RA module 21: Structure of GLMs practice problems

(The attached PDF file has better formatting.)

Fox Regression analysis Chapter 15 Structure of Generalized linear models

** Exercise 21.1: Components of generalized linear models

A generalized linear model has three components: a linear predictor, a link function, and a random component (a conditional distribution of the response variable).

A. What is a linear predictor?
B. What is a link function?
C. What is the random component?
D. For classical regression analysis, what are these three elements?

Part A: The linear predictor is a linear function of regressors, $\eta_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj}$. η_j is a function of the fitted value, not necessarily the fitted value itself.

Part B: The link function is smooth and invertible linearizing function $g(\cdot)$, which transforms the expectation of the response variable, $\mu_j = E(Y_j)$, to the linear predictor $g(\mu_j) = \eta_j$.

Illustration: For a log-link function, if the linear predictor $\eta_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj} = 2$, the fitted value is $e^2 = 7.389$, since $ln(7.389) = 2$.

Part C: The random component specifies the conditional distribution of the response variable Y_j (for the jth of n independently sampled observations), given the values of the explanatory variables in the model.

Illustration: For Poissson GLM with a log-link function, if the linear predictor $\eta_j = 2$, Y_j has a Poisson distribution with a mean of $e^2 = 7.389$.

Part D: For classical regression, the linear predictor is the same.
The link function is the identity function: $\eta_j = \mu_j = E(Y_j)$.
The random component is a normal distribution with the same variance at every point.

Jacob: What types of link functions should we know?

Rachel: Know the log link and logit link functions.

Jacob: What types of conditional distributions should we know?

Rachel: Know the Poisson, Gamma, and binomial distributions.

Jacob: The textbook does not give formulas for solving GLMs. Do we have to solve GLMs for the final exam?

Rachel: One can't solve GLMs by pencil and paper. The final exam tests the GLM concepts in the practice problems on the discussion forum; it does not give data and ask for the GLM coefficients.

** Exercise 21.2: Fitting generalized linear models

A. How are generalized linear models fit to observed data?
B. How does this differ from classical regression analysis?

Part A: Fitting a distribution to observed values has two parts:

● Choose the distribution, such as normal, Poisson, Gamma, binomial. This is the conditional distribution of the response variable.
● Choose the parameters of the distribution to maximize the likelihood (the probability) of observing the empirical data.

Jacob: For classical regression analysis, do we choose a conditional distributions of the response variable?

Rachel: Yes, we choose a normal distribution with a constant variance.

Jacob: Are there statistical methods to choose the conditional distribution?

Rachel: We use intuition and the relation of the variance to the mean.

Intuition: For probabilities, we use a binomial distribution. For counts, we might use a Poisson distribution or a negative binomial distribution. For stock prices or claim severities, we might use a lognormal distribution or a Gamma distribution.

Part B: Classical regression analysis assumes the distribution is a normal distribution with the same variance at every point. With this assumption, maximizing the likelihood is the same as minimizing the squared error of the residuals. Ordinary least squares estimation for a normal distribution is maximum likelihood estimation.

Jacob: Do we maximize the likelihood or the loglikelihood?

Rachel: The loglikelihood is a monotonic function of the loglikelihood. If we have a points $f(x_j)$, where f is a function of x, the value x_j which maximizes $f(x_j)$ is also the value which maximizes $ln(f(x_j))$.

Jacob: Is this the same as minimizing the residual deviance?

Rachel: The residual deviance is 2 × (K − loglikelihood $(x_{1,j}, x_{2,j}, \ldots , x_{n,j})$. The set of $(x_{1,j}, x_{2,j}, \ldots , x_{n,j})$ which maximize the likelihood also maximize the loglikelihood and minimize the residual deviance.

** Exercise 21.3: Link function

An actuary uses a generalized linear model to relate the response variable Y_j to two explanatory variables (covariates), X_1 and X_2.

● Let μ_j be the expected value for the response variable at observation j.
● Let η_j be the linear predictor at observation j

The intercept of the GLM is α, and the coefficients of X_1 and X_2 are β_1 and β_2.

A. What is the linear predictor η_j?
B. For a log-link function $g()$, what is the relation between μ_j and the independent variables?
C. For a logit link function $g()$, what is the relation between μ_j and the independent variables?

Part A: The linear predictor at observation $j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j}$.

Jacob: GLMs differ from classical regression they are used for multiplicative models, probability models with dichotomous random variables, and models of skewed distributions. Yet this linear predictor is the same as the formula in classical regression analysis.

Rachel: GLMs have three parts: linear predictor, link function, and conditional distribution of the response variable. The linear predictor is the same as for classical regression analysis.

Part B: $g(\mu_j) = ln(\mu_j) = \eta_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j}$

Jacob: For classical regression analysis, do we say $Y_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j}$?

Rachel: The observed value Y_j is a random variable; it is not equal to an expression of scalars. For classical regression analysis, we write $Y_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \epsilon_j$, where ϵ_j has a normal distribution with a mean of zero and the same variance for all values of the explanatory variables. GLMs use an *identity link function* for classical regression analysis, where g(x) = x:

$g(\mu_j) = \mu_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j}$

Jacob: Why is the log-link function so often used in GLMs?

Rachel: Many relations are multiplicative models. For example, personal auto insurance premiums depend on driver characteristics (like male vs female) and territory (like urban vs rural). The insurance rates are a multiplicative model: the male rate may be twice the female rate and the urban rate may be three times the rural rate. The log-link function gives a multiplicative model:

$ln(\mu_j) = \eta_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} \Rightarrow$
$\mu_j = exp(\alpha + \beta_1 X_{1j} + \beta_2 X_{2j}) \Rightarrow$
$\mu_j = exp(\alpha) \times exp(\beta_1 X_{1j}) \times exp(\beta_2 X_{2j})$

Define new parameters:

● $\alpha' = exp(\alpha)$ = the base rate
● $\beta_1' = exp(\beta_1)$ = the male/female relativity
● $\beta_2' = exp(\beta_2)$ = the urban/rural relativity

Part C: $g(\mu_j) = ln[\ \mu_j / (1 - \mu_j)\ ] = \eta_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j}$

Jacob: What is the rationale for the logit link function? Log odds may be relevant to horse racing or Las Vegas casinos, but they have no intuitive relation to actuarial distributions.

Rachel: That is true, and the logit link function is not appropriate for all actuarial distributions. But the logit link function has the proper form; it converts a range from 0 to 1 to a range from $-\infty$ to $+\infty$.

** Exercuse 21.4: Link function

An actuary uses a generalized linear model with a log-link function to relate the response variable Y_j to two explanatory variables, X_1 and X_2. Let μ_j be the expected value for the response variable at observation j. The intercept of the GLM is α, and the coefficients of X_1 and X_2 are β_1 and β_2.

A. What is the relation of the explanatory variables to the response variable using the link function?
B. What is the relation of the explanatory variables to the response variable using the inverse of the link function?

Part A: $ln(\mu_j) = \alpha + \beta_1 X_1 + \beta_2 X_2$

Part B: $\mu_j = \exp(\alpha + \beta_1 X_1 + \beta_2 X_2)$

Jacob: Why don't we use $\quad ln(Y_j) = \alpha + \beta_1 X_1 + \beta_2 X_2 \quad$ and $\quad Y_j = \exp(\alpha + \beta_1 X_1 + \beta_2 X_2)$?

Rachel: Y is a random variable: the linear predictor adjusted by the inverse of the link function plus a random component.

** Exercise 21.5: Likelihoods

A. What is the range of a likelihood?
B. What is the range of a log-likelihood?
C. What is meant by a saturated model?
D. What is the relation of the likelihood for the saturated model vs any other model?
E. What is the relation of the log-likelihood for the saturated model vs any other model?

Part A: Suppose Y is a function of X. As an example, let Y be the Poisson probability for X: $y = \mu^x e^{-\mu}/x!$

● pdf$(x \mid \mu) = (\mu^x e^{-\mu}) / x!$
● likelihood $(\mu \mid x) = (\mu^x e^{-\mu}) / x!$

The pdf (probability density function) and the likelihood have a range of [0, 1].

Part B: The logarithm of 0 is $-\infty$ and the logarithm of 1 is 0, so the range of the loglikelihood is $(-\infty, 0]$.

Part C: A saturated model has the fitted equal to the observed value at every point. If a regression equation or a GLM has N points, the saturated model has N parameters and 0 degrees of freedom.

Part D: The likelihood is greatest when μ = the observed value. For a given x, the value of $(\mu^x e^{-\mu}) / x!$ is maximized for $\mu = x$. If L_s is the likelihood for the saturated model and L_m is the likelihood for any other model (with the same type of conditional distribution for the response variable but different parameters), then

$$0 \leq L_m \leq L_s \leq 1$$

Part E: If LL_s is the log-likelihood for the saturated model and LL_m is the log-likelihood for any other model (with the same type of conditional distribution for the response variable but different parameters), then

$$-\infty \leq LL_m \leq LL_s \leq 0$$

Note: This exercise uses $\leq$ (less than or equal to). Some exam problems use < (less than). Both relations are correct, as long as the model in question is not the saturated model.