form

$$G_\theta(v) = G_o\left(\frac{v-\mu}{\sigma}\right), \tag{5}$$

where G_o is a standardized cdf. Then the parameter vector is $\theta = \begin{bmatrix}\mu\\ \sigma\end{bmatrix}$, where the location parameter is μ and the scale parameter is σ.

The aim of analysis is to estimate the parameter θ and then make inferences and perform statistical tests on meaningful population parameters such as v_{50}. With G_o chosen such that $G_o(0) = 0.5$, then $G_\theta(\mu) = 0.5$ and $\mu = v_{50}$. Thus, inferences on the parameter μ are, in fact, inferences on v_{50}.

3. The Linear Sensitivity Model

Response (penetration) $y \in \{0,1\}$ has expected value depending on stimulus (velocity) v,

$$\mathrm{E}[y \mid v] = G_o(b_0 + b_1 v). \tag{6}$$

The conditional distribution of y is Bernoulli with the indicated mean, so

$$\mathrm{E}[y \mid v] = G_o(b_0 + b_1 v) = \Pr[y = 1 \mid v]. \tag{7}$$

If G_o is an increasing function, then the response function G_o must be a cdf.

This is a linear parameterization,

$$G_\beta(v) = G_o(b_0 + b_1 v), \tag{8}$$

and the parameter vector is $\beta = \begin{bmatrix}b_0\\ b_1\end{bmatrix}$. To admit additional complexity, these models recognize the argument of G_o as a polynomial and thus have

$$\mathrm{E}[y \mid v] = G_o(b_0 + b_1 v + b_2 v^2 + \cdots + b_k v^k) \tag{9}$$

for finite k, where $k = 1$ in the previous example.

4. Generalized Linear Model (GLM) Estimation

Quantal response function estimation can be implemented using the Generalized Linear Model (GLM) with binomial response distribution and appropriate link functions. (See, for example, McCulloch and Searle.[1])

The response y has expected value depending on stimulus X and β (both vectors)

$$\mathrm{E}[y \mid X] = G_o(X\beta). \tag{10}$$

The distribution of y is Bernoulli with the indicated mean, so

$$\mathrm{E}[y \mid X] = G_o(X\beta) = \Pr[y = 1 \mid X]. \tag{11}$$

For sensitivity modeling the link function G_o is usually taken to be a continuous cdf.

Some authors call the inverse function $Q_o = G_o^{-1}$ the link.

GLM estimates the linear coefficients β in the linear model

$$Q_o\, \mathrm{E}[y \mid X] = X\beta. \tag{12}$$

Common choices for G_o include the normal cdf (*probit* link)

$$G_o\,(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-u^2/2)\, du \tag{13}$$

and the logistic cdf (*logit* link)

$$G_o(z) = \frac{1}{1 + e^{-z}}. \tag{14}$$

5. Reparameterization

Estimation and inference on the location-scale parameter θ is of particular interest, since $\mu = v_{50}$, but the usual GLM estimate provides the linear parameter β and its variance estimate V_β. Since

[1] McCulloch, C.; Searle, S. *Generalized, Linear, and Mixed Models*; John Wiley & Sons: New York, 2001.

$$b_0 + b_1 v = \frac{v - \mu}{\sigma} = -\frac{\mu}{\sigma} + \frac{1}{\sigma} v \,, \tag{15}$$

the location-scale parameter is obtained by

$$\theta = \begin{bmatrix} \mu \\ \sigma \end{bmatrix} = \begin{bmatrix} -b_0/b_1 \\ 1/b_1 \end{bmatrix}. \tag{16}$$

The location-scale variance estimate is given by the variance transformation of equation C-19 in appendix C.

$$V_\theta = \frac{d\theta^t}{d\beta} \cdot V_\beta \cdot \frac{d\theta}{d\beta}\,. \tag{17}$$

The required derivative is

$$\frac{d\theta}{d\beta} = \begin{bmatrix} d\mu/db_0 & d\sigma/db_0 \\ d\mu/db_1 & d\sigma/db_1 \end{bmatrix} = \begin{bmatrix} -1/b_1 & 0 \\ b_0/b_1^2 & -1/b_1^2 \end{bmatrix} = -\frac{1}{b_1^2} \begin{bmatrix} b_1 & 0 \\ -b_0 & 1 \end{bmatrix}. \tag{18}$$

In practice, to avoid numerical instability, computations are conducted using the standardized stimulus

$$u = \frac{v - v_m}{v_s}\,, \tag{19}$$

where v_m and v_s are, respectively, the sample mean and standard deviation of v, and the parameter vector is $\alpha = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$ so that

$$a_0 + a_1 u = \left(a_0 - \frac{a_1 v_m}{v_s}\right) + \frac{a_1}{v_s} v = b_0 + b_1 v = \frac{v - \mu}{\sigma} = -\frac{\mu}{\sigma} + \frac{1}{\sigma} v. \tag{20}$$

Thus, the location-scale parameter θ and its variance estimate V_θ can be recovered from the standardized linear parameter α and its variance estimate V_α. These are

$$\theta = \begin{bmatrix} \mu \\ \sigma \end{bmatrix} = \begin{bmatrix} v_m - v_s a_0/a_1 \\ v_s/a_1 \end{bmatrix} \tag{21}$$

and

$$V_\theta = \frac{d\theta^t}{d\alpha} \cdot V_\alpha \cdot \frac{d\theta}{d\alpha}, \tag{22}$$

where

$$\frac{d\theta}{d\alpha} = \begin{bmatrix} d\mu/da_0 & d\sigma/da_0 \\ d\mu/da_1 & d\sigma/da_1 \end{bmatrix} = \begin{bmatrix} -v_s/a_1 & 0 \\ v_s a_0/a_1^2 & -v_s/a_1^2 \end{bmatrix} = -\frac{v_s}{a_1^2}\begin{bmatrix} a_1 & 0 \\ -a_0 & 1 \end{bmatrix}. \tag{23}$$

Also, if need be, the usual linear parameter β and its variance estimate V_β can be recovered from the standardized linear parameter α and its variance estimate V_α. These are

$$\beta = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} a_0 - a_1 v_m/v_s \\ a_1/v_s \end{bmatrix} \tag{24}$$

and

$$V_\beta = \frac{d\beta^t}{d\alpha} \cdot V_\alpha \cdot \frac{d\beta}{d\alpha}, \tag{25}$$

where

$$\frac{d\beta}{d\alpha} = \begin{bmatrix} db_0/da_0 & db_1/da_0 \\ db_0/da_1 & db_1/da_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -v_m/v_s & 1/v_s \end{bmatrix} = \frac{1}{v_s}\begin{bmatrix} v_s & 0 \\ -v_m & 1 \end{bmatrix}. \tag{26}$$

---

6. Confidence Intervals

First, consider velocity confidence intervals on v_p for fixed probability of penetration p (see figure 1). The response function gives probability of penetration at velocity v,

$$G_\theta(v) = G_o\left(\frac{v-\mu}{\sigma}\right), \quad \text{where} \quad \theta = \begin{bmatrix} \mu \\ \sigma \end{bmatrix}. \tag{27}$$

The nominal point estimate of v_p, the velocity at which the probability of penetration is p, is expressed in terms of the quantile function $Q = G^{-1}$ as

$$v_p = Q_\theta(p) = \mu + \sigma Q_o(p) = K\theta, \quad \text{where} \quad K = [1 \quad Q_o(p)]. \tag{28}$$

Confidence intervals are calculated from the estimator distribution $\theta \sim N_2(M_\theta, V_\theta)$. So

$$K\theta = N(KM_\theta, KV_\theta K^t). \tag{29}$$

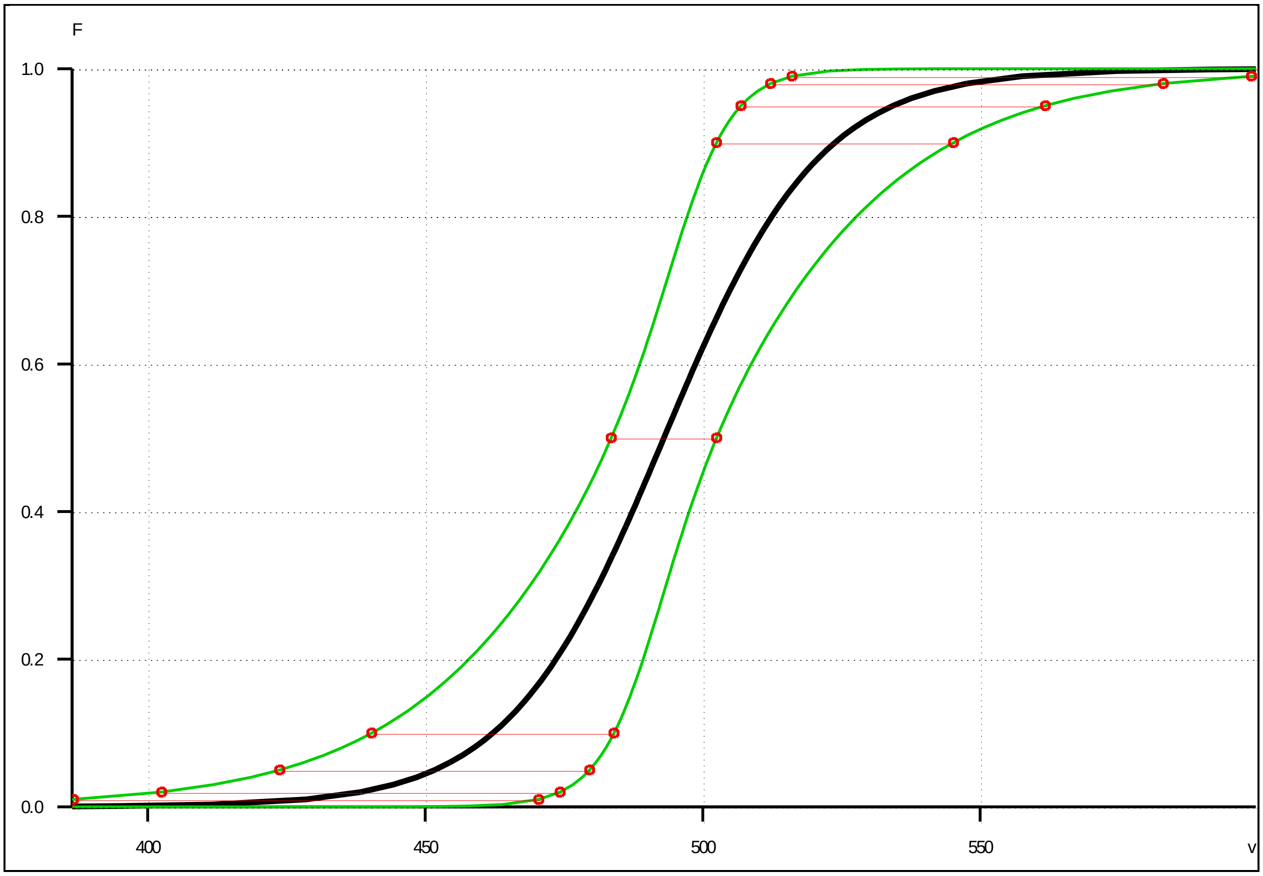

Figure 1. Estimate and normal theory $\boldsymbol{v_p}$ confidence intervals.

Since $K\theta \sim KM_\theta + \sqrt{KV_\theta K^t} \cdot Z$ with $Z \sim N(0,1)$, it follows that a $100(1-\alpha)\%$ two-sided confidence interval on v_p is given in the usual manner by

$$\left[KM_\theta - \sqrt{KV_\theta K^t} \cdot Z_{1-\alpha/2} \text{ and } KM_\theta + \sqrt{KV_\theta K^t} \cdot Z_{1-\alpha/2}\right]. \tag{30}$$

The quantile point Z_q of the standard normal distribution satisfies $\Pr[Z \le Z_q] = q$.

It is also possible to calculate confidence intervals on the probability of penetration p for fixed velocity v (see figure 2). Start with the linear parameterization of the response

$$G_\beta(v) = G_o(b_0 + b_1 v) = G_o(K\beta)\,, \quad \text{where} \quad K = [1 \quad v] \quad \text{and} \quad \beta = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}. \tag{31}$$

Confidence intervals are calculated from the estimator distribution $\beta \sim N_2(M_\beta, V_\beta)$. So

$$K\beta = N\left(KM_\beta, KV_\beta K^t\right). \tag{32}$$

Since $K\beta \sim KM_\beta + \sqrt{KV_\beta K^t} \cdot Z$ where, again, $Z \sim N(0,1)$, it follows that a $100(1-\alpha)\%$ two-sided confidence interval on p is given by

$$\left[G_o\left(KM_\beta - \sqrt{KV_\beta K^t} \cdot Z_{1-\alpha/2}\right) \text{ and } G_o\left(KM_\beta + \sqrt{KV_\beta K^t} \cdot Z_{1-\alpha/2}\right)\right]. \tag{33}$$

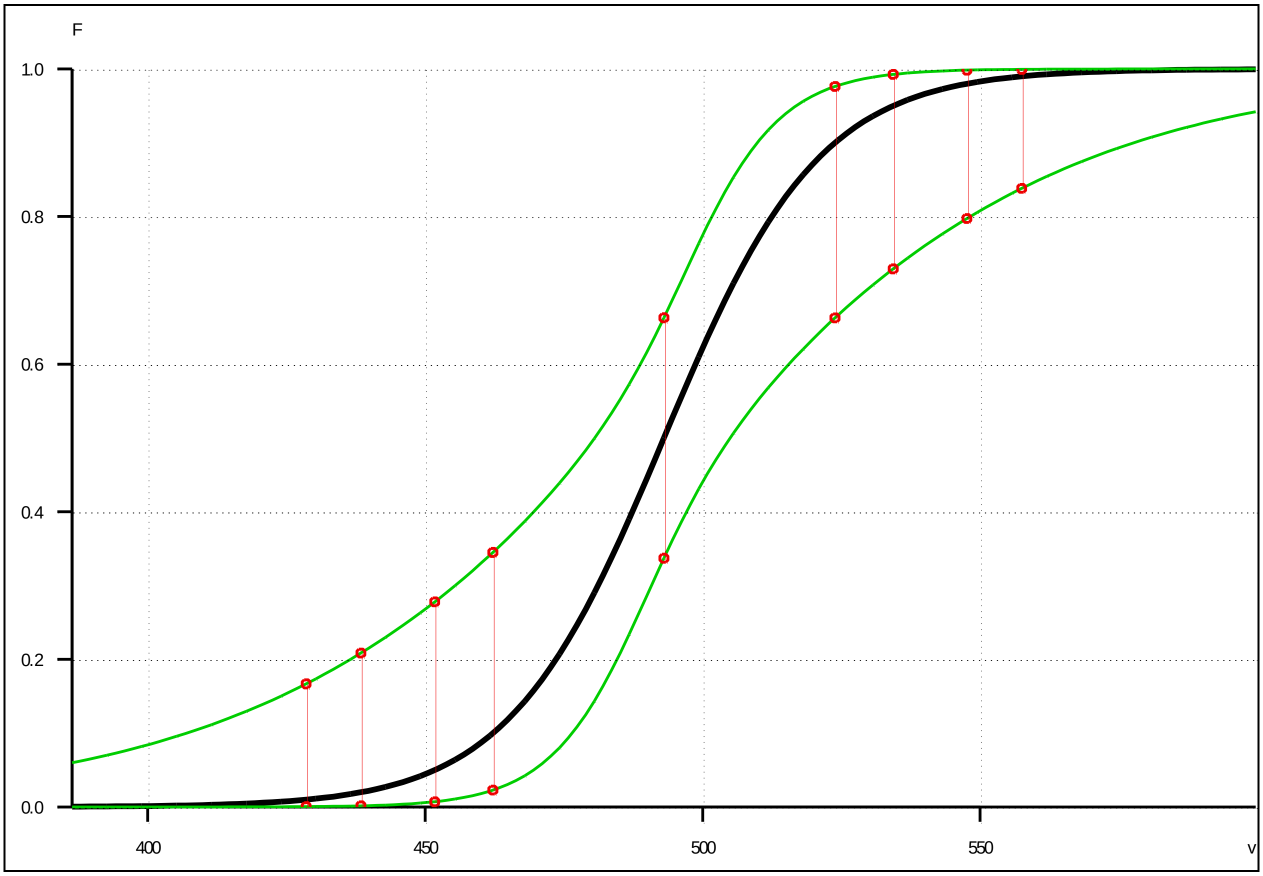

Figure 2. Estimate and normal theory $\boldsymbol{p}$ confidence intervals.

7. Hypothesis Testing

The theory of appendix B provides a means for hypothesis tests on response curve parameters, and, in particular, for comparing v_{50} estimates against each other.

The quadratic forms derived from asymptotic normal distributions of maximum likelihood estimators are known as *Wald's Statistics*, and the tests are called *Wald's Tests*.

Suppose that n experiments and have given n response curve parameter estimates $\{\hat{\theta}_i : i = 1, \ldots, n\}$ with, for example, the probit (cumulative normal) response function in terms of the usual location-scale parameter $\theta = \begin{bmatrix} \mu \\ \sigma \end{bmatrix}$, where $\hat{\theta}_i = \begin{bmatrix} m_i \\ s_i \end{bmatrix}$. So the model for the response is cumulative normal with mean μ and standard deviation σ.

Since the mean is the median, $\mu = v_{50}$. The other parameter σ characterized the steepness of the response curve. Inferences about the parameters are inferences about v_{50} and the steepness of the response. So there are n sets of estimates and their covariance matrices. For $i = 1, \ldots, n$,

$$\hat{\theta}_i = \begin{bmatrix} m_i \\ s_i \end{bmatrix} \quad \text{and} \quad \hat{V}_i = \begin{bmatrix} v_{i11} & v_{i12} \\ v_{i12} & v_{i22} \end{bmatrix}. \tag{34}$$

In terms of appendix B,

$$X = \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_n \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} \hat{V}_1 & & 0 \\ & \ddots & \\ 0 & & \hat{V}_n \end{bmatrix}, \tag{35}$$

and the test is constructed by choosing K (called a contrast matrix in this context) to compare certain elements of X.

$$H_0 : KX = 0 \text{ and } H_1 : KX \neq 0. \tag{36}$$

The test statistic is the associated quadratic form

$$S = (KX)^t (KVK^t)^{-1} KX. \tag{37}$$

Under the null hypothesis, the distribution of S is central chi-square

$$S \sim \chi_r^2 , \tag{38}$$

and the alternative distribution with true parameter value $\theta = M$ is noncentral chi-square

$$S \sim \chi_{r,\delta}^2 , \tag{39}$$

where $r = \operatorname{rank} K$ and the noncentrality parameter is $\delta = (KM)^t (KVK^t)^{-1} KM$.

Note: Any random variable Z has cumulative distribution function (cdf) $F_Z(t) = \Pr\,[Z \leq t]$ and quantile function $Q_Z(u) = \inf\{x : F_Z(x) \geq u\}$.

A test with type I error (probability of rejecting a true H_0) equal to α has critical value S_0, where

$$\alpha = \Pr[S > S_0 \mid H_0] \,, \tag{40}$$

since large values of S are significant, and S exceeds S_0 with H_0-probability α. Thus, the critical value is given by

$$S_0 = Q_{\chi_r^2}\,(1-\alpha). \tag{41}$$

Based on the observed value $\hat{S}$ of S, the p-value of an experiment is

$$p = \Pr[S > \hat{S} \mid H_0] = 1 - F_{\chi_r^2}\,(\hat{S}). \tag{42}$$

The decision rule is to reject H_0 if $\hat{S} > S_0$, or, equivalently, if $p < \alpha$.

For a fixed alternative $\theta = M$ with $KM \neq 0$, the type II error β is the probability of not rejecting the null hypothesis H_0 under H_1, when H_0 is false,

$$\beta = \Pr[S < S_0 \mid H_1] = F_{\chi_{r,\delta}^2}\,(S_0) = F_{\chi_{r,\delta}^2}\,\left(Q_{\chi_r^2}(1-\alpha)\right). \tag{43}$$

The power q of the test is the probability of detecting the alternative,

$$q = 1 - \beta = 1 - F_{\chi_{r,\delta}^2}\,(S_0) = 1 - F_{\chi_{r,\delta}^2}\,\left(Q_{\chi_r^2}(1-\alpha)\right). \tag{44}$$

Illustrations of specific tests follow. The test for v_{50} equality is

$$H_0\colon \mu_1 = \mu_2 \text{ and } H_1\colon \mu_1 \neq \mu_2. \tag{45}$$

The contrast is

$$K = [1 \quad 0 \quad -1 \quad 0 \quad \cdots]. \tag{46}$$

Then

$$KX = m_1 - m_2 \quad \text{and} \quad KVK^t = v_{111} + v_{211} \,, \tag{47}$$

and the quadratic form test statistic is

$$S = \frac{(m_1 - m_2)^2}{v_{111} + v_{211}} \sim \chi_1^2. \tag{48}$$

Under the alternative hypothesis with true difference in mean $\mu_1 - \mu_2 = \Delta$, the test statistic has the noncentral chi-square distribution

$$S \sim \chi^2_{1,\delta}\ , \tag{49}$$

where

$$\delta = \frac{\Delta^2}{v_{111} + v_{211}}, \tag{50}$$

and the test has power

$$q = 1 - \beta = 1 - F_{\chi^2_{1,\delta}}\left(S_0\right) = 1 - F_{\chi^2_{1,\delta}}\left(Q_{\chi^2_1}(1-\alpha)\right). \tag{51}$$

To compare response curves for location and scale, test

$$H_0\colon (\mu_1, \sigma_1) = (\mu_2, \sigma_2) \text{ and } H_1\colon (\mu_1, \sigma_1) \neq (\mu_2, \sigma_2) \tag{52}$$

and use the contrast

$$K = \begin{bmatrix} 1 & 0 & -1 & 0 & \cdots \\ 0 & 1 & 0 & -1 & \cdots \end{bmatrix}. \tag{53}$$

The test statistic S is χ^2_2, with $d_m = m_1 - m_2$ and $d_s = s_1 - s_2$,

$$S = \frac{d_s^2(v_{111} - v_{211}) - 2d_s d_m(v_{112} - v_{212}) + d_m^2(v_{122} - v_{222})}{(v_{111} - v_{211})(v_{122} - v_{222}) - (v_{112} + v_{212})^2}. \tag{54}$$

To compare four v_{50} estimates, one can safely discard the σ information and use $v_i = v_{ii}$ in

$$X = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix} \quad and \quad V = \begin{bmatrix} v_1 & 0 & 0 & 0 \\ 0 & v_2 & 0 & 0 \\ 0 & 0 & v_3 & 0 \\ 0 & 0 & 0 & v_4 \end{bmatrix}. \tag{55}$$

Tests for pairwise comparisons are as those just presented.

$$H_0\colon \mu_i = \mu_j \text{ and } H_1\colon \mu_i \neq \mu_j\ , \tag{56}$$

where

$$S = \frac{\left(m_i - m_j\right)^2}{v_i + v_j} \sim \chi_1^2. \tag{57}$$

To compare all four v_{50} estimates,

$$H_0\colon (\forall i,j)\ \mu_i = \mu_j \text{ and } H_1\colon (\exists i,j) \mu_i \neq \mu_j. \tag{58}$$

Use

$$K = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \tag{59}$$

or, equivalently,

$$K = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}. \tag{60}$$

Both give the same $S \sim \chi_3^2$. The explicit form is tedious, but reasonable software works directly with the matrices anyway.

To compare the first against the mean of the other three,

$$H_0\colon \mu_1 = (\mu_2 + \mu_3 + \mu_4)/3 \text{ and } H_1\colon \mu_1 \neq (\mu_2 + \mu_3 + \mu_4)/3, \tag{61}$$

use

$$K = [3 \quad -1 \quad -1 \quad -1]. \tag{62}$$

Then $S \sim \chi_1^2$.

8. Computation

A historical implementation of quantal response computation has been documented by McKaig and Thomas.[2] The original computations involved were derived from first principles. (See, for example, DARCOM P 706-103.[3]) The code (written in FORTRAN) must be compiled for use on each particular platform.The modern approach in this report uses GLM explicitly. (See appendix E for details on Maximum Likelihood Estimation (MLE) for the GLM.) This technique provides computationally stable solutions for polynomial type models by iterated systems of linear

[2]McKaig, A. E.; Thomas, J. M. Likelihood Program for Sequential Testing Documentation; ARBRL-TR-02481; U.S. Army Ballistics Research Laboratory: Aberdeen Proving Ground, MD, 1983.

[3]DARCOM-P 706-103. *Engineering Design Handbook: Selected Topics in Experimental Statistics with Army Applications*; DARCOM pamphlet, U.S. Army Materiel Development and Readiness Command, 1983.

equations. GLM also provides diagnostics and estimator distributions required for inference. Furthermore, written in Java, the same code runs on Windows, Linux, and Mac OS X.

GLM estimation is implemented by iterative reweighted least squares (IRLS) maximization of the deviance function (which is linearly related to the log likelihood). For the natural link, in this case logit, IRLS is equivalent to the Newton-Raphson method of solving the MLE score equation

$$\mathcal{L}'(\beta) = 0 \, , \tag{63}$$

where $\mathcal{L} = -\log L$. The estimator sequence $\{\beta_i : i = 0,1,2 \ldots\}$ begins with an intial guess β_0. Subsequent elements are generated by solution of the linear differential approximation

$$\mathcal{L}'(\beta_i) + (\beta_{i+1} - \beta_i)\mathcal{L}''(\beta_i) = 0. \tag{64}$$

For the other links, IRLS is equivalent to Fisher scoring, another method for obtaining the MLE, in which the Hessian matrix $\mathcal{L}''$ is replaced by its expected value

$$\mathcal{L}'(\beta_i) + (\beta_{i+1} - \beta_i)\, \mathrm{E}[\mathcal{L}''(\beta_i)] = 0. \tag{65}$$

In all cases, some stopping rule determines termination of the estimation sequence.

This stripped-down Java program illustrates the Fisher update algorithm computation for the logistic model. The linear parameter estimate is `b`, and the information matrix is `M` in the code.

```
////////////////////////////////////////////////////////////////
// QrDemo.java

import static java.lang.Math.*;

public class QrDemo {

    public QrDemo(int n, double[] x, int[] y) {

        double d, dev, dev0=Double.POSITIVE_INFINITY, v, w, A[]={0, 0},
            M[][]={{0,0},{0,0}}, mu[]=new double[n], eta[]=new double[n], b[]={0,0};

        for (int iterations = 1; iterations <= 64; iterations++) {

            dev = 0;
            for (int i = 0; i < n; i++) {
                eta[i] = b[0] + b[1] * x[i];
                mu[i] = 1/(1+exp(-eta[i]));
                dev += y[i] == 1 ? log(mu[i]) : log(1-mu[i]);
            }
            dev *= -2;

            System.out.printf("%2d: dev = %23.16e , b[] = (%23.16e, %23.16e)\n",
                    iterations, dev, b[0], b[1]);

            if (abs((dev0-dev)/dev) < 1e-16) { break; }
            dev0 = dev;

            A[0] = A[1] = M[0][0] = M[0][1] = M[1][0] = M[1][1] = 0;
            for (int i = 0; i < n; i++) {
                v = mu[i] * (1 - mu[i]);
                w = y[i] - mu[i];
                A[0] += w;
                A[1] += w * x[i];
                M[0][0] += v;
                M[0][1] += v * x[i];
                M[1][1] += v * x[i] * x[i];
            }
            d = M[0][0]*M[1][1] - M[0][1]*M[0][1];
            b[0] += (  A[0]*M[1][1] - A[1]*M[0][1] ) / d;
            b[1] += (- A[0]*M[0][1] + A[1]*M[0][0] ) / d;
        }

        System.out.printf("b[] = (%1.5f %1.5f)\n", b[0] , b[1]);
    }

    public static void main(String args[]) {

        int n=10;
        double x[] = { 2620, 2667, 2717, 2718, 2721, 2724, 2744, 2811, 2840, 3020};
        int    y[] = {    0,    0,    0,    0,    1,    1,    0,    1,    1,    1};

        new QrDemo(n, x, y);
    }
}

////////////////////////////////////////////////////////////////
```

Compiling and running the program exposes the deviance convergence and estimator sequence.

```
$ javac QrDemo.java

$ java QrDemo

 1: dev =  1.3862943611198906e+01 , b[] = ( 0.0000000000000000e+00,  0.0000000000000000e+00)
 2: dev =  9.4193618456986190e+00 , b[] = (-3.2157347386093360e+01,  1.1658816396959381e-02)
 3: dev =  8.1337185880127830e+00 , b[] = (-5.9172217539474130e+01,  2.1560777823503036e-02)
 4: dev =  7.6605066309287860e+00 , b[] = (-8.6314382350221580e+01,  3.1517928280382454e-02)
 5: dev =  7.5662084690745290e+00 , b[] = (-1.0466244931233070e+02,  3.8253780626049190e-02)
 6: dev =  7.5591586884309470e+00 , b[] = (-1.1142261623953550e+02,  4.0736192959547696e-02)
 7: dev =  7.5590975468338840e+00 , b[] = (-1.1213096248281227e+02,  4.0996302051673934e-02)
 8: dev =  7.5590975412867530e+00 , b[] = (-1.1213779766541855e+02,  4.0998811895006204e-02)
 9: dev =  7.5590975412867560e+00 , b[] = (-1.1213779829230498e+02,  4.0998812125193504e-02)
10: dev =  7.5590975412867385e+00 , b[] = (-1.1213779829230472e+02,  4.0998812125193410e-02)
11: dev =  7.5590975412867440e+00 , b[] = (-1.1213779829230428e+02,  4.0998812125193250e-02)
12: dev =  7.5590975412867530e+00 , b[] = (-1.1213779829230556e+02,  4.0998812125193710e-02)
13: dev =  7.5590975412867530e+00 , b[] = (-1.1213779829230532e+02,  4.0998812125193630e-02)

b[] = (-112.13780 0.04100)
```

---

9. Applications

The code can be extended to compute parameter, variance, and correlation estimates for different parameterizations and display the results along with a Lower Confidence Bound (LCB) on the v_{50}. This example implements the logit link with Fisher scoring optimization, using data with stimulus (velocity) in the first column and response (penetration) in the second column:

```
600.000  1
579.500  0
580.400  0
616.400  0
626.200  1
627.000  1
599.800  0
614.900  1
613.100  1
575.000  0
```

In practice, one works with the standardized predictor u and linear parameter `a[]` $= \alpha = (a_0, a_1)$ of section 5 and computes its variance `Va` $= V_\alpha$ and correlation `Ra`.

```
a[]  = (-0.19415,   2.2279), Va  = ( 0.88564, -0.35817, . ,   1.7373),
Ra = -0.28876
```

Transformations provide the usual linear and location-scale parameterizations `b[]` $= \beta$ and `ms[]=(mu,sg)` $= \theta$ and their variances `Vb` $= V_\beta$ and `Vms` $= V_\theta$ and correlations.

```
b[]  = ( -73.045,  0.12077), Vb  = (  1881.9,  -3.0988, . , 0.0051048),
```

```
Rb = -0.99978

ms[] = (  604.84,   8.2804), Vms = (  57.348,  -6.3638, . ,   23.998),
Rms = -0.17154

Logit Response Parameters: ( mu , sg ) = (   604.84 ,   8.2804 )
```

The location-scale version is required for confidence bounds on v_{50}, since its standard deviation comes from V_θ. The standard normal 95% quantile `Z_95` =1.645 gives a 95% LCB on v_{50}.

```
V50 estimate           = mu    =   604.84
V50 standard deviation = SD.mu =   7.5728

95% LCB on V50 = mu - SD.mu*Z_95 = 592.38
```

Computations must take place in one of the linear versions, and the standardized version avoids numerical problems caused by collinearity, as indicated by the extreme correlation `Rb` of the raw linear version.

LangMod (Collins and Moss[4]) is a Java GUI implementation of a modified Langlie sequential strategy for quantal response testing. LangMod has been used to support various customer and research programs for testing personal protective equipment as well as ground and aircraft targets. LangMod incorporates (among other things) the logistic regression calculations developed in this report and displays QR calculation results and a graph of the data and response function estimate as shown in figure 3.

[4]Collins, J.; Moss, L. *LangMod Users Manual*; ARL-TN-437; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2011.

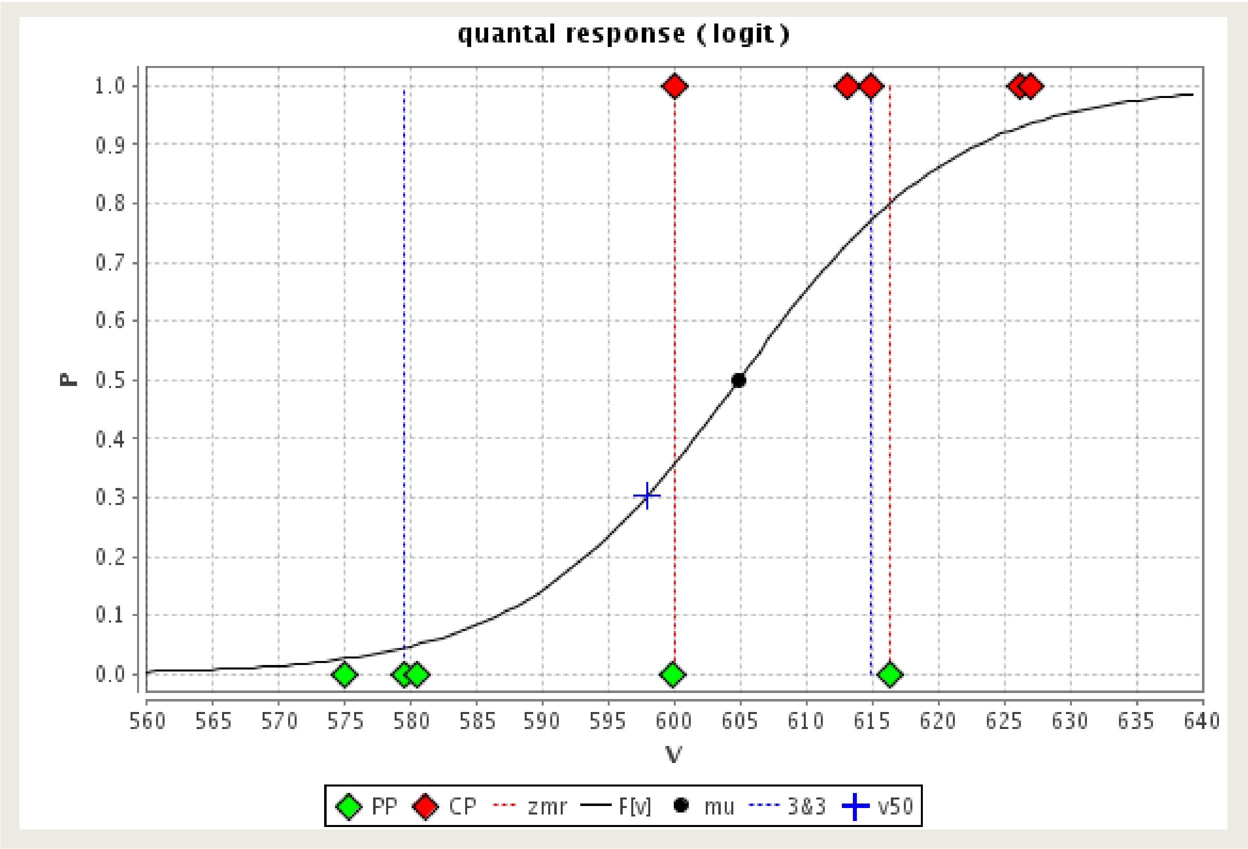

Figure 3. Quantal response from LangMod.

Similar capabilities are implemented in a local S-PLUS library, `libv50`, which has been used in support of various projects (see, for example, Collins et al.[5]). This library uses the native S-PLUS GLM computations. For example, in S-PLUS, the GLM computation can be implemented as

```
fit <- glm(x~v, family=binomial (link=logit), data=z)
```

where the data frame z contains stimulus velocity v and response penetration $x \in \{0,1\}$. Then, `glm` returns model coefficient estimates $\beta = (b_0, b_1)$ in `fit$coefficients` and estimated parameter variance matrix V_β in `summary(fit)$cov.unscaled`. The library implements the reparameterization, hypothesis testing, and confidence interval procedures outlined in this report. The computations and graphics for figures 1 and 2 were generated by `libv50`.

[5]Collins, J., et al. *Sensitivity of Mounting Methods or the Outer Tactical Vest and Shoot Pack Ballistic Limit, Phase I: Current Mounting Methods*; ARL-TR-4116; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2007.

Appendix A. Vector Function Conventions

A.1 Notation

Vectors are columns. With $x \in \mathbb{R}^n$,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \tag{A-1}$$

and if $f: \mathbb{R}^n \to \mathbb{R}^k$, then $f(x) \in \mathbb{R}^k$ and

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_k(x) \end{bmatrix} = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ f_2(x_1, \dots, x_n) \\ \vdots \\ f_k(x_1, \dots, x_n) \end{bmatrix}. \tag{A-2}$$

If A is an $n \times k$ matrix, the a_{ij} denotes the element in row i and column j. The transpose of A,

$$B = A^t\ , \tag{A-3}$$

is a $k \times n$ matrix and has $b_{ij} = a_{ji}$.

A.2 Inner Products and Norms

The usual inner product is

$$\langle x, y \rangle = x^t y = \sum_{i=1}^{n} x_i y_i\ , \tag{A-4}$$

and a weighted inner product is

$$\langle x, y \rangle_W = x^t W y = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i w_{ij} y_j. \tag{A-5}$$

The associated (squared) norms are

$$\|x\|^2 = \langle x, x \rangle = x^t x = \sum_{i=1}^{n} x_i^2 \tag{A-6}$$

and

$$\|x\|_W^2 = x^t W x = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i w_{ij} x_j. \tag{A-7}$$

In the event that W is a diagonal matrix,

$$\langle x, y\rangle_W = x^t W y = \sum_{i=1}^{n} x_i w_{ii} y_i\ , \tag{A-8}$$

and

$$\|x\|_W^2 = \langle x, x\rangle_W = \sum_{i=1}^{n} w_{ii} {x_i}^2. \tag{A-9}$$

A.3 Derivatives

The derivative of f is the $n \times k$ matrix function in which column j is the derivative df_j/dx of the coordinate j scalar field f_j. So the element in row i and column j is df_j/dx_i, and the derivative is

$$\frac{d}{dx} f(x) = \begin{bmatrix} \frac{df_1}{dx_1} & \frac{df_2}{dx_1} & \cdots & \frac{df_k}{dx_1} \\ \frac{df_1}{dx_2} & \frac{df_2}{dx_2} & \cdots & \frac{df_k}{dx_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{df_1}{dx_n} & \frac{df_2}{dx_n} & \cdots & \frac{df_k}{dx_n} \end{bmatrix}. \tag{A-10}$$

In particular, the derivative of a scalar field $f\colon \mathbb{R}^n \to \mathbb{R}^1$ is a column vector

$$\frac{df}{dx} = \begin{bmatrix} \frac{df}{dx_1} \\ \frac{df}{dx_2} \\ \vdots \\ \frac{df}{dx_n} \end{bmatrix}. \tag{A-11}$$

The linearization of f at x_o is

$$f(x_o+\delta) = f(x_o) + \left(\frac{d}{dx}f(x_o)\right)^t \cdot \delta. \tag{A-12}$$

If $f: \mathbb{R}^n \to \mathbb{R}^k$ is given by a linear transformation with $k \times n$ matrix A,

$$f(x) = Ax = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{kj}x_j \end{bmatrix}. \tag{A-13}$$

Then the derivative is the $n \times k$ matrix

$$\frac{df}{dx} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ a_{12} & a_{22} & \cdots & a_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{kn} \end{bmatrix} = A^t , \tag{A-14}$$

and $f(x_o+\delta) = f(x_o) + A\delta$, as expected.

$$f(x) = x^t A = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{nk} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i a_{i1} \\ \sum_{i=1}^n x_i a_{i2} \\ \vdots \\ \sum_{i=1}^n x_i a_{ik} \end{bmatrix}. \tag{A-15}$$

Then the derivative is the $n \times k$ matrix

$$\frac{df}{dx} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{bmatrix} = A. \tag{A-16}$$

So derivatives of inner products and (squared) norms are

$$\frac{d}{dx}\langle x,y\rangle = \frac{d}{dx}\langle y,x\rangle = y , \tag{A-17}$$

$$\frac{d}{dx}\langle x,y\rangle_W = Wy \quad \text{and} \quad \frac{d}{dx}\langle y,x\rangle_W = W^t y , \tag{A-18}$$

$$\frac{d}{dx}\|x\|^2 = 2x , \tag{A-19}$$

$$\frac{d}{dx}\|x\|_W{}^2 = (W + W^t)x\,, \tag{A-20}$$

and

$$\frac{d}{dx}\|x - a\|_W{}^2 = (W + W^t)(x - a). \tag{A-21}$$

Some second derivatives are

$$\frac{d^2}{dxx^t}\|x\|^2 = 2I \tag{A-22}$$

and

$$\frac{d^2}{dxx^t}\|x\|_W{}^2 = W + W^t. \tag{A-23}$$

A.4 The Chain Rule

If x itself is a function with x: $\mathbb{R}^m \to \mathbb{R}^n$, write x = x(u) $\in \mathbb{R}^n$ for u $\in \mathbb{R}^m$.

Then $g = f \circ x : \mathbb{R}\char`^ m \to \mathbb{R}^k$ and the elements of the $m \times k$ matrix $\frac{dg}{du}$ are $\frac{dg_j}{du_i} = \sum_{p=1}^{n} \frac{df_j}{dx_p}\frac{dx_p}{du_i}$ for $i = 1, \dots, m$ and $j = 1, \dots, k$. So, the derivative of g is

$$\frac{dg}{du} = \begin{bmatrix} \sum_{p=1}^{n}\frac{df_1}{dx_p}\frac{dx_p}{du_1} & \sum_{p=1}^{n}\frac{df_2}{dx_p}\frac{dx_p}{du_1} & \cdots & \sum_{p=1}^{n}\frac{df_k}{dx_p}\frac{dx_p}{du_1} \\ \sum_{p=1}^{n}\frac{df_1}{dx_p}\frac{dx_p}{du_2} & \sum_{p=1}^{n}\frac{df_2}{dx_p}\frac{dx_p}{du_2} & \cdots & \sum_{p=1}^{n}\frac{df_k}{dx_p}\frac{dx_p}{du_2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{p=1}^{n}\frac{df_1}{dx_p}\frac{dx_p}{du_m} & \sum_{p=1}^{n}\frac{df_2}{dx_p}\frac{dx_p}{du_m} & \cdots & \sum_{p=1}^{n}\frac{df_k}{dx_p}\frac{dx_p}{du_m} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{dx_1}{du_1} & \frac{dx_2}{du_1} & \cdots & \frac{dx_n}{du_1} \\ \frac{dx_1}{du_2} & \frac{dx_2}{du_2} & \cdots & \frac{dx_n}{du_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dx_1}{du_m} & \frac{dx_2}{du_m} & \cdots & \frac{dx_n}{du_{1m}} \end{bmatrix} \cdot \begin{bmatrix} \frac{df_1}{dx_1} & \frac{df_2}{dx_1} & \cdots & \frac{df_k}{dx_1} \\ \frac{df_1}{dx_2} & \frac{df_2}{dx_2} & \cdots & \frac{df_k}{dx_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{df_1}{dx_n} & \frac{df_2}{dx_n} & \cdots & \frac{df_k}{dx_n} \end{bmatrix} = \frac{dx}{du} \cdot \frac{df}{dx}. \tag{A-24}$$

Thus follows the chain rule for vector fields

$$\frac{d}{du}f\big(x(u)\big) = \frac{d}{du}x(u) \cdot \frac{d}{dx}f(x)_{|x=x(u)}. \tag{A-25}$$

In particular, with $\mathrm{x} = \mathrm{f}^{-1}$ and $\mathrm{f}\big(\mathrm{x}(\mathrm{u})\big) = \mathrm{u}$, then $\frac{\mathrm{df}}{\mathrm{du}} = \mathrm{I}$ and $\frac{\mathrm{d}}{\mathrm{du}}\mathrm{f}^{-1}(\mathrm{u}) = \left(\frac{\mathrm{d}}{\mathrm{dx}}\mathrm{f}(\mathrm{x})\right)^{-1}_{\mathrm{x}=\mathrm{f}^{-1}(\mathrm{u})}$.

Simply write $\frac{\mathrm{df}}{\mathrm{du}} = \frac{\mathrm{dx}}{\mathrm{du}} \cdot \frac{\mathrm{df}}{\mathrm{dx}}$ when f is a function of x, which is, in turn, a function of u and write $\frac{\mathrm{dx}}{\mathrm{du}} = \left(\frac{\mathrm{du}}{\mathrm{dx}}\right)^{-1}$, in general.

For example,

$$\frac{d}{dx}\|Bx\|_W^2 = \frac{d}{dx}(Bx) \cdot (W + W^t)Bx = B^t(W + W^t)Bx\,, \tag{A-26}$$

and

$$\frac{d}{dx}\|Bx + a\|_W^2 = \frac{d}{dx}(Bx) \cdot (W + W^t)(Bx + a) = B^t(W + W^t)(Bx + a). \tag{A-27}$$

INTENTIONALLY LEFT BLANK.

Appendix B. Multivariate Normal Distribution and Quadratic Forms

B.1 Normal Distribution

The vector

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{B-1}$$

has the multivariate normal distribution

$$X \sim N_n(M, V)\ , \tag{B-2}$$

with mean vector

$$M = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix} \tag{B-3}$$

and covariance matrix

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nn} \end{bmatrix} \tag{B-4}$$

when each x_i is normally distributed with mean m_i and variance v_{ii} so that $x_i \sim N(m_i, v_{ii})$ and the x_i are related by $\mathrm{Cov}(x_i, x_j) = v_{ij}$. So, $\mathrm{E}\,X = M$ and $\mathrm{Var}\,X = \mathrm{E}[(X - M)(X - M)^t] = V$.

The probability density function (pdf) of X is $f(x)$, where

$$(2\pi)^{n/2}|V|^{1/2}f(x) = \exp\left[-\frac{1}{2}\ (x - M)^t V^{-1}(x - M)\right] = \exp\left[-\frac{1}{2}\ \|x - M\|^2_{V^{-1}}\right]. \tag{B-5}$$

In particular, the x_i are independent if V is a diagonal matrix. If $V = vI$ where I is the identity matrix, then all x_i have the same variance v. If $V = I$, then all x_i have unit variance. If, additionally, $M = 0$, then the x_i are independent standard normal $N(0,1)$.

For any n-vector A,

$$A = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}. \tag{B-6}$$

X has the translation property

$$X + A \sim N_n(M + A, V)\ . \tag{B-7}$$

For any linear transformation $K: \mathbb{R}^n \to \mathbb{R}^r$,

$$K = \begin{bmatrix} k_{11} & \cdots & k_{1n} \\ \vdots & \ddots & \vdots \\ k_{r1} & \cdots & k_{rn} \end{bmatrix}, \tag{B-8}$$

the distribution of KX is

$$KX \sim N_r(KM, KVK^t). \tag{B-9}$$

Useful conditions are that V is nonsingular (positive-definite) and that $r \leq n$ and rank $K = r$, so that KVK^t is also nonsingular (positive-definite).

B.2 The Quadratic Form

Associated with each nondegenerate multivariate normal X is a standard quadratic form which has a central χ^2 distribution. Any symmetric positive-definite square matrix V has a "square root" U which can be obtained through Choleski, singular value, or spectral decomposition where $V = UU^t$. Then the inverse $W = V^{-1}$ can be written as $W = T^tT$ where $T = U^{-1}$. Now if $X \sim N_n(M, V)$, then $X - M \sim N_n(0, V)$ and $Z = T(X - M) \sim N_n(0, TVT^t)$. Since $TVT^t = TUU^tT^t = I$, it follows that $Z \sim N_n(0, I)$, and the elements of Z are iid $N(0,1)$, so

$$\sum_{i=1}^{n} z_i^2 = Z^tZ \sim \chi_n^2, \tag{B-10}$$

which is chi-squared with n degrees of freedom. In terms of matrices,

$$\begin{aligned} Z^tZ &= \big(T(X - M)\big)^t T(X - M) = (X - M)^t T^t T\ (X - M) \\ &= (X - M)^t W\ (X - M) = (X - M)V^{-1}(X - M) \end{aligned} \tag{B-11}$$

is the quadratic form associated with $X \sim N_n(M, V)$. It has the chi-squared distribution

$$Z^tZ = (X - M)^t V^{-1}(X - M) \sim \chi_n^2 \tag{B-12}$$

with n degrees of freedom.

There is also a noncentral quadratic form associated with X. Suppose now that

$$Y = TX \sim N_n(TM, I) \tag{B-13}$$

and let

$$TM = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix}. \tag{B-14}$$

The y_i are independent normal $N(d_i, 1)$, and

$$\sum_{i=1}^{n} y_i^2 \sim \chi^2_{n,\delta} \,. \tag{B-15}$$

This is a noncentral chi-squared distribution with noncentrality parameter

$$\delta = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} \mathrm{E}[y_i]^2. \tag{B-16}$$

In the literature, the noncentrality parameter is variously considered to be δ or $\delta/2$ or $\sqrt{\delta}$.

Anyway,

$$\sum_{i=1}^{n} y_i^2 = Y^tY = (TX)^tTX = X^tT^tT\,X = X^tWX = X^tV^{-1}X. \tag{B-17}$$

Furthermore,

$$\delta = \sum_{i=1}^{n} d_i^2 = (TM)^tTM = M^tT^tTM = M^tV^{-1}M. \tag{B-18}$$

So the noncentral quadratic form associated with $X \sim N_n(M, V)$ is

$$Y^tY = X^tV^{-1}X \sim \chi^2_{n,M^tV^{-1}M}, \tag{B-19}$$

which has the noncentral chi-square distribution with n degrees of freedom and noncentrality parameter $\delta = M^tV^{-1}M$.

It is possible to combine transformations and quadratic forms. Start with

$$X \sim N_n(M, V). \tag{B-20}$$

The transformed centered vector is

$$K(X - M) \sim N_r(0, KVK^t) \,, \tag{B-21}$$

and its quadratic form is

$$[K(X - M)]^t (KVK^t)^{-1} [K(X - M)] \sim \chi_r^2. \tag{B-22}$$

The transformed (uncentered) vector is

$$KX \sim N_r(KM, KVK^t) \,, \tag{B-23}$$

and its quadratic form is

$$(KX)^t (KVK^t)^{-1} KX \sim \chi^2_{r,(KM)^t (KVK^t)^{-1} KM} \,, \tag{B-24}$$

which is noncentral chi-square with r degreed of freedom and noncentrality parameter

$$\delta = (KM)^t (KVK^t)^{-1} KM. \tag{B-25}$$

Appendix C. Maximum Likelihood Estimator Distributions

In general, the joint density (likelihood) L of a sample $\{X_i\}_{i=1}^{n}$ depends on a k-dimensional parameter $\psi = (\psi_1, \dots, \psi_k)$, as

$$L(X_1, \dots, X_n; \psi). \tag{C-1}$$

The log likelihood is denoted $\mathcal{L} = \log L$. A maximum likelihood estimator $\hat{\psi}$ of ψ satisfies

$$\hat{\psi} = \arg\max L(\psi) = \arg\max \mathcal{L}(\psi). \tag{C-2}$$

The $k \times k$ Fisher Information Matrix for the entire sample

$$M_\psi = \mathrm{E}\left[\left(\frac{d\mathcal{L}}{d\psi}\right) \cdot \left(\frac{d\mathcal{L}}{d\psi}\right)^t\right] \tag{C-3}$$

has inverse

$$V_\psi = M_\psi^{-1}. \tag{C-4}$$

The asymptotic distribution of the MLE is

$$\begin{gathered} \sqrt{n}\left(\hat{\psi} - \psi\right) \to N_k\left(0, nV_\psi\right) \quad \text{as} \quad n \to \infty \\ \hat{\psi} \sim N_k\left(\psi, V_\psi\right). \end{gathered} \tag{C-5}$$

If the samples are independent, then the likelihood is

$$L = \prod_{i=1}^{n} f_i(X_i)\,. \tag{C-6}$$

Its logarithm is

$$\mathcal{L} = \log L = \sum_{i=1}^{n} \log f_i(X_i)\,, \tag{C-7}$$

and the derivative is

$$\frac{d\mathcal{L}}{d\psi} = \sum_{i=1}^{n} \frac{d}{d\psi} \log f_i(X_i)\,. \tag{C-8}$$

The outer product is

$$\left(\frac{d\mathcal{L}}{d\psi}\right) \cdot \left(\frac{d\mathcal{L}}{d\psi}\right)^t = \sum_{i=1}^{n} \left(\frac{d}{d\psi} log\, f_i(X_i)\right) \cdot \left(\frac{d}{d\psi} log\, f_i(X_i)\right)^t$$

$$+ \sum_{i \neq j} \left(\frac{d}{d\psi} log\, f_i(X_i)\right) \cdot \left(\frac{d}{d\psi} log\, f_j(X_j)\right)^t. \tag{C-9}$$

As the samples are independent, the information matrix M_L is

$$\mathrm{E}\left[\left(\frac{d\mathcal{L}}{d\psi}\right) \cdot \left(\frac{d\mathcal{L}}{d\psi}\right)^t\right] = \sum_{i=1}^{n} \mathrm{E}\left[\left(\frac{d}{d\psi} \log f_i(X_i)\right) \cdot \left(\frac{d}{d\psi} \log f_i(X_i)\right)^t\right]$$

$$+ \sum_{i \neq j} \mathrm{E}\left[\left(\frac{d}{d\psi} \log f_i(X_i)\right)\right] \cdot \mathrm{E}\left[\left(\frac{d}{d\psi} \log f_j(X_j)\right)^t\right], \tag{C-10}$$

and since

$$\mathrm{E}\left[\frac{d}{d\psi} \log f\right] = \int \frac{1}{f} \frac{df}{d\psi}\, f = \int \frac{df}{d\psi} = \frac{d}{d\psi} \int f = \frac{d}{d\psi} 1 = 0, \tag{C-11}$$

the information for the entire sample is

$$M_\psi = \sum_{i=1}^{n} M_i\,, \tag{C-12}$$

where M_i is the information matrix for a single observation

$$M_i = \mathrm{E}\left[\left(\frac{d}{d\psi} \log f_i(X_i)\right) \cdot \left(\frac{d}{d\psi} \log f_i(X_i)\right)^t\right]. \tag{C-13}$$

Furthermore, if the samples are identically distributed, then $M_\psi = nM_o$, where

$$M_o = \mathrm{E}\left[\left(\frac{d}{d\psi}\log f(X_o)\right)\cdot\left(\frac{d}{d\psi}\log f(X_o)\right)^t\right]. \tag{C-14}$$

In this case, with $V_o = M_o^{-1}$, the variance is $V_\psi = M_\psi^{-1} = n^{-1}M_o^{-1} = n^{-1}V_o$, and for an iid sample, the result is the usual

$$\begin{gathered}\sqrt{n}\big(\hat{\psi}-\psi\big) \to N_k(0, V_o) \quad as \quad n \to \infty \\ \hat{\psi} \sim N_k(\psi, n^{-1}V_o).\end{gathered} \tag{C-15}$$

Now consider another parameterization ϕ. The chain rule gives

$$\frac{d\mathcal{L}}{d\psi} = \frac{d\phi}{d\psi}\frac{d\mathcal{L}}{d\phi}, \tag{C-16}$$

so the information matrix transformation is

$$\begin{aligned} M_\psi &= \mathrm{E}\left[\left(\frac{d\phi}{d\psi}\frac{d\mathcal{L}}{d\phi}\right)\cdot\left(\frac{d\phi}{d\psi}\frac{d\mathcal{L}}{d\phi}\right)^t\right] \\ &= \frac{d\phi}{d\psi}\mathrm{E}\left[\left(\frac{d\mathcal{L}}{d\phi}\right)\cdot\left(\frac{d\mathcal{L}}{d\phi}\right)^t\right]\frac{d\phi}{d\psi}^t \\ &= \frac{d\phi}{d\psi}M_\phi\frac{d\phi}{d\psi}^t\ . \end{aligned} \tag{C-17}$$

The chain rule also gives the inverse of a derivative matrix as the derivative of the inverse function, so

$$M_\psi{}^{-1} = \left(\frac{d\phi}{d\psi}^t\right)^{-1} M_\phi{}^{-1}\left(\frac{d\phi}{d\psi}\right)^{-1}, \tag{C-18}$$

and the corresponding variance transformation is

$$V_\psi = \frac{d\psi}{d\phi}^t V_\phi \frac{d\psi}{d\phi}. \tag{C-19}$$

INTENTIONALLY LEFT BLANK.

Appendix D. The Linear Model

D.1 The Basic Model

The usual linear model is

$$Y = X\beta + \epsilon \,, \tag{D-1}$$

where Y is an $n \times 1$ response, X is an $n \times p$ independent matrix, β is a $p \times 1$ parameter, and the $n \times 1$ error $\epsilon \sim N(0, \sigma^2 I_n)$. So E $Y = X\beta$ and Var $Y = \sigma^2 I_n$. Each column of X is a linear predictor, and the model is

$$y_i = \sum_{i=1}^{p} X_{ij}\beta_j. \tag{D-2}$$

One works with the p-parameter model for a single predictor v by choosing a set of fixed basis functions $\{f_1, \dots, f_p\}$ and setting $X_{ij} = f_j(v_i)$.

$$y_i = \sum_{i=1}^{p} \beta_j f_j(v_i). \tag{D-3}$$

For example, choice of $f_j(v) = v^{j-i}$ gives the polynomial model

$$y = \beta_0 + \beta_1 v + \beta_2 v^2 + \cdots + \beta_{p-1} v^{p-1}. \tag{D-4}$$

(See equation B-5 in appendix B.) Solution by least squares is equivalent to maximum likelihood for normal error, and the criterion is to choose u that minimizes $Q = \epsilon^t \epsilon = \|\epsilon\|^2$ $= \|Y - X\beta\|^2$ since $\epsilon = Y - X\beta$. This is

$$Q = \|Y\|^2 - 2\beta^2 X^t Y + \|X\beta\|^2. \tag{D-5}$$

The solution follows from setting the derivative to 0,

$$\frac{dQ}{d\beta} = -2X^t Y + 2X^t X\beta = 0 \,, \tag{D-6}$$

to obtain the normal equations

$$X^t X\beta = X^t Y \tag{D-7}$$

with solution

$$\hat{\beta} = (X^t X)^{-1} X^t Y \tag{D-8}$$

and response estimate

$$\hat{Y} = HY, \tag{D-9}$$

where the so-called hat matrix is

$$H = X\,(X^t X)^{-1} X^t. \tag{D-10}$$

Note that E $\hat{\beta} = \beta$ and Var $\hat{\beta} = \sigma^2 (X^t X)^{-1}$.

Modern software for linear least-squares estimation operates on equation D-7 through the response vector Y and the design matrix X. The normal equations are solved efficiently without inverting the design matrix, and software provides parameter estimates and diagnostics such as the parameter variance and hat matrix diagonal.

D.2 The Weighted Model

When the error is $N(0, \Sigma)$, the correct inner product is weighted by the symmetric $W = \Sigma^{-1}$, so $Q = \epsilon^t W \epsilon = \|\epsilon\|_W^2 = \|Y - X\beta\|_W^2$. This is

$$Q = \|Y\|_W^2 - 2\beta^t X^t W Y + \|X\beta\|_W^2. \tag{D-11}$$

Then

$$\frac{dQ}{d\beta} = -2X^t W Y + 2X^t W X\beta = 0. \tag{D-12}$$

The normal equation is

$$X^t W X\beta = X^t W Y. \tag{D-13}$$

The solution is

$$\hat{\beta} = (X^t W X)^{-1} X^t W Y, \tag{D-14}$$

and the response estimate is $\hat{Y} = HY$, where

$$H = X(X^t W X)^{-1} X^t W. \tag{D-15}$$

Note that $\mathrm{E}\,\hat{\beta} = \beta$ and $\mathrm{Var}\,\hat{\beta} = (X^t W X)^{-1}$.

Modern software for linear least-squares estimation operates on equation D-13 through the response vector Y, the weight vector W, and the design matrix X. The normal equations are solved efficiently without inverting the design matrix, and software provides parameter estimates and diagnostics such as the parameter variance and hat matrix diagonal.

Note that for these models, the likelihood function is

$$L(\beta) = (2\pi)^{-n/2} |\mathrm{W}|^{1/2} \exp\left[-\frac{1}{2}\ \|Y - X\beta\|_W^2\right]. \tag{D-16}$$

Its log derivative is

$$\frac{d}{d\beta} \log L(\beta) = X^t W (Y - X\beta), \tag{D-17}$$

and the information matrix is

$$M_\beta = X^t W X. \tag{D-18}$$

INTENTIONALLY LEFT BLANK.

Appendix E. The Generalized Linear Model

E.1 Model Formulation

In the Generalized Linear Model (GLM), the response Y has an arbitrary distribution, $\eta = X\beta$ is a linear function of the k-dimensional parameter β, and the mean response is modeled as

$$\mu = \mathrm{E}[Y \mid X] = G(\eta) \tag{E-1}$$

for some monotone link function G with derivative $g = G'$. (Some authors call G^{-1} the link.)

Response distributions are taken to be from a single-parameter exponential family, with the form

$$f(y, \theta, \psi) = \exp\left[\frac{y\theta - b(\theta)}{a(\psi)} + c(y, \psi)\right]. \tag{E-2}$$

The parameterθ is to be estimated, and ψ is a nuisance parameter.

With $\ell = \log f$, calculate the moments of Y in terms of exponential family components.

$$\mathrm{E}\left[\frac{d}{d\theta}\ell(y, \theta, \psi)\right] = 0 \tag{E-3}$$

because $\int f(y, \theta, \psi)dy = 1$, and under suitable regularity conditions, $0 = \frac{d}{d\theta}\int f(y, \theta, \psi)dy$ $= \int \frac{d}{d\theta} f(y, \theta, \psi)dy = \int \frac{d}{d\theta}\ell(y, \theta, \psi) \cdot f(y, \theta, \psi)dy$. Therefore, $\mathrm{E}[(y - b'(\theta))/a(\psi)] = 0$, and

$$\mathrm{E}[Y] = \mu = G(\eta) = b'(\theta). \tag{E-4}$$

Also, as usual,

$$\mathrm{E}\left[\left(\frac{d}{d\theta}\ell(y, \theta, \psi)\right)^2\right] = -\,\mathrm{E}\left[\frac{d^2}{d\theta^2}\ell(y, \theta, \psi)\right]$$

because $\mathrm{Var}\left[\frac{d}{d\theta}\ell\right] = \mathrm{E}\left[\left(\frac{d}{d\theta}\ell\right)^2\right] = \int \left(\frac{d}{d\theta}\ell\right)^2 \cdot f\, dy = \int \frac{d}{d\theta}\ell \cdot \frac{d}{d\theta} f\, dy = \frac{d}{d\theta}\ell \cdot f\Big|_{-\infty}^{\infty}$
$$- \int \frac{d^2}{d\theta^2}\ell \cdot f\, dy = -\,\mathrm{E}\left[\frac{d^2}{d\theta^2}\ell\right]. \tag{E-5}$$

Therefore, $\mathrm{E}[(y - \mu)^2/a(\psi)^2] = -\,\mathrm{E}[-b''(\theta)/a(\psi)]$, and

$$\operatorname{Var} Y = v(\mu)a(\psi) = b''(\theta)a(\psi), \tag{E-6}$$

where $v(\mu) = \operatorname{Var}[Y]/a(\psi) = b''(\theta) = g(\eta)\frac{d\eta}{d\theta}$.

In the case that $\eta = \theta$, and hence $\mu = G(\eta) = G(\theta)$, G is called the canonical link function. Then $\mu = b'(\theta) = G(\theta) = G(\eta)$ and $v(\mu) = b''(\theta) = g(\theta) = g(\eta)$.

E.2 Estimation

Let $[x_1 \quad \cdots \quad x_n] = X^t$, so the (column) vector x_i is row i of X, $\eta_i = x_i^t\beta$, and $\theta_i = \theta(\eta_i)$.

Maximum likelihood estimation for the GLM is accomplished by maximizing the log likelihood function

$$\mathcal{L} = \sum_{i=1}^{n} \ell(y_i, \theta_i, \psi) = \sum_{i=1}^{n} \left[\frac{y_i\theta_i - b(\theta_i)}{a(\psi)} + c(y_i, \psi)\right]. \tag{E-7}$$

This is a weighted least-squared problem where the design and weight depend on the unknown parameter, and it can be solved iteratively by the Newton-Raphson method.

The Newton-Raphson Method

In one dimension, a zero of F is obtained by linearizing and updating the current argument x_o to x, solving $F(x_o) + (x - x_o)F'(x_o) = 0$ to get $x = x_o - F(x_o)/F'(x_o)$. Optimize F by setting $F'' = 0$, so the update is $x = x_o - F'(x_o)/F''(x_o)$.

The vector version is $x = x_o - \left[\frac{d^2}{dxx^t}F(x_o)\right]^{-1}\frac{d}{dx}F(x_o)$. Taking $x = x_o + \delta$, the increment $\delta = x - x_o$ satisfies $\left[\frac{d^2}{dxx^t}F(x_o)\right]\delta = -\frac{d}{dx}F(x_o)$. Some derivatives are required.

Gradient

Differentiating gives the gradient (vector of first derivatives)

$$\mathcal{D}(\beta) = \frac{d\mathcal{L}}{d\beta} = a(\psi)^{-1} \cdot \sum_{i=1}^{n} [y_i - b'(\theta_i)] \cdot \frac{d\theta_i}{d\beta}. \tag{E-8}$$

Since $\frac{d}{d\beta}b'(\theta_i) = b''(\theta_i)\frac{d\theta_i}{d\beta} = v(\mu_i)\frac{d\theta_i}{d\beta}$ and $\frac{d}{d\beta}b'(\theta_i) = \frac{d\mu_i}{d\beta} = \frac{d}{d\beta}G(\eta_i) = \frac{d}{d\beta}G(x_i^t\beta)$ $= g(\eta_i)x_i$, then

$$\frac{d\theta_i}{d\beta} = \frac{g(\eta_i)}{v(\mu_i)} x_i. \tag{E-9}$$

So the gradient is

$$\mathcal{D}(\beta) = a(\psi)^{-1} \cdot \sum_{i=1}^{n} (y_i - \mu_i) \cdot \frac{g(\eta_i)}{v(\mu_i)} x_i = a(\psi)^{-1} \cdot X^t W_F U_F, \tag{E-10}$$

where, for $i = 1, \dots, n$, the diagonal weight matrix $W_F$has elements $w_{Fi} = g(\eta_i)\hat{}2/v(\mu_i)$,

$$W_F = \text{diag}\left[\frac{g(\eta_1)^2}{v(\mu_1)} \quad \cdots \quad \frac{g(\eta_n)^2}{v(\mu_n)}\right], \tag{E-11}$$

and the centered/scaled response vector U_F has elements $u_{Fi} = (y_i - \mu_i)/g(\eta_i)$,

$$U_F = \left[\frac{y_1 - \mu_1}{g(\eta_1)} \quad \cdots \quad \frac{y_n - \mu_n}{g(\eta_n)}\right]. \tag{E-12}$$

Hessian

Using $\frac{d}{d\beta} v(\mu_i) = v'(\mu_i) \frac{d}{d\beta} G(x_i^t \beta) = v'(\mu_i) g\left(x_{i\beta}^t\right) x_i$,

$$\frac{d^2}{d\beta\beta^t} \theta_i = \frac{g'(\eta_i) v(\mu_i) - g(\eta_i)^2 v'(\mu_i)}{v(\mu_i)^2} x_i x_i^t, \tag{E-13}$$

and the Hessian (matrix of second derivatives) is

$$\begin{aligned}\mathcal{H}(\beta) &= \frac{d^2}{d\beta\beta^t} \mathcal{L} = a(\psi)^{-1} \cdot \sum_{i=1}^{n} \left[-b''(\theta_i) \frac{d\theta_i}{d\beta} \frac{d\theta_i^t}{d\beta} + (y_i - \mu_i) \frac{d^2}{d\beta\beta^t} \theta_i\right] \\ &= a(\psi)^{-1} \cdot \sum_{i=1}^{n} \left[-\frac{g(\eta_i)^2}{v(\mu_i)} + (y_i - \mu_i) \frac{g'(\eta_i) v(\mu_i) - g(\eta_i)^2 v'(\mu_i)}{v(\mu_i)^2}\right] x_i x_i^t \\ &= a(\psi)^{-1} \cdot X^t W_N X,\end{aligned} \tag{E-14}$$

where

$$W_N = W_F - W_D\,, \tag{E-15}$$

and W_D is a diagonal matrix with diagonal elements

$$w_{Di} = (y_i - \mu_i)\frac{g'(\eta_i)v(\mu_i) - g(\eta_i)^2 v'(\mu_i)}{v(\mu_i)^2}. \tag{E-16}$$

Since $\mathrm{E}[W_D] = 0$, the expected value of the Hessian is

$$\mathrm{E}\,\mathcal{H}(\beta) = -a(\psi)^{-1} \cdot X^t W_F X. \tag{E-17}$$

The Fisher Information Matrix is

$$M_\beta = \mathrm{E}\left[\frac{d}{d\beta}\mathcal{L} \cdot \frac{d}{d\beta}\mathcal{L}^t\right] = -\,\mathrm{E}\left[\frac{d^2}{d\beta\beta^t}\mathcal{L}\right] = -\,\mathrm{E}\,\mathcal{H}(\beta) = -a(\psi)^{-1} \cdot X^t W_F X, \tag{E-18}$$

and the asymptotic estimator distribution is $N_k\,(\beta, a(\psi) \cdot (X^t W_F X\,)^{-1})$.

Newton-Raphson

Now, apply the Newton-Raphson algorithm iteratively to solve the optimization.

For GLM, the Newton-Raphson update is $\beta = \beta_o + \delta$, where

$$\begin{aligned} \mathcal{H}(\beta_o) &= -\mathcal{D}(\beta_o) \\ (X^t W_N X)\delta &= X^t W_F U_F \\ (X^t W_N X)\delta &= X^t W_N W_N^{-1} W_F U_F \\ (X^t W_N X)\delta &= X^t W_N U_N \end{aligned} \tag{E-19}$$

with

$$U_N = W_N^{-1} W_F U_F. \tag{E-20}$$

These are the normal equations for minimization of $Q = \|U_N - X\delta\|_{W_N}^2$. Both W_N and U_N depend on β_o. The normal equations can be solved iteratively with an initial guess β_o by calculating $\eta = X\beta_o$, $\mu = G(\eta)$, g, g', v, v', W_F, U_F, W_D, W_N, and U_N. Then solve for δ. The updated solution is $\beta = \beta_o + \delta$. Now replace β_o with β, and repeat. This is iteratively reweighted least squares with Newton-Raphson update.

Fisher Scoring

For the GLM, the Fisher Scoring update uses $\mathrm{E}\,\mathcal{H}$ in place of $\mathcal{H}$ to get

$$\mathrm{E}\,\mathcal{H}(\beta_o)\delta = -\mathcal{D}(\beta_o) \text{ and } (X^t W_F X)\delta = X^t W_F U_F, \tag{E-21}$$

which are the normal equations for $Q = \|U_F - X\delta\|_{W_F}^2$. Both W_F and U_F depend on β_o. The normal equations can be solved iteratively with an initial guess β_o by calculating $\eta = X\beta_o$, $\mu = G(\eta)$, g, v, W_F, and U_F. Then solve for δ, update, and repeat. This is iteratively reweighted least squares with Fisher scoring.

E.3 Canonical Link

For the canonical link, $w_{Fi} = g(\eta_i) = v(\mu_i)$ and $u_{Fi} = (y_i - \mu_i)/w_{Fi}$. Also, since $d\theta_i/d\beta = x_i$ and $d^2\theta_i\,/d\beta\beta^t = 0$, it follows that $\mathrm{E}\,\mathcal{H}(\beta) = \mathcal{H}(\beta)$. So Newton-Raphson and Fisher scoring are equivalent.

E.4 Confidence Intervals

Normal-approximation $100\gamma\%$ confidence intervals on the mean response are given by

$$G\left(x\beta \pm \Phi_{1-(1-\gamma)/2}\sqrt{x^t V x}\right), \tag{E-22}$$

where Φ is a standard normal quantile, V is the estimated parameter variance matrix, and x is a row of an X matrix corresponding to the desired level. For the basis implementation, this is

$$x = \Big(f_1(v), \dots, f_p(v)\Big). \tag{E-23}$$

E.5 Bernoulli Response

Suppose the response $Y \in \{0,1\}$ is Bernoulli with $\Pr[Y = 1] = \mu = 1 - \Pr[Y = 0]$. The Bernoulli probability is

$$f(y) = \mu^y(1-\mu)^{1-y} = \exp\left[y \log\frac{\mu}{1-\mu} + \log(1-\mu)\right], \tag{E-24}$$

so $a = 1$, $c = 0$, and there is no nuisance parameter. Furthermore, $\theta = \log\big(\mu/(1-\mu)\big)$ and $\mu = 1/\big(1 + e^{-\theta}\big)$, and so $b(\theta) = -\log(1-\mu) = \log\big(1 + e^{\theta}\big)$. Note that $\mathrm{E}\,Y = b'(\theta)$ $= e^{\theta}/\big(1 + e^{\theta}\big) = \mu$ and $\mathrm{Var}\,Y = v(\mu) = b''(\theta) = e^{-\theta}/\big(1 + e^{-\theta}\big)^2 = \mu(1-\mu)$ as expected.

With $\eta = \theta$ and $\mu = G(\theta)$, the canonical link for Bernoulli response is seen to be the logistic cdf $G(\eta) = 1/(1 + e^{-\eta}) = \mu$. Note that $g(\eta) = \mu(1-\mu) = v(\mu)$, so $w_{Fi} = \mu_i(1-\mu_i)$ and $u_{Fi} = (y_i - \mu_i)/\big(\mu_i(1-\mu_i)\big)$. The resulting model is logistic regression, or the logit model.

For an arbitrary link cdf G, take $\eta = X\beta$, $\mu = G(X\beta)$, $v(\mu) = \mu(1-\mu)$, and $g(\eta) = g(X\beta)$.

Use of the standard normal cdf $G = \Phi$ with pdf $g = \phi$ gives the probit model.

Because the likelihood function is $L = \prod \mu_i^{y_i} \cdot (1-\mu_i)^{1-y_i}$, it follows that $L_{\text{full}} = 1$ for the Bernoulli model and the deviance is $\Delta = -2\log L$.

As an example, consider the usual two-parameter model with predictor $(x_1, \dots, x_n)$ and response $(y_1, \dots, y_n)$. The increment $\delta = (d_0, d_1)$ is the solution of $M\delta = A$, where

$$M = X^t W X = \begin{bmatrix} \Sigma w_i & \Sigma w_i x_i \\ \Sigma w_i x_i & \Sigma w_i x_i^2 \end{bmatrix} \quad and \quad A = X^t W U = \begin{bmatrix} \Sigma w_i u_i \\ \Sigma w_i u_i x_i \end{bmatrix}. \tag{E-25}$$

To do the Fisher update of section E.2, calculate the linear response $\eta_i = b_0 + b_1 x_i$, mean $\mu_i = G(\eta_i)$, derivative $g_i = g(\eta_i)$, variance $v_i = \mu_i(1-\mu_i)$, transformed response $u_i = u_{Fi} = (y_i - \mu_i)/g_i$, weight $w_i = w_{Fi} = g_i^2/v_i$, and weighted transformed response $w_i u_i = w_{Fi} u_{Fi} = (y_i - \mu_i) g_i / v_i$.

For the Newton-Raphson update, $w_{Di} = (y_i - \mu_i)(g_i' v_i - g_i^2 v_i')/v_i^2$ and $w_{Ni} = w_{Fi} - w_{Di}$ and $u_i = u_{Ni} = w_{Fi}/w_{Ni} \cdot (y_i - \mu_i)/g_i$. Then $w_i = w_{Ni}$ and $w_i u_i = w_{Ni} u_{Ni} = w_{Fi} u_{Fi}$.

For the canonical link $w_i = w_{Fi} = g_i = v_i$ and $w_i u_i = w_{Fi} u_{Fi} = y_i - \mu_i$.

NO. OF COPIES	ORGANIZATION
1 (PDF only)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FORT BELVOIR VA 22060-6218
1	DIRECTOR US ARMY RESEARCH LAB IMNE ALC HRR 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB RDRL CIO LL 2800 POWDER MILL RD ADELPHI MD 20783-1197

NO. OF COPIES	ORGANIZATION
1	USARL RDRL SLE R FLORES WSMR NM 88002-5513
	ABERDEEN PROVING GROUND
1	DIR US ARMY EVALUATION CTR HQ TEAE SV P A THOMPSON 2202 ABERDEEN BLVD 2ND FL APG MD 21005-5001
3 (2 HC 1 PDF)	DIR USARL RDRL SL J BEILFUSS P TANENBAUM RDRL SLB A M PERRY (PDF only)

NO. OF COPIES	ORGANIZATION
18	DIR USARL RDRL SLB R BOWEN RDRL SLB D J COLLINS (5 CPS) J EDWARDS R GROTE T HOLDREN R KINSLER L MOSS E SNYDER RDRL SLB E M MAHAFFEY RDRL SLB G P MERGLER RDRL SLB S S SNEAD RDRL SLB W L ROACH RDRL WML A A THOMPSON D WEBB

INTENTIONALLY LEFT BLANK.