

(Revised 6/29/2016)

Quasi-Likelihood

Quasi-likelihood refers to a method of solving for the parameters in a Generalized Linear Model when the assumption that a distribution follows the exponential family is not supported. Solving for the parameters for a model using quasi-likelihood requires assuming there is a relationship between the mean and variance, but the underlying distribution need not be known. Quasi-likelihood equations start with a variation on the derivative of the likelihood equations, which leads to the name quasi-likelihood.

An example of where quasi-likelihood would be used is the case of over dispersion. Over dispersion, where the variance is larger than the mean, for Poisson regression is mentioned in sections 3.2 and 9.8 of the Dobson & Barnett textbook. Over dispersion is also mentioned in the case of the Binomial distribution where the variance observed is larger than the variance calculated using the Binomial distribution formula as described in section 7.7. Using the Poisson distribution requires assuming that the variance equals the mean, but the results of the modeling exercise may indicate that the mean is materially different than the variance. Through the quasi-likelihood approach, one can solve for a constant of proportionality between the mean and variance at the same time one solves for the betas in the Poisson regression. A more general case where the quasi-likelihood method is appropriate is where there is some uncertainty about the underlying distribution for the dependent variable, but assuming a constant of proportionality between the mean and variance is warranted.

Regression Through the Origin

Regression through the origin occurs when the intercept term in the linear equation linking the explanatory variables to the dependent variable is zero or is left out of the equation. Normally, the intercept is included by default in the linear equation linking the dependent variable to the explanatory variables, but the test of significance for the intercept term could indicate that the value is not significantly different than zero and the modeler could choose to drop the intercept from the equation. Leaving the intercept out of the equation can lead to non-intuitive results though. The easiest case in which to visualize this is for a general linear model, where setting the intercept to zero implies that if all of the values for explanatory variables are zero the dependent variable will be zero which may not yield sensible results.

QQ and Box Plots to test distribution assumptions

In several cases, the Dobson text refers to Normal probability plot, which is used to test Normality of the residuals. This is an example of a QQ plot, which is built using order statistic

concepts. The algorithm used to build a QQ plot (quantile-quantile) is described in Section 4.4.1 of Hogg, McKean and Craig 7th Edition.

Box plots are also described in section 4.4.1 and make use of order statistics to describe how the observations behave as an aid to evaluating how well a model fits the data at hand.

Tweedie Distributions

The Tweedie Distributions refers to a family of distributions where the variance is related to the mean through a power function $\text{Var}[Y] = a[E(Y)]^p$.

One can characterize members of the exponential family of distributions using this formula:

- Normal distribution, $p = 0$
- Poisson distribution, $p = 1$
- Compound Poisson–gamma distribution, $1 < p < 2$
- Gamma distribution, $p = 2$
- Inverse Gaussian distribution, $p = 3$

The Tweedie Distribution is particularly useful in the case where one has a data set that combines observations with zero for the dependent variable with continuous non-zero value with the result that one can view the distribution as being a Compound Poisson-Gamma Distribution. In some software packages, one can specify the Tweedie distribution, which will in turn cause the software package to find the best value for the power, p , parameter when solving for the parameters (betas) in the linear equation.

The formula above is consistent with the general feature for the Exponential Distribution family that the variance is a function of the mean.

Canonical Link Function

The inverse of the link function transforms the linear formula for the mean result to the original data set scale. The canonical form of the link function varies by the distribution selected to model the explanatory variable. If one uses the canonical link function, the estimate from the GLM is unbiased. For example, if one uses the log link function in conjunction with the Poisson distribution, the estimates from that GLM are unbiased estimates.

The canonical form of the link function can be derived for each member of the Exponential Family of distributions using the general formula and the goal that the estimates of the expected value from the GLM should be unbiased. While one can use a non-canonical link function, the result from the GLM will not be an unbiased estimate of the mean. Some common canonical link functions are shown below.

Distribution	Canonical Link Function
Normal	Identity
Binomial	Logit
Poisson	Natural Logarithm
Gamma	Inverse

One should note that the table above is not exhaustive and the candidate should work through the problems at the end of the relevant chapters in the textbooks to have a better understanding of the link function for each distribution as well as the relationship between the mean and variance for members of the Exponential Distribution Family.

Exponential Family Formulas

The information below is included to supplement the table shown in Section 3.2 of Dobson & Barnett and may be useful to the candidate when working through the questions at the end of Chapter 3 in Dobson & Barnett.

A random variable Y whose probability distribution depends on a single parameter θ (all other parameters are treated as known) is a member of the exponential family if it can be written in the following form:

$$f(y; \theta) = \exp [a(y)b(\theta) + c(\theta) + d(y)]$$

Distribution	θ	$a(y)$	$b(\theta)$	$c(\theta)$	$d(y)$
Binomial	q	y	$\log\left(\frac{q}{1-q}\right)$	$m \log(1-q)$	$\log \binom{m}{y}$
Normal	μ	y	$\frac{\mu}{\sigma^2}$	$-\frac{\mu^2}{2\sigma^2}$	$-\frac{y^2}{2\sigma^2} - \frac{1}{2} \log 2\pi\sigma^2$
Poisson	λ	y	$\log \lambda$	$-\lambda$	$-\log y!$
Exponential	θ	y	$-\frac{1}{\theta}$	$-\log \theta$	0
Gamma	θ	y	$-\frac{1}{\theta}$	$-\alpha \log \theta$	$(\alpha - 1) \log y - \log \Gamma(\alpha)$
Inverse Gaussian	μ	y	$-\frac{\theta}{2\mu^2}$	$\frac{\theta}{\mu}$	$-\frac{\theta}{2y} + \frac{1}{2} \log \frac{\theta}{2\pi y^3}$
Negative Binomial	β	y	$\log\left(\frac{\beta}{1+\beta}\right)$	$-r \log(1+\beta)$	$\log \binom{r+y-1}{r-1}$