Quantum Generalized Linear Models: A Proof of Concept

Uchenna Chukwu and Colleen M. Farrelly, Quantopo LLC

Big Picture Summary

Overview and Implications

- Generalized linear models are the simplest instance of link-based statistical models, which are based on the underlying geometry of an outcome's underlying probability distribution (typically from the exponential family).
- Machine learning algorithms provide alternative ways to minimize a model's sum of square error (error between predicted values and actual values of a test set).
- However, some deep results regarding the exponential family's relation to affine connections in differential geometry provide a possible alternative to link functions:
    1. Algorithms that either continuously deform the outcome distribution from known results
    2. Algorithms that superpose all possible distributions and collapse to fit a dataset
    - Leveraging the fact that some quantum computer gates, such as the non-Gaussian transformation gate, essentially perform (1) natively and in a computationally-efficient way!
- This project provides a proof-of-concept for leveraging specific hardware gates to solve the affine connection problem, with benchmarking at state-of-the-art levels.
- Results can be extended to many other, more complicated statistical models, such as generalized estimating equations, hierarchical regression models, and even homotopy-continuation problems.

Generalized Linear Model Background and Business Usage

An Introduction to Tweedie Models

Generalized Linear Models

- Generalized linear modeling (GLM) as extension of linear regression to outcomes with probability distributions that are not Gaussian
  - Includes binomial outcomes, Poisson outcomes, gamma outcomes, and many more
  - Link functions to transform distribution of these outcomes to a normal distribution to fit a linear model
    - $E(Y) = \mu = g^{-1}(\boldsymbol{X}\beta)$
    - $\mathrm{Var}(Y) = Var(\mu) = Var(g^{-1}(\boldsymbol{X}\beta))$
    - Where Y is a vector of outcome values, μ is the mean of Y, **X** is the matrix of predictor values, g is a link function (such as the log function), and β is a vector of predictor weights in the regression equation.
  - Many statistical extensions:
    - Generalized estimating equations (longitudinal data modeling)
    - Generalized linear mixed models (longitudinal data with random effects)
    - Generalized additive models (in which the predictor vectors can be transformed within the model)
    - Cox regression and Weibull-based regression (survival data modeling)
    - Very high computational cost for many of these extensions

Important Applications of GLMs

- Ubiquitous in part failure modeling, medical research, actuarial science, and many other problems
  - Example problems:
    - Modeling likelihood of insurance claims and expected payout (worldwide, a $5 trillion industry)
    - Understanding risk behavior in medical research (daily heroin usage, sexual partners within prior month…)
    - Modeling expected failure rates and associated conditions for airplane parts or machine parts within a manufacturing plant (~$4 trillion industry in the USA alone)
    - Modeling expected natural disaster impacts and precipitating factors related to impact extent
  - Many supervised learning algorithms are extensions of generalized linear models and have link functions built into the algorithm to model different outcome distributions
    - Boosted regression, Morse-Smale regression, dgLARS, Bayesian model averaging…
    - Optimization algorithm to find minimum sum of square error differ among machine learning methods and with respect to GLMs, which use a least square error algorithm
    - Methods like deep learning and classical neural networks attempt to solve this problem in a general way through a series of general mappings leading to a potentially novel link function
    - Exploiting the geometric relationships between distributions through a superposition of states collapsed to the "ideal" link would present an optimal solution to the problem
  - Tweedie regression as a general framework that handles many distributions in the exponential family and the problem of overdispersion of model/outcome variance
    - Very nice geometric properties
    - Connected to many common exponential family distributions

Details of Tweedie Regression

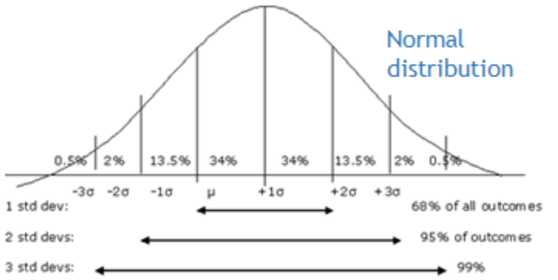
- Many common distributions of the exponential family converge to Tweedie distributions and can be formulated through Tweedie distributions, formally defined as:
  - $E(Y) = \mu$
  - $Var(Y) = \varphi\mu^{\xi}$
    - where φ is the dispersion parameter, and ξ is the Tweedie parameter (or shape parameter)
- Tweedie distributions themselves enjoy a variety of useful properties:
  - Reproductive properties that allow distributions to be added together to form new distributions that are themselves Tweedie
  - Varying Tweedie parameter and dispersion parameter to recover many exponential family distributions used in GLMs:
    - Tweedie parameter of 0 for normal distribution
    - Tweedie parameter of 1 for Poisson distribution
    - Tweedie parameter of 2 for gamma distribution
    - Dispersion parameter for 0-inflated models and outliers, similar to negative binomial regression models

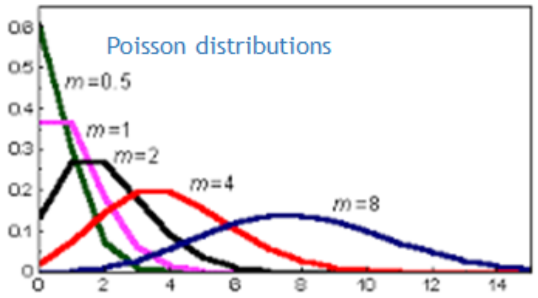
The Problem of Overdispersion in Tweedie Models

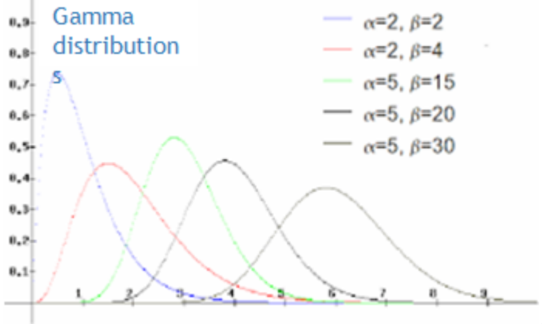
- Well-known statistical problem involving dispersion parameters, which relate to the variance of a outcome
- Many GLMs and their machine learning extensions struggle on problems of overdispersion
  - Simulations show this behavior, particularly as dispersion parameter increases substantially (values of 4+)
  - Empirical datasets with 0-inflation and long tails
- Recent paper exploring bagged KNN models
  - Demonstrates problem in simulations
  - Demonstrates with open-source datasets, such as UCI's Forest Fire dataset
  - Models that work well, such as the KNN ensemble with varying k parameters, tend to take a long time to compute

Common Tweedie Models

Family Distribution	Dispersion (extra 0’s and tail fatness)	Power (variance proportional to mean: 1/Power)
Normal	1	0
Poisson	**1**	**1**
Compound Poisson	1	>1 and <2
Gamma	1	2
Inverse-Gaussian	1	3
Stable	1	>2 (Extreme >3)
Negative Binomial	>1	1
Underdispersion Poisson	<1	1
Unique Tweedie	>=1	>=0





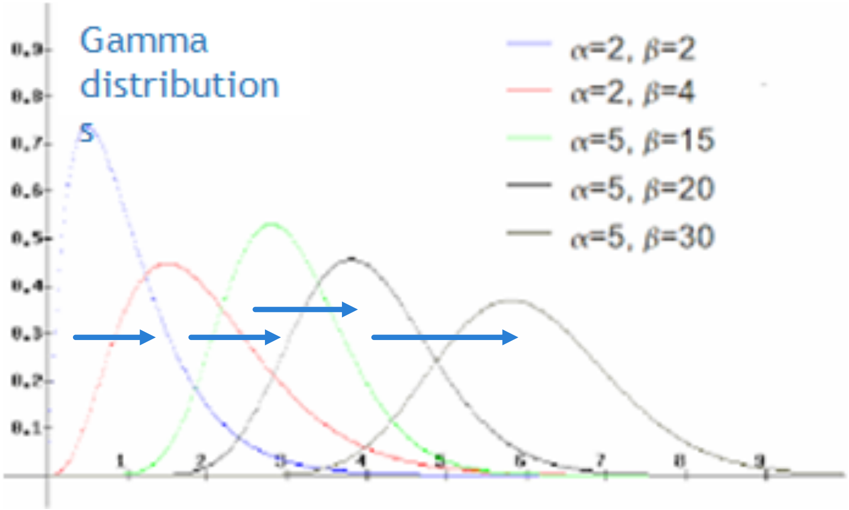


Connection of GLMs to Differential Geometry

Motivation for Implementation on Xanadu System

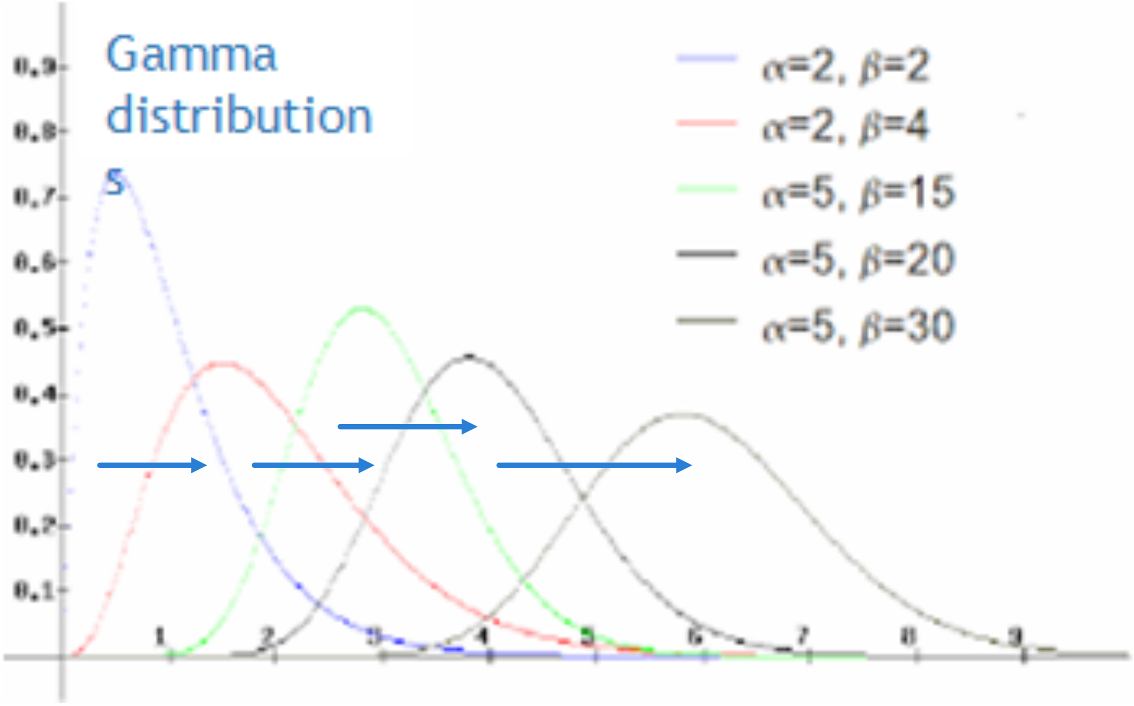
Differential Geometry and the Exponential Family

- Possible to formulate exponential family distributions and their parameterizations to form a series of curves on a 2-dimensional surface
- Each curve defined by 2 points at either end of the probability function, 0 and 1, connected by a line that follows a shortest path following parameterization of the distribution, called a geodesic
- Because the exponential family can be generalized into Tweedie distributions through continuous transformations, the geodesic connecting 0 and 1 can flow across distributions defining the 2-dimensional surface in a continuous manner (much like homotopy continuation methods).
- This is an affine connection, and the morphing of the line as it passes parameters transforms one distribution to another.



Consequences of Exponential Family Geometry

- Analytically-derived results/equation for one distribution morphed to fit another distribution through continuous transformations!
- Limit theorems derived by continuous deformations of either moment generating functions or characteristic functions



Xanadu Technology and Suitability to GLMs

- Xanadu's qumode formulation makes ideal for implementing quantum GLMs
  - Ability to perform linear algebra operations on physical data representations
    - GLMs and their extensions all based on simple matrix operations
    - $Mean(Y) = g^{-1}(\boldsymbol{X}\beta + \varepsilon)$
    - Matrix multiplication and addition for the linear model $(\boldsymbol{X}\beta + \varepsilon)$ coupled with a continuous transformation of the model results to fit the outcome distribution
  - Non-Gaussian transformation gate provides perfect avenue to perform the affine transformation related to the outcome distribution without a need to specific a link function to approximate the geometry
    - Should be able to approximate any continuous outcome's distribution, creating potential new "link functions" through this gate through affine transformation of the wavefunctions representing the data
    - Removes the need for approximations by easy-to-compute link transformations
    - In theory, should approximate any continuous distribution, including ones that aren't included in common statistical packages implementing GLMs and their longitudinal/survival data extensions
    - Thus, Xanadu's system provides a general solution to the linear regression equation with many potential extensions to more sophisticated regression models!

Methods and Results on Example Cases

Simulated overdispersion dataset and UCI Forest Fire dataset

Methodology

- Simulation
  - Similar to simulations used in the KNN ensemble paper
  - 1000 observations with a 70/30 test/train split
  - Tweedie outcome related to 3 predictors (1 interaction term, 1 main effect) with added noise
    - Tweedie parameter=1, dispersion parameter=8
  - 1 noise variable added
- Empirical dataset
  - UCI Repository's Forest Fire dataset
    - Notoriously difficult to beat the mean model with machine learning algorithms
    - 12 predictors (2 spatial coordinates of location, month, day, FFMC index, DMC index, DC index, ISI index, temperature, relative humidity, wind, and rain) and 517 observations
    - t-SNE was used to reduce the dimensionality of the predictor set to 4 components so as to make it compatible with Xanadu's capabilities.
    - 70/30 test/train split
- Comparison methods
  - Boosted regression
  - Random forest (tree-based bagged ensemble)
  - DGLARS (tangent-space-based least angle regression model)
  - BART (Bayesian-based tree ensemble)
  - HLASSO (homotopy-based LASSO model)
  - Poisson regression (GLM without any modifications)

Data Preprossessing

- Dimensionality reduction through t-SNE to create a set of 4 predictors and 1 outcome, such that predictors are uncorrelated when entered into models.
  - Easier for systems to calculate with fewer variables.
  - Decorrelation helps most regression methods, including linear models and tree models.
  - Other dimensionality reduction methods are possible, including the introduction of factors from factor analytic models or combinations of linear/nonlinear, global/local dimensionality reduction algorithms.
- Scaling of outcome to a scale of -3 to 3, such that the Xanadu simulation can effectively model and process the data in qumodes.
  - Slight warping of the most extreme values, but these are generally less than 5 observations per dataset.
  - Other types of scaling might be useful to explore.

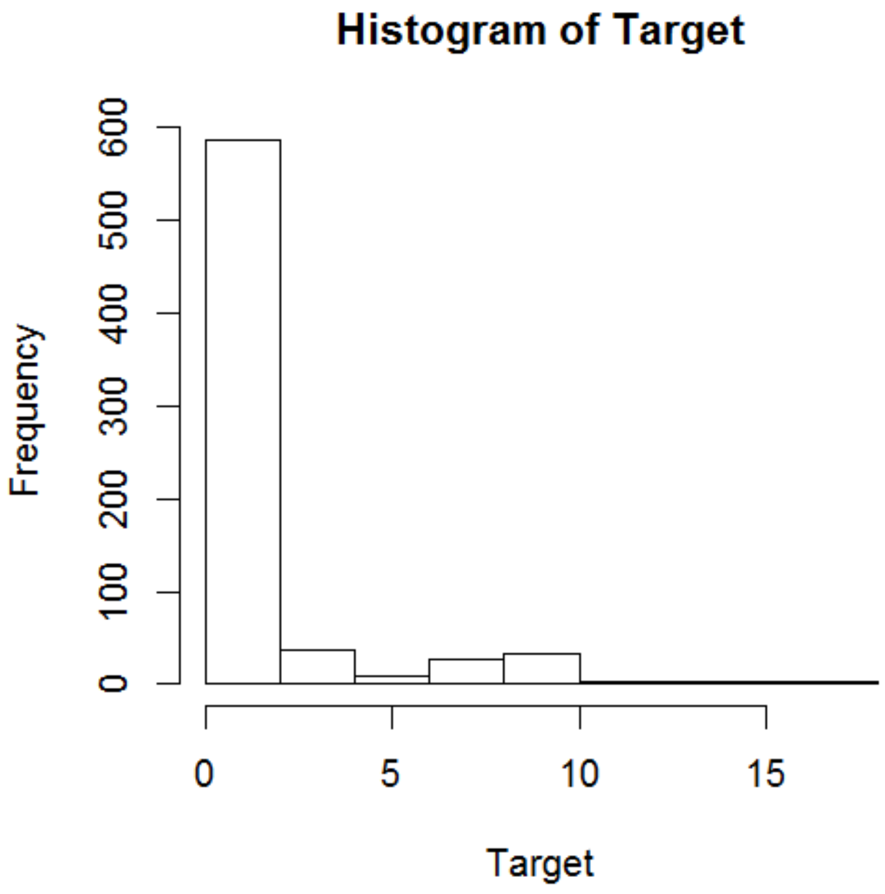
Qumodes Circuit Details

- GLMs can be embedded within Xanadu's qumode quantum computer simulation software (and qumode computer) with a singular value decomposition of the β coefficient in the formulation:
  - $Mean(Y) = g^{-1}(X\beta)$
- This translates to $\beta = O_1 \Sigma O_2$, which can be modeled through a series of quantum circuit gates:
  - Multiplication of X and an orthogonal matrix:
    - $|O_1 X\rangle \cong U_1 |X\rangle$, which corresponds to a linear interferometer gate (U_1) acting on X
  - Multiplication of that result by a diagonal matrix:
    - $|\Sigma O_1 X\rangle \propto S(r) |O_1 X\rangle$, which corresponds to a squeezing gate that acts on a single qumode
  - Multiplication of X and an orthogonal matrix:
    - $|O_2 \Sigma O_1 X\rangle \cong U_2 |\Sigma O_1 X\rangle$, which corresponds to a linear interferometer gate (U_2) acting on the result
  - Multiplication by a nonlinear function on this result:
    - $|g^{-1}(O_2 \Sigma O_1 X)\rangle \cong \Phi |O_2 \Sigma O_1 X\rangle$, which corresponds to the non-Gaussian gate acting on the result
- This gives a final result of gates acting upon the dataset as:
  - $\Phi * \mathcal{U}_2 * \mathcal{S} * \mathcal{U}_1 |X\rangle \propto |g^{-1}(X\beta)\rangle$

Qumodes Parameter Settings

- The algorithm simulation was created through Strawberry Fields.
  - The deep learning framework already existed.
    - Hidden layers and bias terms were removed to collapse to a generalized linear model framework.
    - The loss function optimized was mean square error, which corresponds to the loss functions specified in the comparison algorithms.
    - Qumode cut-off dimension was set to 10.
    - Optimization via least squares was not available, so gradient descent was used with a learning rate of 0.1 over 80 iterations.
  - This gave a qumodes implementation of a quantum generalized linear model with a boosting feel to it.
- Because the quantum computing component is inherently probabilitistic, algorithms were run on the same training and test set multiple times to average out quantum effects.

Results: Simulation of Overdispersion



Algorithm	Scaled Model MSE
Random Forest	0.80
BART	**0.78**
Boosted Regression	**0.78**
DGLARS	0.81
HLASSO	0.81
GLM	0.81
QGLM	0.82
Mean	0.85

QGLMs yield slightly worse prediction on the simulated dataset. However, their performance is not far off from state-of-the art algorithms, and some random error is expected from the quantum machinery.

Results: Forest Fire Dataset

Algorithm	Scaled Model MSE
Random Forest	0.125
BART	0.125
Boosted Regression	0.119
DGLARS	0.114
HLASSO	0.120
GLM	0.119
QGLM	**0.106**
Mean	0.115

QGLMs emerge as the best-performing algorithm on a difficult, real-world dataset (Forest Fire dataset in the UCI repository). QGLMs provide ~10% gain over the next best algorithm on this dataset. This suggests that they work well on real data and difficult problems.

Conclusions

- This suggests that the qumodes formulation with its unique operators can eliminate the need for link functions within linear models by exploiting the geometry of the models and still give good prediction.
  - Better than state-of-the-art prediction for a difficult Tweedie regression dataset (UCI Forest Fire)
  - Around state-of-the-art prediction for a simulated dataset
- This has the potential to bring statistical modeling into quantum computing, by leveraging the underlying geometry and the connection between model geometry and the geometry of quantum physics.
  - Generalized estimating equations/generalized linear mixed models
  - Structural equation models/hierarchical regression models
  - Also a potential avenue through which to implement the homotopy continuation method common in dynamic systems research and some machine learning models (such as homotopy-based LASSO), which take a known problem's solution and continuously deform it to fit the problem of interest.
    - Currently a computational challenge
    - Limited to small datasets

References

- Amari, S. I. (1997). Information geometry. Contemporary Mathematics, 203, 81-96.
- Bulmer, M. G. (1974). On fitting the Poisson lognormal distribution to species-abundance data. Biometrics, 101-110.
- Buscemi, F. (2012). Comparison of quantum statistical models: equivalent conditions for sufficiency. *Communications in Mathematical Physics*, *310*(3), 625-647.
- Cortez, P., & Morais, A. D. J. R. (2007). A data mining approach to predict forest fires using meteorological data.
- De Jong, P., & Heller, G. Z. (2008). Generalized linear models for insurance data (Vol. 10). Cambridge: Cambridge University Press.
- Farrelly, C. M. (2017). KNN Ensembles for Tweedie Regression: The Power of Multiscale Neighborhoods. *arXiv preprint arXiv:1708.02122*.
- Farrelly, C. M. (2017). Topology and Geometry in Machine Learning for Logistic Regression.
- Fehm, L., Beesdo, K., Jacobi, F., & Fiedler, A. (2008). Social anxiety disorder above and below the diagnostic threshold: prevalence, comorbidity and impairment in the general population. Social psychiatry and psychiatric epidemiology, 43(4), 257-265.
- Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2006). Cannabis use and other illicit drug use: testing the cannabis gateway hypothesis. Addiction, 101(4), 556-569.
- Frees, E. W., Lee, G., & Yang, L. (2016). Multivariate frequency-severity regression models in insurance. Risks, 4(1), 4.
- Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychological bulletin, 118(3), 392.
- Herings, R., & Erkens, J. A. (2003). Increased suicide attempt rate among patients interrupting use of atypical antipsychotics. Pharmacoepidemiology and drug safety, 12(5), 423-424.

References

- Jorgensen, B. (1997). The theory of dispersion models. CRC Press.
- Jørgensen, B., Goegebeur, Y., & Martínez, J. R. (2010). Dispersion models for extremes. Extremes, 13(4), 399-437.
- Killoran, N., Bromley, T. R., Arrazola, J. M., Schuld, M., Quesada, N., & Lloyd, S. (2018). Continuous-variable quantum neural networks. *arXiv preprint arXiv:1806.06871.*
- Killoran, N., Izaac, J., Quesada, N., Bergholm, V., Amy, M., & Weedbrook, C. (2018). Strawberry Fields: A Software Platform for Photonic Quantum Computing. *arXiv preprint arXiv:1804.03159.*
- Luitel, B. N. (2016). Prediction of North Atlantic tropical cyclone activity and rainfall (Doctoral dissertation, The University of Iowa).
- Marriott, P. (1990). Applications of differential geometry to statistics (Doctoral dissertation, University of Warwick).
- Mills, K. L., Teesson, M., Ross, J., & Peters, L. (2006). Trauma, PTSD, and substance use disorders: findings from the Australian National Survey of Mental Health and Well-Being. American Journal of Psychiatry, 163(4), 652-658.
- Nielsen, F., & Garcia, V. (2009). Statistical exponential families: A digest with flash cards. arXiv preprint arXiv:0911.4863.
- Osborne, M. R., Presnell, B., & Turlach, B. A. (2000). A new approach to variable selection in least squares problems. *IMA journal of numerical analysis*, *20*(3), 389-403.
- Pistone, G., & Rogantin, M. P. (1999). The exponential statistical manifold: mean parameters, orthogonality and space transformations. Bernoulli, 5(4), 721-760.
- Sawalha, Z., & Sayed, T. (2006). Traffic accident modeling: some statistical issues. Canadian Journal of Civil Engineering, 33(9), 1115-1124.
- Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. In Statistics: Applications and new directions: Proc. Indian statistical institute golden Jubilee International conference (Vol. 579, p. 6o4).