A Primer on the Exponential Family of Distributions

David R. Clark, FCAS, MAAA, and Charles A. Thayer

A PRIMER ON THE EXPONENTIAL FAMILY OF DISTRIBUTIONS

David R. Clark and Charles A. Thayer

2004 Call Paper Program on Generalized Linear Models

Abstract

Generalized Linear Model (GLM) theory represents a significant advance beyond linear regression theory, specifically in expanding the choice of probability distributions from the Normal to the Natural Exponential Family. This Primer is intended for GLM users seeking a handy reference on the model's distributional assumptions. The Exponential Family of Distributions is introduced, with an emphasis on variance structures that may be suitable for aggregate loss models in property casualty insurance.

A PRIMER ON THE EXPONENTIAL FAMILY OF DISTRIBUTIONS

INTRODUCTION

Generalized Linear Model (GLM) theory is a significant advance beyond linear regression theory. A major part of this advance comes from allowing a broader family of distributions to be used for the error term, rather than just the Normal (Gaussian) distribution as required in linear regression.

More specifically, GLM allows the user to select a distribution from the *Exponential Family*, which gives much greater flexibility in specifying the variance structure of the variable being forecast (the "response variable"). For insurance applications, this is a big step towards more realistic modeling of loss distributions, while preserving the advantages of regression theory such as the ability to calculate standard errors for estimated parameters. The Exponential family also includes several discrete distributions that are attractive candidates for modeling claim counts and other events, but such models will not be considered here

The purpose of this Primer is to give the practicing actuary a basic introduction to the Exponential Family of distributions, so that GLM models can be designed to best approximate the behavior of the insurance phenomenon.

Insurance Applications

Two major application areas of GLM have emerged in property and casualty insurance. The first is classification ratemaking, which is very clearly illustrated in the papers by Zehnwirth and Mildenhall. The second is in loss reserving, also given an excellent treatment in papers by England & Verrall. In 1991, Mack pointed out a connection between these two applications, so it is not surprising that a common modeling framework works in both contexts.

Both classification ratemaking and reserving seek to find the "best" fitted values μ_i to the observed values y_i. In both cases the <u>response variable</u>, Y_i, of which the observed values y_i are realizations, is measured in units of aggregate loss dollars The response is dependent on predictor variables called <u>covariates</u>. Following Mack, classification ratemaking is performed using at least two covariates, which might include territory and driver age. In the reserving application, the covariates might include accident year and development year.

For our discussions, the choice of covariates used as predictors will not be important, but it will always be assumed that the response variable Y represents aggregate loss dollars. Some of the desirable qualities of the distribution for Y, driven by this assumption, are:

- The distribution is unbiased, or "balanced" with the observed values
- It allows zero values in the response with non-zero probability
- It is positively skewed

Before seeing how specific distributions in the Exponential Family measure up to these desirable qualities, some basic definitions are needed.

DEFINING THE EXPONENTIAL FAMILY

The General and Natural Forms

The general exponential family includes all distributions, whether continuous, discrete or of mixed type, whose probability function or density can be written as follows:

<u>General Form</u> (ignoring parameters other than θ_i):

$$f(y_i;\theta_i) = \exp[d(\theta_i)\ e(y_i) + g(\theta_i) + h(y_i)]$$

where d, e, g, h are all known functions that have the same form for all y_i.

For GLM, we make use of a special subclass called the Natural Exponential Family, for which $d(\theta_i)=\theta_i$ and $e(y_i)=y_i$. Following McCullagh & Nelder, the "natural form" for this family includes an additional dispersion parameter ϕ that is constant for all y_i

Natural Form:

$$f(y_i;\theta_i,\phi)=\exp[\{\theta_i\ y_i-b(\theta_i)\}/a(\phi)+c(y_i,\phi)]$$

where a, b, c are all known functions that have the same form for all y_i.

For each form, θ_i is called the canonical parameter for Y_i.

Appendix A shows how the moments are derived for the Natural Exponential Family.

The natural form can also be written in terms of the mean μ_i rather than θ_i by means of a simple transformation: $\mu_i=\tau(\theta_i)=E[y_i;\theta_i]$. This mean value parameterization of the density function, in which μ_i is an explicit parameter, will be the form used in the rest of the paper and the Appendices.

Mean Value Natural Form:

$$f(y_i;\mu_i,\phi)=\exp[\{\tau^{-1}(\mu_i)\ y_i-b(\tau^{-1}(\mu_i))\}/a(\phi)+c(y_i,\phi)]$$

To put this in context, a GLM setup based on Y consists of a linear component, which resembles a linear model with several independent variables, and a link function that relates the linear part to a function of the expected value μ_i of Y_i, rather than to μ_i itself. In the GLM, the variables are called covariates, or factors if they refer to qualitative categories. The function $\theta=\tau^{-1}(\mu)$ used in the mean value form is called the canonical link function for a GLM setup based on Y, because it gives the best estimators for the model parameters. Other link functions can be used successfully, so there is no need to set aside practical considerations to use the canonical link function for Y.

For most of this paper, every parameter of the distribution of Y, apart from μ itself, will be considered a known constant. The derogatory-sounding term nuisance parameter is used to identify all parameters that are not of immediate interest.

The Dispersion Function a(ϕ)

The natural form includes a dispersion function $a(\phi)$ rather than a simple constant ϕ. This apparent complication provides an important extra degree of flexibility to model cases in which the Y_i are independent, but not identically distributed. The distributions of the Y_i have the same form, but not necessarily the same mean and variance.

We do not need to assume that every point in the historical sample of n observations has the same mean and variance. The mean μ_i is estimated as a function of a linear combination of predictors (covariates). The variance around this mean can also be a function of external information by making use of the dispersion function $a(\phi)$.

One way in which a model builder might make use of a dispersion function to help improve a model is to set $a(\phi) = \phi / w_i$, where ϕ is constant for all observations and w_i is a weight that may vary by observation. The values w_i are a priori weights based on external information that are selected in order to correct for unequal variances among the observations that would otherwise violate the assumption that ϕ is constant.

Now that we have seen how a non-constant dispersion function can be used to counteract non-constant variance in the response variable, we will assume that the weights are equal to unity, so that each observation is given equal weight.

The Variance Function Var(Y_i), and Uniqueness

Before looking at some specific distributions in the Natural Exponential Family, we define a uniqueness property of the variance structure in the natural exponential family. This property, presented concisely on page 51 of Jorgensen, states that the relationship between the variance and the mean (ignoring dispersion parameter ϕ) uniquely identifies the distribution.

In the notation of Appendix A, we write *Var(Y_i)* in terms of μ_i as $Var(Y_i) = a(\phi) \cdot V(\mu_i)$, so that the variance is composed of two components: one that depends on ϕ and external factors, and a second that relates the variance to the mean. The function $V(\mu)$, called the unit variance function, is what determines the form of a distribution, given that it is from the natural exponential family with parameters from a particular domain.

The upshot of this result is that, among continuous distributions in this family, $V(\mu) = 1$ implies we have a Normal with mean μ and variance $\phi = \sigma^2$, that $V(\mu) = \mu^2$ arises from a Gamma, and $V(\mu) = \mu^3$ from an Inverse Gaussian. For a discrete response, $V(\mu) = \mu$ means we have a Poisson

Uniqueness Property: *The unit variance function $V(\mu)$ uniquely identifies its parent distribution type within the natural exponential family.*

The implications of this Uniqueness Property are important for model design in GLM because it means that once we have defined a variance structure, we have specified the distribution form. Conversely, if a member of the Exponential Family is specified, the variance structure is also determined.

BASIC PROPERTIES OF SPECIFIC DISTRIBUTIONS

Our discussion of the natural exponential family will focus on five specific distributions:

- Normal (Gaussian)
- Poisson
- Gamma
- Inverse Gaussian
- Negative Binomial

The natural exponential family is broader than the specific distributions discussed here. It includes the Binomial, Logarithmic and Compound Poisson/Gamma (sometimes called "Tweedie" - see Appendix C) curves. The interested reader should refer to Jørgensen for details of additional members of the exponential family.

Many other distributions can be written in the general exponential form, if one allows for enough nuisance parameters. For instance, the Lognormal is seen to be a member of the general family by using $e(y) = \ln(y)$ instead of $e(y) = y$, but that excludes it from the natural exponential family. Using a Normal response variable in a GLM with a log link function applied to μ is quite different from applying a log transform to the response itself. The link function relates μ_i to the linear component; it does not apply to Y itself.

In the balance of this discussion, it is assumed that the variable Y is being modeled in currency units. The function $f(y)$ represents the probability or density function over a range of aggregate loss dollar amounts.

Appendix B gives "cheat sheet" summaries of the key characteristics of each distribution.

The Normal (Gaussian) Distribution

The Normal distribution occupies a central role in the historical development of statistics. Its familiar bell shape seems to crop up everywhere. Most linear regression theory depends on Normal approximations to the sampling distribution of estimators. Techniques used in parameter estimation, analysis of residuals, and testing model adequacy are guided largely by intuitions about the Normal curve and its properties.

The Normal has been criticized as a distribution for insurance losses because:

- Its range includes both negative and positive values.
- It is symmetrical, rather than skewed.
- The degree of dispersion supported by the Normal is quite limited.

Besides these criticisms, we should also note that a GLM with an unadjusted Normal response implies that the variance is constant, regardless of the expected loss volume. That is, if a portfolio with a mean of $1,000,000 has a standard deviation of $500,000, a larger portfolio with a $100,000,000 mean would have the same standard deviation.

A weighted dispersion function $a(\phi)=\phi/w_i$ can be used to provide more flexibility in adjusting for non-constant variance. The weights w_i can be set so that the variance for each predicted value μ_i is proportional to some exposure base such as on-level premium or revenue.

For the Normal distribution, this amounts to using weighted least squares. The parameters that minimize the sum of squares are equal to the parameters that maximize the likelihood. The least squares expression then becomes:

Ordinary Least Squares = $\sum(y_i-\mu_i)^2$

Weighted Least Squares = $\sum w_i \cdot (y_i-\mu_i)^2$

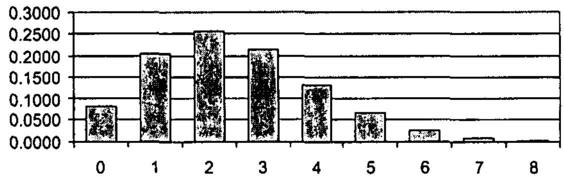
where w_i = 1 / Exposures for category i

Poisson and Over-Dispersed Poisson Distributions

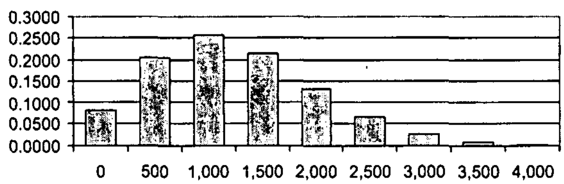
The Poisson distribution is a discrete distribution ranging over the non-negative integers. It has a mean equal to its variance.

The Over-Dispersed Poisson distribution is a generalization of the Poisson, in which the range is a constant ϕ times the positive integers. That is, the variable Y can take on values $\{0, 1\phi, 2\phi, 3\phi, 4\phi, \ldots\}$. It has a variance equal to ϕ times the mean.

Poisson Distribution



Over-Dispersed Poisson Distribution $\phi = 500$



The first important point to make concerning the Poisson is that, even though it is a discrete distribution, it can still be used as an approximation to a distribution of aggregate losses. There is no need to interpret the probabilities as anything other than a discretized version of an aggregate distribution. In fact, the Poisson immediately shows an advantage over the Normal:

- It is defined only over positive values
- It has positive skewness

An additional advantage of the Poisson is that it allows for a mass point at zero. The assumption that the ratio of the variance to the mean is constant is reasonable for insurance applications. Essentially, this means that when we add together independent random variables, we can add their means and variances. A very convenient property of the Over-Dispersed Poisson (ODP) is that the sum of ODP's that share a common scale parameter ϕ will also be ODP.

Gamma Distribution

The Gamma distribution is defined over positive values and has a positive skew. The probability density function, written in the natural exponential form, is:

$$f(y) = \exp\left[\alpha \cdot \left(\left(\frac{-y}{\mu}\right) - \ln(\mu)\right) + (\alpha - 1) \cdot \ln(\alpha \cdot y) + \ln\left(\frac{\alpha}{\Gamma(\alpha)}\right)\right]$$

From its form, we see that the Gamma belongs to one-parameter natural exponential family, but only if we assume that the shape parameter α is fixed and known By holding α constant, we treat the CV of the response variable as constant regardless of loss volume. As such, portfolios with expected losses of \$1,000,000 and \$100,000,000 would have the same CV. This seems unrealistic for many casualty insurance applications, although the Gamma may work well in high-volume lines of business, where GLM-based classification rating plans and bulk loss reserving models work best.

The Gamma distribution is closed under convolution in certain cases. When the PDF is written in the form below, the sum of two Gamma random variables $X_1 \sim Gamma(\alpha_1, \theta)$ and $X_2 \sim Gamma(\alpha_2, \theta)$ is also Gamma-distributed with $X_{1+2} \sim Gamma(\alpha_1 + \alpha_2, \theta)$, if they have a common θ. Unfortunately, we cannot capitalize on this property in GLM, since we require α to be constant and θ_i to vary.

$$f(y) = \frac{y^{\alpha-1}}{\theta^{\alpha} \cdot \Gamma(\alpha)} \cdot e^{-y/\theta}$$

Inverse Gaussian Distribution

The Inverse Gaussian distribution is occasionally recommended as a model for insurance losses, especially since its shape is very similar to the Lognormal.

The probability density function, written in the natural exponential form is:

$$f(y) = \exp\left[\left\{\left(\frac{-y}{2\mu^2}\right)+\left(\frac{1}{\mu}\right)\right\}\cdot\frac{1}{\phi}-\left(\frac{1}{2\phi y}+\ln\left(\sqrt{2\pi\phi\, y^3}\right)\right)\right].$$

In this form, the ϕ parameter is again treated as fixed and known. The variance is equal to $\phi\cdot\mu^3$. In other words, the variance is proportional to the mean loss amount cubed. This implies that the CV of a portfolio of losses would <u>increase</u> as the volume of loss increases, which is an unreasonable assumption for insurance phenomena.

The Inverse Gaussian distribution also has a practical difficulty that is worth noting. The difficulty is seen when the cumulative distribution function (CDF) is written:

$$F(y) = \text{NORMSDIST}\left(\frac{(y-\mu)}{\sqrt{y\cdot\mu}}\cdot\frac{1}{CV}\right)+\text{EXP}\left(\frac{2}{CV^2}\right)\cdot\text{NORMSDIST}\left(-\frac{(y+\mu)}{\sqrt{y\cdot\mu}}\cdot\frac{1}{CV}\right)$$

For small values of CV, this expression requires a very accurate evaluation for both EXP($\cdot$) and the tails of NORMSDIST($\cdot$) function. In practice, this represents a problem since commonly used software often does not provide values in the extreme tails.

The Negative Binomial Distribution

The Negative Binomial distribution, like the Poisson, is a discrete distribution that can be used to approximate aggregate loss dollars. As in the Over-Dispersed Poisson, we can add a scale parameter ϕ to increase the flexibility of the curve.

The Negative Binomial distribution has a variance function equal to:

$$Var(y) \quad = \quad \phi \cdot \mu + \frac{\phi}{k} \cdot \mu^2 \text{ with unit variance } V(\mu) = \mu \cdot \left(1 + \frac{\mu}{k}\right).$$

The variance can be interpreted as the sum of an unsystematic (or "random") component $\phi \cdot \mu$, and a systematic component $\frac{\phi}{k} \cdot \mu^2$. The inclusion of a systematic component implies that some relative variability, as measured by a coefficient of variation, remains even as the mean grows very large. That is,

$$\lim_{\mu \to \infty} CV \quad = \quad \lim_{\mu \to \infty} \frac{\sqrt{Var(y)}}{E[y]} \quad = \quad \lim_{\mu \to \infty} \sqrt{\frac{\phi}{\mu} + \frac{\phi}{k}} \quad = \quad \sqrt{\frac{\phi}{k}}.$$

We would expect the variance of a small portfolio of risks to be driven by random elements represented by the unsystematic component. As the portfolio grows by adding more and more similar risks, the variance would become dominated by the systematic component The parameter k can be interpreted as the expected size of loss μ for which the systematic and unsystematic components are equal.

Stated differently, the k parameter is a selected dollar amount. When the expected loss is below the amount k, the variance is closer to being proportional to the mean and the distribution starts to resemble the Poisson. When the expected loss is above the amount k, the variance is closer to being proportional to the mean squared and the distribution approaches a Gamma shape

This variance structure finds a close parallel to the concept of "mixing", as used in the Heckman-Meyers collective risk model. The unsystematic risk is then typically called the "process variance" and the systematic risk the "parameter variance".

$$Total\ Variance = \underbrace{E[Var(y)]}_{\substack{\text{Process} \\ \text{Variance}}} + \underbrace{Var(\mu)}_{\substack{\text{Parameter} \\ \text{Variance}}}$$

A practical calculation problem arises if we wish to simultaneously estimate the k and μ parameters. The k parameter is imbedded in a factorial function and is not independent of the scale parameter ϕ, as shown in the probability function below. Because of this complexity, the k will need to be set by the model user separately from the fit of μ. This can be repeated for different values, with a final selection made by the user.

$$\text{Prob}(Y = y) \quad = \quad \exp\left[\left(\ln\left(\frac{\mu}{\mu+k}\right)\cdot y + \ln\left(\frac{k}{\mu+k}\right)k\right)/\phi + \ln\binom{(k+y)/\phi - 1}{y/\phi}\right]$$

The Lognormal Distribution – Not!

Because of its popularity in insurance applications, it is worthwhile to include a brief discussion of the Lognormal distribution.

The Lognormal distribution is a member of the general exponential family, but its density cannot be written in the natural form:

$$f(y) \quad = \quad \exp\left[\left(\mu\cdot\ln(y) - \mu^2/2\right)/\phi - \left(\frac{\ln(y)^2}{2\phi} + \ln\left(\sqrt{2\pi\phi}\right) + \ln(y)\right)\right].$$

To employ a Lognormal model for insurance losses Y, we apply a log transform to the observed values of the response, and fit a Normal distribution assumption to the transformed data. The response variable is therefore $\ln(Y)$ rather than Y.

While it initially seems attractive to be able to use the lognormal along with GLM theory, there are a number of problems with this approach. The first is purely practical. Since we are applying a logarithmic transform $\ln(y)$ to our observed y_i, any zero or negative values make the formula unworkable. One possible workaround is to add a constant amount to each y_i in order to ensure that the logarithms exist.

A second problem is that while the estimate of $\hat{\mu}_i$ (the mean of $\ln(y_i)$) will be unbiased, we cannot simply exponentiate it to estimate the mean of y_i in the original scale of dollars. A bias correction is needed on the GLM results.

A third potential problem arises from the fact that the lognormal model implicitly assumes, as does the Gamma, that all loss portfolios have the same CV. If we believe that the y_i come from distributions with identical CV's, then the GLM model with the Gamma assumption can be used as an alternative to the Lognormal model. This would allow us to steer clear of the first two problems.

HIGHER MOMENT PROPERTIES OF SPECIFIC DISTRIBUTIONS

Now that we have reviewed the basic properties for five specific members of the natural exponential family, including their variance structure, we will examine the overall shape of the curves being used.

Moments

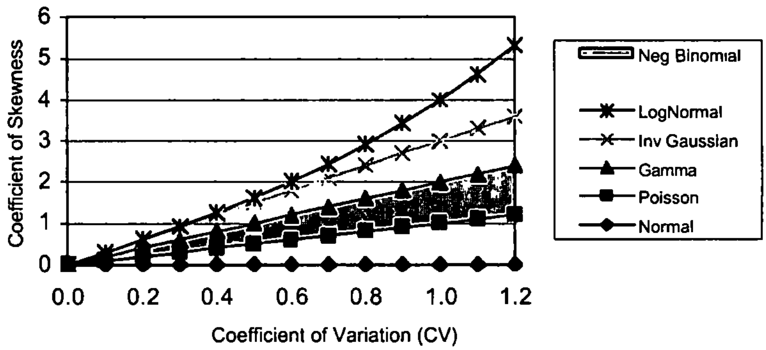
The variances for the natural exponential family members described in the previous section may be summarized as follows

Distribution	Variance
Normal	$Var(y) = \phi$
[Over-Dispersed] Poisson	$Var(y) = \phi \cdot \mu$ (constant V/M)
[Over-Dispersed] Negative Binomial	$Var(y) = \phi \cdot \mu + \frac{\phi}{k} \cdot \mu^2$
Gamma	$Var(y) = \phi \cdot \mu^2$ (constant CV)
Inverse Gaussian	$Var(y) = \phi \cdot \mu^3$

Two higher moments, representing skewness and kurtosis, can be represented in a similar sequence as functions of the CV.

	Skewness	Kurtosis
	$\frac{E\left[(Y-\mu)^3\right]}{Var(Y)^{3/2}}$	$\frac{E\left[(Y-\mu)^4\right]}{Var(Y)^{4/2}}$
Normal	0	3
Poisson	CV	$3+CV^2$
Negative Binomial	$(2-p)\cdot CV$	$3+\left(6\cdot(1-p)+p^2\right)\cdot CV^2$
Gamma	$2\cdot CV$	$3+6\cdot CV^2$
Inverse Gaussian	$3\cdot CV$	$3+15\cdot CV^2$

The Negative Binomial distribution can be seen to represent values in the range between the Poisson and Gamma distributions, since $0 < p < 1$. The graph below shows the relationship between the CV and the skewness coefficient.



The Lognormal distribution is shown for comparison sake, and has a coefficient of skewness equal to $(3 + CV^2) \cdot CV$.

Measuring Tail Behavior: The Unit Hazard Function $h_w(y)$

In order to evaluate tail behavior of the curves in the exponential family, we will examine the hazard function $h_w(y)$, the average hazard rate over an interval of fixed width "w".

Unit Hazard Function

$$h_w(y) = \frac{F(y+w) - F(y)}{1 - F(y)}$$ for continuous distributions, w = layer width

$$h_w(y) = \frac{\Pr(y < Y \leq y + w)}{\Pr(Y > y)}$$ for discrete distributions, w = fixed integer.

The more familiar hazard function $h(y) = f(y)/[1-F(y)]$ presented in Klugman [2003] is sometimes called the "failure rate", because it represents the conditional probability or density of a failure in a given instant of time, given that no failure has yet taken place. The unit hazard function measures the change in $F(y)$ over a small interval of width w, rather than a rate at a given instant in time.

The unit hazard function has a useful interpretation in insurance applications. It is roughly the probability of a partial limit loss in an excess layer. For example, in a layer of $10,000,000 excess $90,000,000, we seek the probability that a loss will not exceed $100,000,000, given that it is in the layer. A high value for $h_w(y)$ would mean that a loss above $90,000,000 would be unlikely to exhaust the full $10,000,000 layer

For most insurance applications, we would expect a decreasing unit hazard function. That is, as we move to higher and higher layers, the chance of a partial loss would decrease. For instance, if we consider a layer such as $10,000,000 xs $990,000,000 we would expect that any loss above $990,000,000 would almost certainly be a full-limit loss. This would imply $h_w(y) \to 0$.

The decreasing hazard function is not what we generally find in the exponential family. For the Normal and Poisson, the hazard function approaches 1, implying that full-limit losses become less likely on higher layers – exactly the opposite of what our understanding of insurance phenomena would suggest. The Negative Binomial, Gamma and Inverse Gaussian distributions asymptotically approach constant amounts, mimicking the behavior of the exponential distribution

The table below shows the asymptotic behavior as we move to higher attachment points for a layer of width w.

Distribution	Limiting Form of $h_w(y)$	Comments
Normal	$\lim_{y \to \infty} h_w(y) = 1$	No loss exhausts the limit
Poisson	$\lim_{y \to \infty} h_w(y) = 1$	
Negative Binomial	$\lim_{y \to \infty} h_w(y) = 1 - (1-p)^w$	
Gamma	$\lim_{y \to \infty} h_w(y) = 1 - e^{-w\alpha \cdot \mu}$	
Inverse Gaussian	$\lim_{y \to \infty} h_w(y) = 1 - e^{-w \cdot (2\phi\mu^2)}$	
Lognormal	$\lim_{y \to \infty} h_w(y) = 0$	Every loss is a full-limit loss

From this table, we see that the members of the natural exponential family have tail behavior that does not fully reflect the potential for extreme events in heavy casualty insurance. It would seem that the natural exponential distributions used with GLM are more appropriate for insurance lines without much potential for extreme events or natural catastrophes.

SMALL SAMPLE ISSUES

The results calculated in Generalized Linear Models generally rely on asymptotic behavior assuming a large number of observations are available. Unfortunately, this is not always the case in Property & Casualty insurance. For instance, in per-risk or per-occurrence excess of loss reinsurance, there may not be a large enough volume of losses to rely upon asymptotic approximations.

While we include here a brief discussion of the uncertainty in our parameter estimates, this is an area in which much more research is needed.

Including Uncertainty in the Mean μ

Most of our discussion of the exponential family has focused on the distribution of future losses around an estimated mean μ. However, the actuary is more often asked to provide a confidence interval around the estimated value of the mean $\hat{\mu}$. The estimate $\hat{\mu}$ is also a random variable, with a mean, variance and higher moments. However, GLM models generally produce an approximation to this distribution by making use of the asymptotic behavior of the coefficients $\hat{\beta}$ in the linear predictor being Normal.

The calculation of the variance in the parameter estimates, which leads to the confidence interval around the estimated mean $\hat{\mu}$, is accomplished using the matrix of second derivatives of the loglikelihood function. A comprehensive discussion of that calculation can be found in McCullagh & Nelder or Klugman [1998].

In general, the distribution of the estimator $\hat{\mu}$ will not be the same exponential family form as that of Y. In other words, the process and parameter variances are variances of different distribution forms. As a practical solution, the actuary will want to select a reasonable curve form (e.g., a gamma or lognormal) with mean and variance that match the estimated $\hat{\mu}$ and $Var(\hat{\mu})$ from the model.

Including Uncertainty in the Dispersion ϕ

In all of the discussion to this point, the dispersion parameter ϕ has been assumed to be fixed and known. It is estimated as a side calculation, separate from the estimate of the parameters $\hat{\beta}$ used to estimate the mean $\hat{\mu}$.

So long as the separate estimate of the dispersion parameter is based on a large number of observations, this approximation is reasonable A problem arises in certain insurance applications where there are relatively few observations, and our estimate of the dispersion is far from certain.

In normal linear regression, the uncertainty in the dispersion parameter (σ^2 instead of ϕ) is modeled by using a Student-t distribution rather than a Normal distribution. The use of a Student-t distribution is equivalent to an assumption that the parameter σ^2 (or ϕ) is distributed as Inverse Gamma with a shape parameter equal to its degrees of freedom ν. That is:

$$g(\phi) = \frac{\lambda^{\frac{\nu}{2}} \cdot e^{-\lambda/\phi}}{\phi^{\frac{\nu}{2}+1} \cdot \Gamma(\nu/2)} \qquad E[\phi^k] = \lambda^k \frac{\Gamma(\nu/2-1)}{\Gamma(\nu/2)}, \text{ for } 2k < \nu,$$

where ν = degrees of freedom.

A similar "mixing" of the dispersion parameter can be made for curves other than the Normal. It is not always easy to explicitly calculate the mixed distribution, but the moments can be found with the formula above.

For calculation purposes, if the distribution is used in a simulation model, the mixing can be accomplished in a two-step process. First we simulate a value for ϕ from an Inverse Gamma distribution. Second we simulate a value from the loss distribution conditional on the simulated ϕ.

The real difficulty with the uncertainty in the dispersion parameter is that it has a significant effect on the higher moments on the distribution, and therefore on the tail – the part of the distribution where the actuary may have the greatest concern. As the formula for the moments of the Inverse Gamma shows, many of the higher moments will not exist.

Another important note on the uncertainty in the dispersion parameter relates to the use of the Lognormal distribution. When the log transform is applied to the observed data in order to use linear regression, we have uncertainty in the dispersion of the logarithms $\ln(y_i)$. When the transformed data $\ln(y_i)$ has a Student-t distribution, the untransformed data y_i follows a Log-T distribution The Log-T has been recommended by Kreps and Murphy for use in estimating confidence intervals in reserving applications.

What neither author noted, however, is that none of the moments of the Log-T distribution exists. We are able to calculate percentiles, but not a "confidence interval" around the mean, because the mean itself does not exist.

CONCLUSIONS

The use of the Natural Exponential Family of distributions in GLM allows for more realistic variance structures to be used in modeling insurance phenomena This is a real advance beyond linear regression models, which are restricted to the Normal distribution.

The Natural Exponential Family also allows the actuary to work directly with their loss data in units of dollars, without the need for logarithmic or other transformations.

However, these advantages do not mean that GLM has resolved all issues for actuarial modeling. The curve forms are generally thin-tailed distributions and should be used with caution in insurance applications with potential for extreme events, or with a small sample of historical data.

REFERENCES

Dobson, Annette J., An Introduction to Generalized Linear Models, Second Edition, Chapman & Hall, 2002.

England, Peter D., and Richard J. Verrall, *A Flexible Framework for Stochastic Claims Reserving*, CAS Proceedings Vol. LXXXVIII, 2001.

Halliwell, Leigh J., *Loss Prediction by Generalized Least Squares*, CAS Proceedings Vol. LXXXIII, 1996.

Jørgensen, Bent, The Theory of Dispersion Models, Great Britain: Chapman & Hall, 1997.

Klugman, Stuart A., "Estimation, Evaluation, and Selection of Actuarial Models," CAS Exam 4 Study Note, 2003.

Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot, Loss Models: From Data to Decisions, New York: John Wiley & Sons, Inc , 1998.

Kreps, Rodney E., *Parameter Uncertainty in (Log)Normal Distributions*, CAS Proceedings Vol. LXXXIV, 1997.

Mack, Thomas *A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves*, ASTIN Bulletin, Vol. 21, No. 1, 1991.

McCullagh, P. and J.A Nelder, Generalized Linear Models, Second Edition, Chapman & Hall/CRC, 1989.

Mildenhall, Stephen J., *A Systematic Relationship Between Minimum Bias and Generalized Linear Models*, CAS Proceedings Vol. LXXXVI, 1999.

Murphy, Daniel, *Unbiased Loss Development Factors*, CAS Proceedings Vol. LXXXI, 1994.

Zehnwirth, Ben, *Ratemaking: From Bailey and Simon (1960) to Generalized Linear Regression Models*, CAS Forum Including the 1994 Ratemaking Call Papers, 1994.

Appendix A: Deriving Moments for the Natural Exponential Family

As stated in this paper, the probability density function $f(y)$ for the natural exponential family is given by:

$$f(y;\theta,\phi) = \exp\left[\,(\theta \cdot y - b(\theta))/a(\phi) + c(y,\phi)\right]$$

In the natural form, a, b, c are suitable known functions, θ is the canonical parameter for Y, and ϕ is the dispersion parameter. The unit cumulant function $b(\theta)$, which is useful in computing moments of Y, does not depend on y or ϕ. Likewise, the dispersion function $a(\phi)$ does not depend on y or θ. The catch-all function $c(y, \phi)$ has no dependence on θ.

The unit cumulant function $b(\theta)$ is so named because it can be used to calculate *cumulants*, which are directly related to the random variable's moments

We recall from Statistics that the Moment Generating Function $MGF(t)$ is defined as:

$$MGF(t) \;=\; \int_{-\infty}^{\infty} e^{ty} \cdot f(y)\,dy \qquad \text{for continuous variables}$$

and that

$$E\left[Y^r\right] \;=\; \left.\frac{\partial^r\, MGF(t)}{\partial t^r}\right|_{t=0}.$$

The Cumulant Generating Function $K(t)$ is defined as $\ln\left[MGF(t)\right]$, and the cumulants:

$$\kappa_r \;=\; \left.\frac{\partial^r\, K(t)}{\partial t^r}\right|_{t=0}.$$

There is an easy mapping between the first four cumulants and the moments:

$$\kappa_1 = E\left[y^1\right] = \mu \qquad\qquad \kappa_3 = E\left[(y-\mu)^3\right]$$

$$\kappa_2 = E\left[(y-\mu)^2\right] = Var(y) \qquad\qquad \kappa_4 = E\left[(y-\mu)^4\right] - 3\cdot Var(y)^2$$

For the Natural Exponential Family, the Cumulant Generating Function can be written in a very convenient form:

$$K(t) \quad = \quad \frac{b(\theta + a(\phi)\cdot t) - b(\theta)}{a(\phi)} \text{ , so that}$$

$$\kappa_r \quad = \quad b^{(r)}(\theta)\cdot a(\phi)^{r-1} \qquad \text{where } b^{(r)}(\theta) \quad = \quad \frac{\partial^r b(\theta)}{\partial \theta^r}.$$

In the mean value form, where $\theta = \tau^{-1}(\mu)$, the chain rule is used to find derivatives in terms of μ. The function $b''(\theta)$ is the <u>unit variance function,</u> denoted $V(\mu)$ when expressed in terms of μ.

Mean $E[Y;\theta] = b'(\theta) = \mu$

Variance $Var[Y;\theta] = b''(\theta)\cdot a(\phi) = V(\mu)\cdot a(\phi)$

$$Skewness = \frac{b^{(3)}(\theta)\cdot a(\phi)^2}{Var[Y;\theta]^{3/2}} = \frac{d}{d\mu}[V(\mu)]\cdot\sqrt{\frac{a(\phi)}{V(\mu)}}$$

$$Kurtosis = 3 + \frac{b^{(4)}(\theta)\cdot a(\phi)^3}{Var[Y;\theta]^2} = 3 + \left[\frac{d^2}{d\mu^2}[V(\mu)]\; V(\mu) + \left(\frac{d}{d\mu}V(\mu)\right)^2\right]\cdot\frac{a(\phi)}{V(\mu)}$$

Appendix B1: Normal Distribution

Density Function:

$$f(y) = \frac{1}{\sqrt{2\pi\phi}} \cdot \exp\left(\frac{-(y-\mu)^2}{2\phi}\right)$$

$$y \in (-\infty, \infty)$$

Natural Form:

$$f(y) = \exp\left[(\mu y - \mu^2/2)/\phi - \left(\frac{y^2}{2\phi} + \ln\left(\sqrt{2\pi\phi}\right)\right)\right]$$

Cumulative Distribution Function in Excel® Notation:

$$F(y) = \text{NORMDIST}\left(y, \mu, \sqrt{\phi}, 1\right)$$

Moments:

$$E[Y] = \mu$$

$$Var(Y) = \phi$$

$$Skewness = \frac{E\left[(Y-\mu)^3\right]}{Var(Y)^{3/2}} = 0$$

$$Kurtosis = \frac{E\left[(Y-\mu)^4\right]}{Var(Y)^{4/2}} = 3$$

Convolution of independent Normal random variables:

$$N_x(\mu_x, \phi_x) \otimes N_y(\mu_y, \phi_y) \Rightarrow N_{x+y}(\mu_x + \mu_y, \phi_x + \phi_y)$$

Appendix B2: Over-Dispersed Poisson

Probability Function:

$$\text{Prob}(Y = y) = \left(\frac{\mu}{\phi}\right)^{y/\phi} \frac{e^{-\mu/\phi}}{(y/\phi)!}$$

$$y \in (0,\ 1\phi,\ 2\phi,\ 3\phi,\ 4\phi,\ \ldots)$$

Natural Form

$$\text{Prob}(Y = y) = \exp\left[(\ln(\mu)\ y - \mu)/\phi - y \cdot \ln(\phi)/\phi - \ln((y/\phi)!)\right]$$

Cumulative Distribution Function in Excel® Notation:

$$\text{Prob}(Y \le y) = 1 - \text{GAMMADIST}\left(\frac{\mu}{\phi}, \frac{y}{\phi} + 1, 1, 1\right)$$

Moments:

$$E[Y] = \mu$$

$$Var(Y) = \phi\ \mu \qquad CV = \sqrt{\frac{\phi}{\mu}}$$

$$Skewness = \frac{E\left[(Y-\mu)^3\right]}{Var(Y)^{3/2}} = \sqrt{\frac{\phi}{\mu}} = CV$$

$$Kurtosis = \frac{E\left[(Y-\mu)^4\right]}{Var(Y)^{4/2}} = 3 + CV^2$$

Convolution of independent Over-Dispersed Poisson random variables:

$$ODP_x(\mu_x, \phi) \otimes ODP_y(\mu_y, \phi) \Rightarrow ODP_{x+y}(\mu_x + \mu_y, \phi)$$

where ϕ is a constant variance/mean ratio

Appendix B3: Gamma

Density Function: $$f(y) = \left(\frac{y\,\alpha}{\mu}\right)^{\alpha}\left(\frac{1}{y}\right)\cdot\frac{e^{-y\alpha/\mu}}{\Gamma(\alpha)}$$

$$y \in (0, \infty)$$

Natural Form: $$f(y) = \exp\left[\alpha\cdot\left(\left(\frac{-y}{\mu}\right)-\ln(\mu)\right)+(\alpha-1)\ \ln(\alpha\cdot y)+\ln\left(\frac{\alpha}{\Gamma(\alpha)}\right)\right]$$

Cumulative Distribution Function in Excel® Notation:

$$F(y) = \text{GAMMADIST}\left(\frac{y\cdot\alpha}{\mu}, \alpha, 1, 1\right)$$

Moments:

$$E[Y] = \mu$$

$$Var(Y) = \frac{\mu^2}{\alpha} \qquad CV = \sqrt{\frac{1}{\alpha}}$$

$$Skewness = \frac{E\left[(Y-\mu)^3\right]}{Var(Y)^{3/2}} = \frac{2}{\sqrt{\alpha}} = 2\cdot CV$$

$$Kurtosis = \frac{E\left[(Y-\mu)^4\right]}{Var(Y)^{4/2}} = 3+6\cdot CV^2$$

Convolution of independent Gamma random variables:

$$G_x(\mu_x, \alpha_x = \mu_x/\beta)\otimes G_y(\mu_y, \alpha_y = \mu_y/\beta) \Rightarrow G_{x+y}(\mu_x+\mu_y, \alpha_x+\alpha_y)$$

where β is a constant variance/mean ratio

Appendix B4: Inverse Gaussian

Density Function. $$f(y) = \frac{1}{\sqrt{2\pi\phi y^3}} \cdot \exp\left(\frac{-(y-\mu)^2}{2\phi\mu^2 y}\right)$$

$$y \in (0, \infty)$$

Natural Form: $$f(y) = \exp\left[\left\{\left(\frac{-y}{2\mu^2}\right) + \left(\frac{1}{\mu}\right)\right\} \cdot \frac{1}{\phi} - \left(\frac{1}{2\phi y} + \ln\left(\sqrt{2\pi\phi y^3}\right)\right)\right]$$

Cumulative Distribution Function in Excel® Notation:

$$F(y) = \text{NORMSDIST}\left(\frac{(y-\mu)}{\mu \cdot \sqrt{\phi \cdot y}}\right) + \text{EXP}\left(\frac{2}{\phi \cdot \mu}\right) \cdot \text{NORMSDIST}\left(-\frac{(y+\mu)}{\mu \cdot \sqrt{\phi \cdot y}}\right)$$

Moments:

$$E[Y] = \mu$$

$$Var(Y) = \phi \cdot \mu^3 \qquad CV = \sqrt{\phi \cdot \mu}$$

$$Skewness = \frac{E\left[(Y-\mu)^3\right]}{Var(Y)^{3/2}} = 3 \cdot \sqrt{\phi \cdot \mu} = 3 \cdot CV$$

$$Kurtosis = \frac{E\left[(Y-\mu)^4\right]}{Var(Y)^{4/2}} = 3 + 15 \cdot CV^2$$

Convolution of independent Inverse Gaussian random variables:

$$IG_x\left(\mu_x, \phi_x = \beta / \mu_x^2\right) \otimes IG_y\left(\mu_y, \phi_y = \beta / \mu_y^2\right) \Rightarrow IG_{x+y}\left(\mu_x + \mu_y, \phi_{x+y} = \beta / (\mu_x + \mu_y)^2\right)$$

where β is a constant variance/mean ratio

Appendix B5: [Over-Dispersed] Negative Binomial

Probability Function:

$$\text{Prob}(Y=y) = \binom{(k+y)/\phi - 1}{y/\phi} \cdot p^{k/\phi} \cdot (1-p)^{y/\phi}$$

$$y \in (0,\ 1\phi,\ 2\phi,\ 3\phi,\ 4\phi,\ \ldots)$$

Natural Form:

$$\text{Prob}(Y=y) = \exp\left[\left(\ln\left(\frac{\mu}{\mu+k}\right)\cdot y + \ln\left(\frac{k}{\mu+k}\right)\cdot k\right)/\phi + \ln\binom{(k+y)/\phi - 1}{y/\phi}\right]$$

Cumulative Distribution Function in Excel® Notation:

$$\text{Prob}(Y \le y) = \text{BETADIST}\left(\frac{k}{\mu+k}, \frac{\mu}{\phi}, \frac{y}{\phi}+1\right)$$

Moments.

$$E[Y] = k\cdot\frac{(1-p)}{p} = \mu \qquad \text{so } p = \frac{k}{\mu+k}$$

$$Var(Y) = \phi\cdot k\cdot\frac{(1-p)}{p^2} = \phi\cdot\mu + \frac{\phi}{k}\cdot\mu^2$$

$$CV = \sqrt{\frac{\phi}{\mu}+\frac{\phi}{k}}$$

$$Skewness = \frac{E\left[(Y-\mu)^3\right]}{Var(Y)^{3/2}} = (2-p)\cdot CV$$

$$Kurtosis = \frac{E\left[(Y-\mu)^4\right]}{Var(Y)^{4/2}} = 3 + \left(6\cdot(1-p)+p^2\right)\cdot CV^2$$

Convolution of independent Over-Dispersed Negative Binomial random variables:

$$NB_x(\mu_x, \phi, p) \otimes NB_y(\mu_y, \phi, p) \Rightarrow NB_{x+y}(\mu_x+\mu_y, \phi, p)$$

Appendix C: Compound Poisson/Gamma (Tweedie) Distribution

The Tweedie distribution can be interpreted as a collective risk model with a Poisson frequency and a Gamma severity.

Probability Function:

$$f(y \mid \lambda, \theta, \alpha) = \begin{cases} e^{-\lambda} & y = 0 \\ \displaystyle\sum_{k=1}^{\infty} \underbrace{\frac{\lambda^k e^{-\lambda}}{k!}}_{\text{Poisson}} \cdot \underbrace{\frac{y^{k\alpha - 1} e^{-y/\theta}}{\theta^{k\alpha} \cdot \Gamma(k\alpha)}}_{\text{Gamma}} & y > 0 \end{cases} \qquad y \in (0, \infty)$$

This form appears complicated, but can be re-parameterized to follow the natural exponential family form.

We set: $\alpha = \dfrac{2-p}{p-1}$ $\quad \lambda = \dfrac{\mu^{2-p}}{\phi \cdot (2-p)}$ $\quad \theta = \phi \cdot (p-1) \cdot \mu^{p-1}$

and $1 < p < 2$, since $p = \dfrac{\alpha + 2}{\alpha + 1}$ and $\alpha > 0$

$$f(y \mid \mu, \phi, p) = \exp\left[\left(\frac{\mu^{2-p}}{(2-p)} + \frac{y}{(p-1) \cdot \mu^{p-1}}\right) \cdot \frac{-1}{\phi}\right] c(y, \phi)$$

where

$$c(y, \phi) = \begin{cases} 1 & y = 0 \\ \displaystyle\sum_{k=1}^{\infty} \frac{y^{k(2-p)/(p-1)-1}}{[\phi \cdot (2-p)]^k \cdot [\phi \cdot (p-1)]^{k(2-p)/(p-1)} \cdot \Gamma[k(2-p)/(p-1)] \cdot k!} & y > 0 \end{cases}$$

The density function $f(y \mid \mu, \phi, p)$ can then be seen to follow the "natural form" for the exponential family.

Moments: $E[Y] = \lambda \cdot \theta \cdot \alpha = \mu$

$$Var(Y) = \lambda \cdot \theta^2 \alpha (\alpha+1) = \phi \mu^p$$

$$CV = \sqrt{\frac{1}{\lambda} + \frac{1}{\lambda\alpha}} = \sqrt{\frac{\phi}{\mu^{2-p}}}$$

$$Skewness = \frac{E[(Y-\mu)^3]}{Var(Y)^{3/2}} = \frac{\lambda \cdot \theta^3 \cdot \alpha \cdot (\alpha+1) \cdot (\alpha+2)}{(\lambda \cdot \theta^2 \cdot \alpha \cdot (\alpha+1))^{3/2}} = p \cdot CV$$

$$Kurtosis = \frac{E[(Y-\mu)^4]}{Var(Y)^{4/2}} = 3 + p \cdot (2p-1) \ CV^2$$

For GLM, a p value in the (1, 2) range must be selected by the user. The mean μ and dispersion ϕ are then estimated by the model.

The Compound Poisson/Gamma is a continuous distribution, with a mass point at zero. The evaluation of the cumulative distribution function (CDF) is somewhat inconvenient, but can be accomplished using any of the collective risk models available to actuaries.

Finally, we may note that the convolution of independent Tweedie random variables:

$$TW_x(\lambda_x, \theta, \alpha) \otimes TW_y(\lambda_y, \theta, \alpha) \Rightarrow TW_{x+y}(\lambda_x + \lambda_y, \theta, \alpha)$$