



Search

Try Premium Free
for 1 Month

Income over \$102,306? - Get \$2MM in Life Insurance from \$63/month*. As seen on WSJ, CNN



Yuanyuan Liu • 3rd

Director, Statistical Machine Learning, AIG

London, United Kingdom



AIG



University of Oxford



See contact info



500+ connections

• Comprehensive academic background in Statistics (15+ years) • Expertise in Quantitative Analytics and AI design for Insurance and Investment • Interested in developing and integrating AI/ML components into Finance and building AI-first strategies/products/platforms

Highlights



You both worked at AIG

Yuanyuan worked at AIG after you started

[Say hello](#)

Articles & activity

1,275 followers



Why AIG? Why Science?



Yuanyuan Liu

Published on LinkedIn

Every now and then, people ask Why AIG? Why Science? Except for its global reputation, there are many unique features embedded in this great company. From personal perspective, AIG's pioneer...

[...see more](#)

25 Likes



Like



Comment



Share

[See all articles](#)

Looking forward to join Jürgen Schmidhuber and #andrewng at next...

Yuanyuan shared this

31 Likes

[See all activity](#)

Experience



AIG

5 yrs 6 mos

Promoted

\$500/hr Advisory
Roles

Many companies are looking for paid Business Advisors. Are you interested?

[Learn more](#)Vanderbilt's Online
Ed.D.

Earn an Ed.D. in leadership and learning in organizations online.

[Learn more](#)

People Also Viewed



Nils Mork-Ulness • 3rd

Head Of Strategy, Investments AI at AIG



Stefan Zohren • 3rd

Machine Learning Researcher I
Quantitative Strategist

Johannes Bergström • 3rd

Machine Learning Scientist at AIG



Danijela Horak • 3rd

Senior Research Scientist at AIG



Jessica Cheng FIA • 3rd

Actuary at AIG

Wang Han • 2nd

Data Scientist, Investments AI at AIG



Reza Khorshidi, D.Phil. (Oxon)...

Chief Scientist, Global at AIG



John Greenall • 3rd

Director, Investments AI at AIG



Sid Kumar • 3rd

AI Product Manager at AIG Science



Diana Turcsany • 3rd

Researcher with experience in Deep Learning, Machine Learning, Computer Vision, NLP & Neural Computation

Learn the skills Yuanyuan has

Advanced SAS
Programming for R Users,
Part 2
Viewers: 4,803

SAS Essential Training: 2
Regression Analysis for
Healthcare Research

Business Math
Viewers: 32,935

[See more courses](#)[Messaging](#)



Try Premium Free
for 1 Month

Fully AI-powered Robo Trading Based on Deep Neural Networks and Alternative Data



- Senior Manager, Quantitative Analytics**
Apr 2016 – Mar 2017 · 1 yr
London, United Kingdom
Deep Recommendation System
Deep Insurance Claim / Credit Default Inference
 Adversarial variational inference for Tweedie...
- Manager, Quantitative Analytics**
Apr 2015 – Mar 2016 · 1 yr
London, United Kingdom
Submission Prioritization
Opportunity Mapping (Recommender System)
Client Lifetime Value
- Senior Quantitative Analyst**
Dec 2013 – Mar 2015 · 1 yr 4 mos
London, United Kingdom
SME Loss-Risk
Retention



Class Tutor and Teaching Assistant
University of Oxford
Sep 2009 – Jun 2013 · 3 yrs 10 mos
Oxford, Oxfordshire, United Kingdom



Marketing Manager Assistant (Intern)
Anheuser-Busch
Nov 2005 – Jan 2006 · 3 mos
London, Greater London, United Kingdom

Education



University of Oxford
DPhil, Statistics (Machine Learning)
2009 – 2013
Activities and Societies: Vice-President of CSSA, 2009-2010 (University of Oxford Society)
Research topic: "Monte Carlo inference for classification trees"
Supervisor: Prof. Brian D. Ripley



University of Oxford
MSc, Applied Statistics, Distinction
2008 – 2009
Dissertation: "Internet traffic classification using k-nearest neighbours, trees, and random forests"





University of Bristol
BSc, Mathematics with Statistics, First Class Honours (Top 1st)
2004 – 2008
Activities and Societies: President of CSSA, 2005-2006 (University of Bristol Society)







Licenses & Certifications



Graduate Statistician

 Messaging





Try Premium Free for 1 Month

Skills & Endorsements

R · 37

Endorsed by Hamaad Shah and 3 others who are highly skilled at this

Endorsed by 5 of Yuanyuan's colleagues at AIG

Statistics · 28

Endorsed by Reza Khorshidi, D.Phil. (Oxon) and 2 others who are highly skilled at this

Endorsed by 7 of Yuanyuan's colleagues at AIG

Data Mining · 24

Endorsed by Reza Khorshidi, D.Phil. (Oxon), who is highly skilled at this

Endorsed by 7 of Yuanyuan's colleagues at AIG


Show more

Accomplishments


2 Languages

Chinese · English


Interests




Data Mining, Statistics, Big Data, D...
182,803 members




The R Project for Statistical Compu...
117,497 members




University of Oxford
433,620 followers



Pattern Recognition & Machine Le...
49,244 members



QuantumBlack
11,204 followers



Winton
17,155 followers


See all


LinkedIn

About
Careers
Advertising
Small Business

Talent Solutions
Marketing Solutions
Sales Solutions
Safety Center

Community Guidelines
Privacy & Terms
Mobile


 **Questions?**
Visit our Help Center.

 **Manage your account and privacy.**
Go to your Settings.

Select Language

English (English)

LinkedIn Corporation © 2019

 **Messaging**

3 of 3

5/10/2019, 8:20 AM

[SPEAKERS \(/EVENTS/DEEP-LEARNING-FINANCE-SUMMIT-LONDON-2016/SPEAKERS\)](/EVENTS/DEEP-LEARNING-FINANCE-SUMMIT-LONDON-2016/SPEAKERS)[SCHEDULE \(/EVENTS/DEEP-LEARNING-FINANCE-SUMMIT-LONDON-2016/SCHEDULE\)](/EVENTS/DEEP-LEARNING-FINANCE-SUMMIT-LONDON-2016/SCHEDULE)[VENUE \(/EVENTS/DEEP-LEARNING-FINANCE-SUMMIT-LONDON-2016/VENUE\)](/EVENTS/DEEP-LEARNING-FINANCE-SUMMIT-LONDON-2016/VENUE)[PARTNERS \(/EVENTS/DEEP-LEARNING-FINANCE-SUMMIT-LONDON-2016#PARTNERS_ANCHOR\)](/EVENTS/DEEP-LEARNING-FINANCE-SUMMIT-LONDON-2016#PARTNERS_ANCHOR)[SPONSORSHIP \(/SPONSORSHIP-INQUIRY/NEW?EVENT=DEEP+LEARNING+IN+FINANCE+SUMMIT\)](/SPONSORSHIP-INQUIRY/NEW?EVENT=DEEP+LEARNING+IN+FINANCE+SUMMIT)[REGISTER \(/EVENTS/DEEP-LEARNING-FINANCE-SUMMIT-LONDON-2016/REGISTER\)](/EVENTS/DEEP-LEARNING-FINANCE-SUMMIT-LONDON-2016/REGISTER)

Yuanyuan Liu

Overview of Artificial Intelligence in Finance with Applications in Insurance

As many have argued, the rise of artificial intelligence could be seen as the second machine age. The exponential growth in computing power, memory capacity, distributional modules, and most importantly the data (structured and unstructured) created, all require algorithmic innovations to dig out meaningful business insights. Machine learning, especially deep learning is undoubtedly a crucial key to future. The latest development in computer vision, natural language processing, and reinforcement learning, could easily improve / alter our current insurance structures. Generative models have the potential to reveal deep insights of clients hence leads to deep underwriting / customized healthcare strategy. Sequential models together with reinforcement learning could provide deep understanding of financial time series to assist investment decisions

Yuanyuan Liu joined AIG Science in 2013. During the past 5 years, he has led multiple global projects such as SME loss-risk analysis, client lifetime value model, submission prioritization, and recommender systems using advanced machine learning techniques. He and his team has conducted edge-cutting researches and published a series of paper in the world's top machine learning conferences including ICML, NIPS, and AAAI etc. Most recently, Yuanyuan is working on AIG's innovative R&D projects to apply deep learning algorithms in insurance and investment, using generative models, sequential models, computer vision, and reinforcement learning. Yuanyuan graduated from Oxford with a DPhil in Statistical Data Mining and an MSc in Applied Statistics. Prior to that, he studied a Mathematics with Statistics major in Bristol

 (<https://uk.linkedin.com/in/yuanyuan-liu-a6506725>)



Join RE•WORK Slack channel (https://join.slack.com/t/rework-dl/shared_invite/enQtNDAYOTE1MzAxNDI4LTJhMmVhOTQ5ZjZGY4MzNhNGIxN2Y0Yzc2ZDlwYzEzZjQyMzMzMjMDdhYjUyNjAzOTg0MWRmYjk2ODc5OTMxMmY)

ADVERSARIAL VARIATIONAL BAYES METHODS FOR TWEEDIE COMPOUND POISSON MIXED MODELS

Yaodong Yang*, Rui Luo, Yuanyuan Liu

American International Group Inc.

ABSTRACT

The Tweedie Compound Poisson-Gamma model is routinely used for modeling non-negative continuous data with a discrete probability mass at zero. Mixed models with random effects account for the covariance structure related to the grouping hierarchy in the data. An important application of Tweedie mixed models is pricing the insurance policies, e.g. car insurance. However, the intractable likelihood function, the unknown variance function, and the hierarchical structure of mixed effects have presented considerable challenges for drawing inferences on Tweedie. In this study, we tackle the Bayesian Tweedie mixed-effects models via variational inference approaches. In particular, we empower the posterior approximation by implicit models trained in an adversarial setting. To reduce the variance of gradients, we reparameterize random effects, and integrate out one local latent variable of Tweedie. We also employ a flexible hyper prior to ensure the richness of the approximation. Our method is evaluated on both simulated and real-world data. Results show that the proposed method has smaller estimation bias on the random effects compared to traditional inference methods including MCMC; it also achieves a state-of-the-art predictive performance, meanwhile offering a richer estimation of the variance function.

Index Terms— Tweedie model, variational inference, insurance policy pricing

1. INTRODUCTION

Tweedie models [1, 2] are special members in the exponential dispersion family; they specify a power-law relationship between the variance and the mean: $\text{Var}(Y) = \text{E}(Y)^{\mathcal{P}}$. For arbitrary positive \mathcal{P} , the index parameter of the variance function, outside the interval of $(0, 1)$, Tweedie corresponds to a particular stable distribution, e.g., Gaussian ($\mathcal{P} = 0$), Poisson ($\mathcal{P} = 1$), Gamma ($\mathcal{P} = 2$), Inverse Gaussian ($\mathcal{P} = 3$). When \mathcal{P} lies in the range of $(1, 2)$, the Tweedie model is equivalent to Compound Poisson-Gamma Distribution [3], hereafter *Tweedie* for simplicity. Tweedie serves as a special Gamma mixture model, with the number of mixtures determined by a Poisson-distributed random variable, parameterized by $\{\lambda, \alpha, \beta\}$ and denoted as: $Y = \sum_{i=1}^T G_i, T \sim$

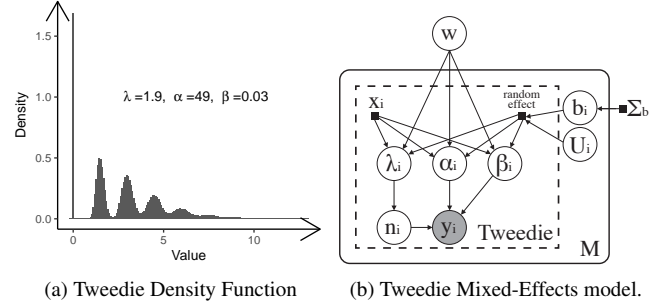


Fig. 1: Graphical model of Tweedie mixed-effect model.

$\text{Poisson}(\lambda), G_i \stackrel{i.i.d}{\sim} \text{Gamma}(\alpha, \beta)$. Tweedie is heavily used for modeling non-negative heavy-tailed continuous data with a discrete probability mass at zero (see Fig. 1a). As a result, Tweedie gains its importance from multiple domains [4, 5], including actuarial science (aggregated loss/premium modeling), ecology (species biomass modeling), meteorology (precipitation modeling). On the other hand, in many field studies that require manual data collection, for example in insurance underwriting, the sampling heterogeneity from a hierarchy of groups/populations has to be considered. Mixed-effects models can represent the covariance structure related to the grouping hierarchy in the data by assigning common random effects to the observations that have the same level of a grouping variable; therefore, estimating the random effects is also an important component in Tweedie modeling.

Despite the importance of Tweedie mixed-effects models, the intractable likelihood function, the unknown variance index \mathcal{P} , and the hierarchical structure of mixed-effects (see Fig. 1b) hinder the inferences on Tweedie models. Unsurprisingly, there has been little work devoted to full-likelihood based inference on the Tweedie mixed model, let alone Bayesian treatments. In this work, we employ variational inference to solve Bayesian Tweedie mixed models. The goal is to introduce an accurate and efficient inference algorithm to estimate the posterior distribution of the fixed-effect parameters, the variance index, and the random effects.

2. RELATED WORK

To date, most practice of Tweedie modeling are conducted within the quasi-likelihood (QL) framework [6] where the density function is approximated by the first and second order

*Correspondence to: <yaodong.yang@aig.com>.

moments. QL does not come with the schema of estimating the variance function, even though the estimation of the index parameter plays an important role in hypothesis testing and uncertainty evaluation [7]. Extended quasi-likelihood (EQL) [8], together with the profile likelihood method [9] is proposed to fill the gap of index estimation for QL. Nonetheless, EQL is not capable of handling exact zero values.

It was not until recently that numerical approximations to the Tweedie density function appeared. Well-known approaches include series expansion method [10] and Fourier inversion method [11]. Yang et al. [12] and Chen et al. [13] applied the gradient boosted trees that try to incorporate non-linear interactions in estimating the mean value of the Tweedie model. Qian et al. [14] investigated the grouped elastic net and Lasso methods for the Tweedie model. Nonetheless, the dependence on density approximations also makes the above algorithms vulnerable to numerical issues. Both series expansion and Fourier inversion methods are noted to suffer from unbounded increment in the number of required terms.

When the Tweedie model is expended to incorporate the random effects, the profiled quasi-likelihood (PQL) method still suffers from the numerical restrictions. Likelihood-based inference methods require approximation of the integral over the distribution of the random effect. Advances in numerical approximation like Laplace approximation [15] and adaptive Gauss-Hermite quadrature (AGQ) [16] are then applied to solve the mixed model. However the complex density of the Tweedie model still bothers the profiled likelihood methods.

The expectation of a latent variable for an intractable likelihood function can typically be estimated using the EM algorithm in an empirical Bayes approach or MCMC methods in full Bayesian approach. Simsekli et al. [3, 17] proposed solving the Tweedie mixed-effect models by explicitly exploiting the latent variable in the Bayesian formulation. Both MCMC and Monte Carlo EM methods have been applied. Zhang et al. [17] compared the latent variable approach with the density function approximations aforementioned, and found that MCMC are computationally demanding on the high dimensional dataset, and dedicated supervisions are needed in order to adjust the Monte Carlo error and the sampling variance [18].

3. TWEEDIE MIXED-EFFECTS MODELS

Tweedie EDM $f_Y(y|\mu, \phi, \mathcal{P})$ equivalently describes the compound Poisson–Gamma distribution when $\mathcal{P} \in (1, 2)$. Tweedie assumes Poisson arrivals of the events, and Gamma-distributed “cost” for each individual event. Judging on whether the aggregated number of arrivals N is zero, the joint density for each observation can be written as:

$$P(Y, N = n|\lambda, \alpha, \beta) = d_0(y) \cdot e^{-\lambda} \cdot \mathbb{1}_{n=0} + \frac{y^{n\alpha-1} e^{-y/\beta}}{\beta^{n\alpha} \Gamma(n\alpha)} \cdot \frac{\lambda^n e^{-\lambda}}{n!} \cdot \mathbb{1}_{n>0}, \quad (1)$$

where $d_0(\cdot)$ is the Dirac Delta function at zero. The connection of the parameters $\{\lambda, \alpha, \beta\}$ of Tweedie Compound

Poisson model with the parameters of the general Tweedie EDM model, parameterized by the mean, the dispersion and the index parameter $\{\mu, \mathcal{P}, \phi\}$ is denoted as:

$$\begin{cases} \lambda = \frac{\mu^{2-\mathcal{P}}}{\phi(2-\mathcal{P})} \\ \alpha = \frac{2-\mathcal{P}}{\mathcal{P}-1} \\ \beta = \phi(\mathcal{P}-1)\mu^{\mathcal{P}-1} \end{cases}, \quad \begin{cases} \mu = \lambda\alpha\beta \\ \mathcal{P} = \frac{\alpha+2}{\alpha+1} \\ \phi = \frac{\lambda^{1-\mathcal{P}}(\alpha\beta)^{2-\mathcal{P}}}{2-\mathcal{P}}. \end{cases} \quad (2)$$

A mixed-effects model contains both fixed effects and random effects; graphical model is shown in Fig. 1b. We denote the mixed model as: $\kappa(\mu_i) = f_w(\mathbf{X}_i) + \mathbf{U}_i^\top \mathbf{b}_i, \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \Sigma_b)$ where \mathbf{X}_i is the i -th row of the design matrix of fixed-effects covariates, w are parameters of the fixed-effects model which could represent linear function or deep neural network. \mathbf{U}_i is the i -th row of the design matrix associated with random effects, \mathbf{b}_i is the coefficients of the random effect which is usually assumed to follow a multivariate normal distribution with zero mean and covariance Σ_b , $\kappa(\cdot)$ is the link function, and μ_i is the mean of the i -th response variable Y . In this work, we have considered the random effects on the intercept.

In the context of conducting Bayesian inference on Tweedie mixed models, we define 1) the observed data $D = (\mathbf{x}_i, \mathbf{u}_i, y_i)_{i=1, \dots, M}$. 2) global latent variables $\{\mathbf{w}, \sigma_b\}$, we assume Σ_b is a diagonal matrix with its diagonal elements σ_b ; 3) local latent variables $\{n_i, \mathbf{b}_i\}$, indicating the number of arrivals and the random effect. The parameters of Tweedie is thus denoted by $(\lambda_i, \alpha_i, \beta_i) = f_{\lambda, \alpha, \beta}(\mathbf{x}_i, \mathbf{u}_i|\mathbf{w}, \sigma_b)$. The latent variables thus contain both local and global ones $\mathbf{z} = (\mathbf{w}, n_i, \mathbf{b}_i, \sigma_b)_{i=1, \dots, M}$, and they are assigned with prior distribution $P(\mathbf{z})$. The joint log-likelihood is computed by summing over the number of observations M by $\sum_{i=1}^M \log [P(y_i|n_i, \lambda_i, \alpha_i, \beta_i) \cdot P(n_i|\lambda_i) \cdot P(\mathbf{b}_i|\sigma_b)]$. The goal here is to find the posterior distribution of $P(\mathbf{z}|D)$ and make future predictions via $P(y_{\text{pred}}|D, \mathbf{x}_{\text{pred}}, \mathbf{u}_{\text{pred}}) = \int P(y_{\text{pred}}|\mathbf{z}, \mathbf{x}_{\text{pred}}, \mathbf{u}_{\text{pred}})P(\mathbf{z}|D) d\mathbf{z}$.

4. METHODS

Adversarial Variational Bayes. Variational Inference (VI) [19] approximates the posterior distribution, often complicated and intractable, by proposing a class of probability distributions $Q_\theta(\mathbf{z})$ (so-called inference models), and then finding the best set of parameters θ by minimizing the KL divergence between the proposal and the true distribution, *i.e.*, $\text{KL}(Q_\theta(\mathbf{z})||P(\mathbf{z}|D))$. Minimizing the KL divergence is equivalent to maximizing the evidence of lower bound (ELBO) [20], expressed as Eq. 3. Optimizing the ELBO requires the gradient information of $\nabla_\theta \text{ELBO}$. In our experiments, we find that the model-free gradient estimation method – REINFORCE [21] fails to yield reasonable results due to the unacceptably-high variance issue [22], even equipped with *baseline* trick [23] or *local expectation* [24]. We also attribute the unsatisfactory results to the over-simplified proposal distribution in traditional VI. Here we try to solve these two issues by employing the

implicit inference models with variance reduction tricks.

$$\theta^* = \arg \max_{\theta} \mathbb{E}_{Q_{\theta}(\mathbf{z})} \left[-\log \frac{Q_{\theta}(\mathbf{z})}{P(\mathbf{z})} + \log P(D|\mathbf{z}) \right] \quad (3)$$

AVB [25, 26] empowers the VI methods by using neural networks as the inference models; this allows more efficient and accurate approximations to the posterior distribution. Since a neural network is black-box, the inference model is implicit thus have no closed form expression, even though this does not bother to draw samples from it. To circumvent the issue of computing the gradient from implicit models, the mechanism of adversarial learning is introduced; an additional discriminative network $T_{\phi}(\mathbf{z})$ is used to model the first term in Eq. 3. By building a model to distinguish the latent variables that are sampled from the prior distribution $p(\mathbf{z})$ from those that are sampled from the inference network $Q_{\theta}(\mathbf{z})$, namely, optimizing the blow equation (where $\sigma(x)$ is the sigmoid function):

$$\phi^* = \arg \min_{\phi} \left[-\mathbb{E}_{Q_{\theta}(\mathbf{z})} \log \sigma(T_{\phi}(\mathbf{z})) - \mathbb{E}_{P(\mathbf{z})} \log(1 - \sigma(T_{\phi}(\mathbf{z}))) \right], \quad (4)$$

the ratio is estimated as $\mathbb{E}_{Q_{\theta}(\mathbf{z})} [\log \frac{Q_{\theta}(\mathbf{z})}{P(\mathbf{z})}] = \mathbb{E}_{Q_{\theta}(\mathbf{z})} [T_{\phi^*}(\mathbf{z})]$. AVB considers optimizing Eq. 3 and Eq. 4 as a two-layer mini-max game. We apply stochastic gradient descent alternately to find a Nash-equilibrium. Such Nash-equilibrium, if reached, is a global optimum of the ELBO [26].

Reparameterizable Random Effect. In mixed-effects models, the random effect is conventionally assumed to be $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{b}})$. In fact, they are reparameterisable. As such, we model the random effects by the reparameterization trick [27]; \mathbf{b}_i is now written as $\mathbf{b}(\epsilon_i) = \mathbf{0} + \sigma_{\mathbf{b}} \odot \epsilon_i$ where $\epsilon_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The $\sigma_{\mathbf{b}}$ is a set of latent variables generated by the inference network and the random effects become *deterministic* given the auxiliary noise ϵ_i .

Theorem 1 (Reparameterizable Random Effect). *Random effect is naturally reparameterizable. With the reparameterization trick, the random effect is no longer restricted to be normally distributed. For any “location-scale” family distributions (Students’t, Laplace, Elliptical, etc.)*

Proof. See the Section 2.4 in [27] \square

Note that when computing the gradients, we no longer need to sample the random effect directly, instead, we can back-propagate the path-wise gradients which could dramatically reduce the variance [28].

Hyper Priors. The priors of the latent variables are fixed in traditional VI. Here we however parameterize the prior $P_{\psi}(\mathbf{w})$ by ψ and make ψ also trainable when optimizing Eq. 3. We refer ψ as a kind of *hyper prior* to \mathbf{w} . The intuition is to not constraint the posterior approximation by one over-simplified prior. We would like the prior to be adjusted so as to make the posterior $Q_{\theta}(\mathbf{z})$ close to a set of prior distributions, or a self-adjustable prior; this could further ensure the expressiveness of $Q_{\theta}(\mathbf{z})$. We can also apply the same trick if the class of prior is reparameterizable.

Variance Reduction. We find that integrating out the latent variable n_i by Monte Carlo in Eq. 1 gives significantly lower variance in computing the gradients. As n_i is a Poisson generated sample, the variance will explode in the cases where Y_i is zero but the sampled n_i is positive, and Y_i is positive but the sampled n_i is zero. This also accounts for why the direct application of REINFORCE algorithm fails to work. In practice, we find limiting the number of Monte Carlo samples between 2 – 10, dependent on the dataset, has the similar performance as summing over to larger number.

$$P(y_i|\mathbf{w}, \mathbf{b}, \Sigma) = P(\mathbf{b}, \Sigma) \cdot \sum_{j=1}^T P(y_i|n_j, \mathbf{w}, \mathbf{b}, \Sigma) \cdot P(n_j|\mathbf{w}, \mathbf{b}, \Sigma) \quad (5)$$

Algorithm 1 AVB for Bayesian Tweedie mixed model

```

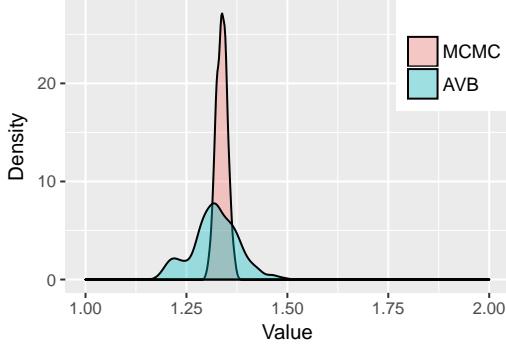
1: Input: data  $D = (\mathbf{x}_i, \mathbf{u}_i, y_i), i = 1, \dots, M$ 
2: while  $\theta$  not converged do
3:   for  $N_T$ : do
4:     Sample noises  $\epsilon_{i,j=1,\dots,M} \sim N(\mathbf{0}, \mathbf{I})$ ,
5:     Map noise to prior  $\mathbf{w}_j^P = P_{\psi}(\epsilon_j) = \mu + \sigma \odot \epsilon_j$ ,
6:     Reparameterize random effect  $\mathbf{b}_j^P = \mathbf{0} + \sigma_{\mathbf{b}} \odot \epsilon_j$ 
7:     Map noise to posterior  $\mathbf{z}_i^Q = (\mathbf{w}_i, \mathbf{b}_i)^Q = Q_{\theta}(\epsilon_i)$ ,
8:     Minimize Eq. 4 over  $\phi$  via the gradients:
        $\nabla_{\phi} \frac{1}{M} \sum_{i=1}^M [-\log \sigma(T_{\phi}(\mathbf{z}_i^Q)) - \log(1 - \sigma(T_{\phi}(\mathbf{z}_i^P)))]$ 
9:   end for
10:  Sample noises  $\epsilon_{i=1,\dots,M} \sim N(\mathbf{0}, \mathbf{I})$ ,
11:  Map noise to posterior  $\mathbf{z}_i^Q = (\mathbf{w}_i, \mathbf{b}_i)^Q = Q_{\theta}(\epsilon_i)$ ,
12:  Sample a mini-batch of  $(\mathbf{x}_i, y_i)$  from  $D$ ,
13:  Minimize Eq. 3 over  $\theta, \psi$  via the gradients:
        $\nabla_{\theta, \psi} \frac{1}{M} \sum_{i=1}^M [-T_{\phi^*}(\mathbf{z}_i^Q) + \sum_{j=1}^T P(y_i|n_j, \mathbf{w}, \mathbf{b}, \sigma_{\mathbf{b}}; \mathbf{x}_i, \mathbf{u}_i) \cdot P(n_j|\mathbf{w}; \mathbf{x}_i, \mathbf{u}_i) \cdot P(\mathbf{b} | \sigma_{\mathbf{b}})]$ 
14: end while
```

5. EXPERIMENTS AND RESULTS

We compare our method with six traditional inference methods on Tweedie. We evaluate our method on two public datasets of modeling the aggregate loss for auto insurance policies [31], and modeling the length density of fine roots in ecology [32]. We separate the dataset in 50%/25%/25% for train/valid/test respectively. Considering the sparsity and right-skewness of the data, we use the ordered Lorenz curve and its corresponding Gini index [33, 34] as the metric. Assuming for the i_{th}/N observations, y_i is the ground truth, p_i to be the results from the baseline predictive model, \hat{y}_i to be the predictive results from the model. We sort all the samples by the relativity $r_i = \hat{y}_i/p_i$ in an increasing order, and then compute the empirical cumulative distribution as $(\hat{F}_p(r) = \frac{\sum_{i=1}^n p_i \cdot \mathbb{1}(r_i \leq r)}{\sum_{i=1}^n p_i}, \hat{F}_y(r) = \frac{\sum_{i=1}^n y_i \cdot \mathbb{1}(r_i \leq r)}{\sum_{i=1}^n y_i})$. The plot of $(\hat{F}_p(r), \hat{F}_y(r))$ is an ordered Lorenz curve, and Gini index is the twice the area between the Lorenz curve and the 45°

Table 1: The pairwise Gini index comparison with standard error based on 20 random splits

MODEL	GLM[29]	PQL [29]	LAPLACE [15]	AGQ[16]	MCMC[17]	TDBOOST[30]	AVB
GLM (AUTOCLAIM)	/	-2.97 _{6.28}	1.75 _{5.68}	1.75 _{5.68}	-15.02 _{7.06}	1.61 _{6.32}	9.84 _{5.80}
PQL	7.37 _{5.67}	/	7.50 _{6.26}	7.50 _{5.72}	6.73 _{5.95}	0.81 _{6.22}	9.02 _{6.07}
LAPLACE	2.10 _{4.52}	-1.00 _{5.94}	/	8.84 _{5.36}	4.00 _{4.61}	21.45 _{4.84}	20.61 _{4.54}
AGQ	2.10 _{4.52}	-1.00 _{5.94}	8.84 _{5.36}	/	4.00 _{4.61}	21.45 _{4.84}	20.61 _{4.54}
MCMC	14.75 _{6.80}	-1.06 _{6.41}	3.12 _{5.99}	3.12 _{5.99}	/	7.82 _{5.83}	11.88 _{5.50}
TDBOOST	17.52 _{4.80}	17.08 _{5.36}	19.30 _{5.19}	19.30 _{5.19}	11.61 _{4.58}	/	20.30 _{4.97}
AVB	-0.17 _{4.70}	0.04 _{5.62}	3.41 _{4.94}	3.41 _{4.94}	0.86 _{4.62}	11.49 _{4.93}	/
GLM (FINERROOT)	/	23.18 _{9.30}	35.87 _{6.52}	35.87 _{6.52}	-15.73 _{10.63}	35.71 _{6.52}	35.64 _{6.53}
PQL	-21.61 _{8.34}	/	30.90 _{8.86}	30.90 _{8.86}	-21.45 _{8.38}	24.19 _{8.36}	28.97 _{8.92}
LAPLACE	-12.55 _{7.37}	-14.72 _{8.81}	/	15.07 _{6.41}	-12.55 _{7.37}	15.33 _{7.36}	10.61 _{7.20}
AGQ	-12.55 _{7.37}	-14.72 _{8.81}	15.07 _{6.41}	/	-12.55 _{7.37}	15.33 _{7.36}	10.61 _{7.20}
MCMC	17.27 _{10.25}	22.53 _{9.31}	35.10 _{6.54}	35.10 _{6.54}	/	35.10 _{6.54}	34.87 _{6.55}
TDBOOST	22.47 _{6.80}	8.50 _{9.09}	22.63 _{6.80}	22.63 _{6.80}	11.61 _{6.80}	/	22.39 _{6.80}
AVB	-8.26 _{17.66}	-10.88 _{8.98}	2.13 _{7.28}	2.13 _{7.28}	-8.26 _{7.66}	11.00 _{7.74}	/

**Fig. 2:** Posterior estimation of the index parameter \mathcal{P} on the AutoClaim data.

line. The Gini index is not symmetrical if different baselines are used. In short, a model with a larger Gini index indicates its superior capability of separating the observations. In the insurance context, Gini index profiles the model’s capability of distinguishing policy holders with different risk levels.

Gini comparison is presented in Table. 1. Here only fixed effects are considered. AVB method shows the state-of-the-art performance on both AutoClaim and FineRoot datasets, even comparatively better than the boosting tree method that is considered to have a strong performance. By inspecting the posterior estimation of \mathcal{P} in Fig. 2, we find AVB shows a richer representation of the index parameter, with two modals at 1.20 and 1.31. As \mathcal{P} uniquely characterizes a Tweedie distribution, compared with one single value that traditional profiled likelihood method offers, flexible choices of \mathcal{P} enable customized descriptions for insurance policies that may have different underlying risk profiles. Also note that AVB uses a neural sampler that does not involve any rejection procedures; unlike MCMC, it holds the potentials for large-scale predictions on high dimensional dataset.

To compare the estimations on random effects, we add “CARTYPE” in the AutoClaim data as the random effect with 6 groups and “PLANT” in the FineRoot data with 8 groups.

Param.	$\mathcal{P}_{\text{auto}}$	$\sigma_{b,\text{auto}}^2$	$\mathcal{P}_{\text{root}}$	$\sigma_{b,\text{root}}^2$
Algo.				
PQL	/	$8.94 \cdot 10^{-5}$	/	$1.0346 \cdot 10^{-3}$
LAPLACE	1.3394	$4.06 \cdot 10^{-4}$	1.4202	$6.401 \cdot 10^{-3}$
AGQ	1.3394	$4.027 \cdot 10^{-4}$	1.4202	$6.401 \cdot 10^{-3}$
MCMC	1.3458	$4.575 \cdot 10^{-3}$	1.4272	$3.490 \cdot 10^{-2}$
AVB	1.3403	$3.344 \cdot 10^{-3}$	1.4237	$2.120 \cdot 10^{-2}$

Table 2: Estimation on \mathcal{P} and σ_b^2 for both AutoClaim and FineRoot data with random effects.

From table 2, we can see that all the algorithms estimate \mathcal{P} in a reasonable similar range while the results from MCMC and AVB are closer. The estimation of σ_b^2 has a large variance. The estimation from AGQ and Laplace methods are around ten times smaller than MCMC and AVB results while PQL produces even smaller estimations. This is consistent with our finding in the simulation study that AGQ and Laplace methods tends to underestimate the random effect variance. The AVB and MCMC estimations are considered more accurate; as the estimated variance is comparatively smaller than MCMC, the AVB estimation can be believed to be more accurate.

6. CONCLUSIONS

We present Adversarial Variational Bayes in solving Bayesian Tweedie mixed models. To empower the posterior approximation, we employ implicit models and training in an adversarial setting. To reduce variance of the gradients, we sum over the local latent variable of Tweedie model that is not parameterizable. We also make the prior distribution trainable to guarantee the expressiveness of the posterior. Our method outperforms all traditional methods on both simulated data and real-world data; meanwhile, the estimations on the variance parameter and the random effects are competitively accurate and more stable. Tweedie model is one fundamental model that is widely used in insurance policy pricing. As far as we are concerned, this is the first work that leverages recent progresses of deep learning methods to tackle insurance problems.

7. REFERENCES

- [1] Bent Jørgensen, *The theory of dispersion models*, CRC Press, 1997.
- [2] Bent Jørgensen, “Exponential dispersion models,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 127–162, 1987.
- [3] Umut Simsekli, Ali Taylan Cemgil, and Yusuf Kenan Yilmaz, “Learning the beta-divergence in tweedie compound poisson matrix factorization models,” in *ICML (3)*, 2013, pp. 1409–1417.
- [4] Bent Jørgensen, *Exponential dispersion models*, Number 48. Instituto de matematica pura e aplicada, 1991.
- [5] Gordon K Smyth and Bent Jørgensen, “Fitting tweedie’s compound poisson model to insurance claims data: dispersion modelling,” *Astin Bulletin*, vol. 32, no. 01, pp. 143–157, 2002.
- [6] Robert WM Wedderburn, “Quasi-likelihood functions, generalized linear models, and the gaussnewton method,” *Biometrika*, vol. 61, no. 3, pp. 439–447, 1974.
- [7] Marie Davidian and Raymond J Carroll, “Variance function estimation,” *Journal of the American Statistical Association*, vol. 82, no. 400, pp. 1079–1091, 1987.
- [8] John A Nelder and Daryl Pregibon, “An extended quasi-likelihood function,” *Biometrika*, vol. 74, no. 2, pp. 221–232, 1987.
- [9] DR Cox and N Reid, “A note on the calculation of adjusted profile likelihood,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 467–471, 1993.
- [10] Peter K Dunn and Gordon K Smyth, “Series evaluation of tweedie exponential dispersion model densities,” *Statistics and Computing*, vol. 15, no. 4, pp. 267–280, 2005.
- [11] Peter K Dunn and Gordon K Smyth, “Evaluation of tweedie exponential dispersion model densities by fourier inversion,” *Statistics and Computing*, vol. 18, no. 1, pp. 73–86, 2008.
- [12] Yi Yang, Wei Qian, and Hui Zou, “A boosted tweedie compound poisson model for insurance premium,” *arXiv preprint arXiv:1508.06378*, 2015.
- [13] Tianqi Chen and Carlos Guestrin, “Xgboost: A scalable tree boosting system,” *arXiv preprint arXiv:1603.02754*, 2016.
- [14] Wei Qian, Yi Yang, and Hui Zou, “Tweedies compound poisson model with grouped elastic net,” *Journal of Computational and Graphical Statistics*, vol. 25, no. 2, pp. 606–625, 2016.
- [15] Luke Tierney and Joseph B Kadane, “Accurate approximations for posterior moments and marginal densities,” *Journal of the american statistical association*, vol. 81, no. 393, pp. 82–86, 1986.
- [16] Qing Liu and Donald A Pierce, “A note on gausshermite quadrature,” *Biometrika*, vol. 81, no. 3, pp. 624–629, 1994.
- [17] Yanwei Zhang, “Likelihood-based and bayesian methods for tweedie compound poisson linear mixed models,” *Statistics and Computing*, vol. 23, no. 6, pp. 743–757, 2013.
- [18] Dani Gamerman and Hedibert F Lopes, *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*, CRC Press, 2006.
- [19] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul, “An introduction to variational methods for graphical models,” *Machine learning*, vol. 37, no. 2, pp. 183–233, 1999.
- [20] Christopher M Bishop, *Pattern recognition and machine learning*, springer, 2006.
- [21] Ronald J Williams, “Simple statistical gradient-following algorithms for connectionist reinforcement learning,” *Machine learning*, vol. 8, no. 3–4, pp. 229–256, 1992.
- [22] Geoffrey E Hinton, Peter Dayan, Brendan J Frey, and Radford M Neal, “The” wake-sleep” algorithm for unsupervised neural networks,” *Science*, vol. 268, no. 5214, pp. 1158, 1995.
- [23] Andriy Mnih and Karol Gregor, “Neural variational inference and learning in belief networks,” *arXiv preprint arXiv:1402.0030*, 2014.
- [24] Michalis Titsias RC AUEB and Miguel Lázaro-Gredilla, “Local expectation gradients for black box variational inference,” in *Advances in Neural Information Processing Systems*, 2015, pp. 2638–2646.
- [25] Ferenc Huszár, “Variational inference using implicit distributions,” *arXiv preprint arXiv:1702.08235*, 2017.
- [26] Lars Mescheder, Sebastian Nowozin, and Andreas Geiger, “Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks,” *arXiv preprint arXiv:1701.04722*, 2017.
- [27] Diederik P Kingma and Max Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
- [28] Alp Kucukelbir, Dustin Tran, Rajesh Ranganath, Andrew Gelman, and David M Blei, “Automatic differentiation variational inference,” *arXiv preprint arXiv:1603.00788*, 2016.
- [29] Peter McCullagh, “Generalized linear models,” *European Journal of Operational Research*, vol. 16, no. 3, pp. 285–292, 1984.
- [30] Yi Yang, Wei Qian, and Hui Zou, “Insurance premium prediction via gradient tree-boosted tweedie compound poisson models,” *Journal of Business & Economic Statistics*, pp. 1–15, 2017.
- [31] Karen CH Yip and Kelvin KW Yau, “On modeling claim frequency data in general insurance with extra zeros,” *Insurance: Mathematics and Economics*, vol. 36, no. 2, pp. 153–163, 2005.
- [32] HN De Silva, AJ Hall, DS Tustin, and PW Gandar, “Analysis of distribution of root length density of apple trees on different dwarfing rootstocks,” *Annals of Botany*, vol. 83, no. 4, pp. 335–345, 1999.
- [33] Edward W Frees, Glenn Meyers, and A David Cummings, “Summarizing insurance scores using a gini index,” *Journal of the American Statistical Association*, vol. 106, no. 495, pp. 1085–1098, 2011.
- [34] Edward W Jed Frees, Glenn Meyers, and A David Cummings, “Insurance ratemaking and a gini index,” *Journal of Risk and Insurance*, vol. 81, no. 2, pp. 335–366, 2014.