

OVERDISPERSED PHOTON-LIMITED SPARSE SIGNAL RECOVERY USING NONCONVEX REGULARIZATION

Yu Lu and Roummel F. Marcia

Department of Applied Mathematics, University of California, Merced, Merced, CA, 95343 USA

ABSTRACT

This paper investigates the application of the ℓ_p quasi-norm, where $0 < p < 1$, in contexts characterized by photon-limited signals such as medical imaging and night vision. In these environments, low-photon count images have typically been modeled using Poisson statistics. In related algorithms, the ℓ_1 norm is commonly employed as a regularization method to promotes sparsity in the reconstruction. However, recent research suggests that using the ℓ_p quasi-norm may yield lower error results. In this paper, we investigate the use of negative binomial statistics, which are more general models than Poisson models, in conjunction with the ℓ_p quasi-norm for recovering sparse signals in low-photon count imaging settings.

Index Terms— Negative binomial distribution, low-count imaging, nonconvex optimization, ℓ_p quasi-norm.

1. INTRODUCTION

Photon-limited signal processing is integral to many fields, including medical imaging [1], night vision [2], epidemiology [3, 4], and network traffic analysis [5, 6]. The Poisson model is well-established for reconstructing discrete low-photon signals, while the negative binomial model has primarily been used in applications such as matrix factorization [7] and data regression [8]. Recently, researchers have begun extending the negative binomial model to reconstruct low-photon signals [9].

The negative binomial probability mass function is given by

$$P(y|r, \beta) = \binom{y+r-1}{y} (1-\beta)^y \beta^r,$$

which can be interpreted as a distribution that models y successful events occurring in a sequence of independent and identically distributed Bernoulli trials before the r -th failure event, with each trial having a probability of failure β . Here, r is commonly referred to as the dispersion parameter. The mean μ is given by $\mu = r(1-\beta)/\beta$. As the negative binomial distribution is a more general model, while the Poisson

distribution is a special case [10, 11], it is anticipated that the negative binomial model will yield lower error results than the Poisson model in general scenarios.

The ℓ_1 norm is widely used in both Poisson and negative binomial models, along with the application of the total variation norm [12]. Recent research has explored the use of the ℓ_p quasi-norm, where $0 < p < 1$ [13, 14, 15], and in particular in the Poisson model for the 1-D signal case [16], with results indicating potential improvements over existing statistical models in signal processing. Consequently, this paper aims to investigate the application of the ℓ_p quasi-norm within the context of the negative binomial model.

2. PROBLEM FORMULATION

In the scenario where the observational data y are independently and identically drawn from the negative binomial distribution, the observation model can be expressed as follows

$$y \sim \text{NB}(r, \beta).$$

Under the negative binomial assumption, the observation model can be reformulated as

$$y_i \sim \text{NB}\left(r, \frac{r}{r + (Af^*)_i}\right),$$

where $f^* \in \mathbb{R}_+^n$ represents the signal or image of interest, and $A \in \mathbb{R}_+^{m \times n}$ denotes a mapping that linearly projects the scene f^* onto a set of measurements Af^* . The expectation on the detector, denoted as μ , is equivalent to Af^* , while the actual observation is $y \in \mathbb{R}_+^m$ (see Fig. 1). Consequently, the probability of observing the data y is

$$P(y|f) = \prod_{i=1}^m \binom{r+y_i-1}{y_i} \left(\frac{r}{r+(Af)_i}\right)^r \left(\frac{(Af)_i}{r+(Af)_i}\right)^{y_i}. \quad (1)$$

It is clear that a higher probability corresponds to a smaller discrepancy between the estimation $\hat{f}$ and the true signal f^* . Therefore, the maximum likelihood principle can be used to optimize the probability of observing the vector y in Eq.(1) by minimizing the corresponding negative binomial log-likelihood function:

$$F(f) \equiv \sum_{i=1}^m (r+y_i) \log(r+(Af)_i) - y_i \log((Af)_i). \quad (2)$$

This work is partly supported by National Science Foundation grants DMS 1840265 and IIS 1741490.

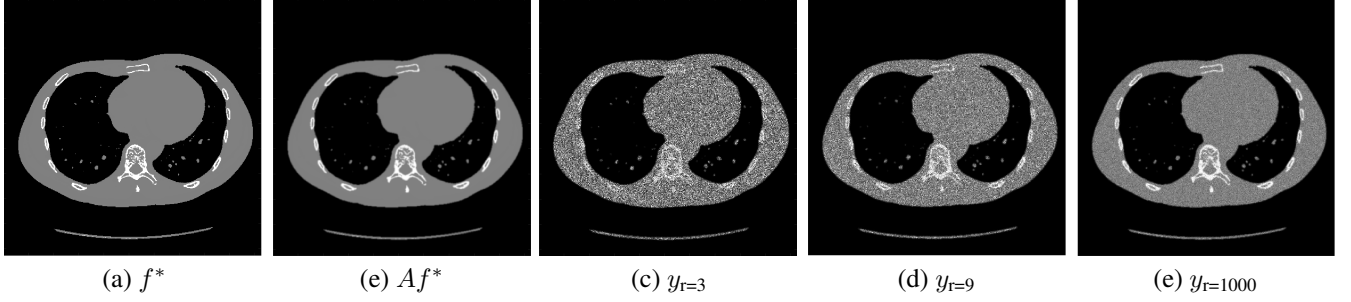


Fig. 1. Example of observation model. (a) The true image f^* . (b) The best observation we expect from detector. (c) Observed measurement $y_{r=3}$ drawn from a negative binomial (NB) distribution with dispersion parameter $r = 3$. (d) Observed measurement $y_{r=9}$ drawn from a NB distribution with $r = 9$. (e) Observed measurement y_{1000} drawn from a NB distribution with $r = 1000$.

Detailed application of this method to the negative binomial distribution can be found in [9].

We also exploit the structure of the signal to improve the accuracy of recovery. In particular, our image is sparse in some space. Consequently, a regularization term, $\text{pen}(f)$, is incorporated. This paper examines the application of the ℓ_p quasi-norm, where $0 < p < 1$, leading to the following formulation for the bounded negative binomial model used for sparse image recovery:

$$\begin{aligned} f^* &= \arg \min_{f \in \mathbb{R}_+^n} \Phi(f) \equiv F(f) + \tau \|f\|_p^p \\ \text{subject to } &0 \leq f, \end{aligned} \quad (3)$$

where $\tau > 0$ is a regularization parameter that balances the data-fidelity term $F(f)$ with the regularization term $\|f\|_p^p$.

In a more general scenario, f may not be sparse but is compressible in some orthonormal basis matrix W , such that $f = W\theta$, where θ is a sparse vector. The sparsity of f itself can be viewed as a special case, where the orthonormal basis matrix W is an identity matrix. Thus, a more generalized version of our problem can be formulated as

$$\begin{aligned} \theta^* &= \arg \min_{\theta \in \mathbb{R}^n} \Phi(\theta) \equiv F(W\theta) + \tau \|\theta\|_p^p \\ \text{subject to } &0 \leq W\theta \\ f^* &= W\theta^* \end{aligned} \quad (4)$$

3. ALGORITHM

The optimization problem (3) can be solved by a sequence of quadratic subproblems using the second-order Taylor series around the current iterate f^j . The derivatives of $F(f)$ are given by the following formulas:

$$\begin{aligned} \nabla F(f) &= \sum_{i=1}^m \left(\frac{r + y_i}{r + (Af)_i} - \frac{y_i}{(Af)_i} \right) A^\top e_i \\ \nabla^2 F(f) &= A^\top \left[\sum_{i=1}^m \left(\frac{r + y_i}{(r + (Af)_i)^2} - \frac{y_i}{(Af)_i^2} \right) e_i e_i^\top \right] A, \end{aligned}$$

where e_i is the i^{th} column of the $m \times m$ identity matrix.

We utilize the Barzilai-Borwein approach [17] to approximate the second derivative of $F(f)$ by a scalar multiple of the identity matrix, denoted as $\nabla^2 F(f) \approx \alpha_j I$, where $\alpha_j > 0$ is a scalar to obtain the quadratic approximation

$$F^j(f) = F(f^j) + (f - f^j)^\top \nabla F(f^j) + \frac{\alpha_j}{2} \|f - f^j\|_2^2.$$

With $f^j = W\theta^j$, the sequence of quadratic subproblems is thus given by

$$\begin{aligned} \theta^{j+1} &= \arg \min_{\theta \in \mathbb{R}^n} F^j(W\theta) + \tau \|\theta\|_p^p \\ \text{subject to } &0 \leq W\theta \\ f^{j+1} &= W\theta^{j+1} \end{aligned} \quad (5)$$

Letting $q^j = f^j - \frac{1}{\alpha_j} \nabla F(f^j)$, the equivalent quadratic subproblems and corresponding iterations are given by

$$\begin{aligned} \theta^{j+1} &= \arg \min_{\theta \in \mathbb{R}^n} \frac{1}{2} \|W\theta - q^j\|_2^2 + \frac{\tau}{\alpha_j} \|\theta\|_p^p \\ \text{subject to } &0 \leq W\theta^{j+1} \\ f^{j+1} &= W\theta^{j+1} \end{aligned} \quad (6)$$

To solve (6), we employ the following technique. First, since W is orthonormal, the objective function in (6) can be written equivalently as

$$\frac{1}{2} \|\theta - W^\top q^j\|_2^2 + \frac{\tau}{\alpha_j} \|\theta\|_p^p, \quad (7)$$

which is separable in θ , meaning it can be minimized separately into the scalar-valued problem

$$\theta_i^{j+1} = \arg \min_{\theta_i \in \mathbb{R}} \frac{1}{2} (\theta_i - (W^\top q^j)_i)^2 + \frac{\tau}{\alpha_j} |\theta_i|^p,$$

which can be accomplished using the Generalized Soft Thresholding (GST) technique [18]. After obtaining all components of θ^{j+1} , we compute $W\theta^{j+1}$ and threshold to ensure that f^{j+1} is feasible.

4. EXPERIMENTS

We conducted three separate experiments, each employing different noise level and realizations. Our investigations were directed towards the exploratio of different p values, the performance comparison of ℓ_p quasi-norm and ℓ_1 norm, and the comparative evaluation of the Poisson and negative binomial models when using the ℓ_p quasi-norm. The specific details of the three experiments are as follows:

- **Experiment 1:** Exploration of optimal solutions in 2D data, drawn from images corrupted by noise at various levels following a negative binomial distribution.
- **Experiment 2:** Comparative analysis of the Poisson and negative binomial models on 2D data extracted from images affected by noise at different levels following a negative binomial distribution.
- **Experiment 3:** Investigation of the performance of the ℓ_p quasi-norm and ℓ_1 norm on 2D data originating from images corrupted by noise at diverse levels following a negative binomial distribution.

All three experiments involved reconstructing a blurry, noisy image (from the MATLAB Medical Imaging Toolbox) with dimensions of 512 by 512 pixels. Three noise levels are presented for each experiment: $r = 3$, $r = 9$, and $r = 1000$, where r corresponds to the dispersion parameter. **Each experiment was conducted 10 times, where the average of these trials are presented.**

In these experiments, our proposed method is compared to three reference models: the Poisson model with an ℓ_1 norm [19], the Poisson model with an ℓ_p quasi-norm [20], and the negative binomial model with an ℓ_1 norm [9]. To measure the disparity between the computed solution $\hat{f}$ and the true solution f^* , we employ the percentage root-mean-square error ($\text{RMSE}(\%) = 100 * \|\hat{f} - f^*\|_2 / \|f^*\|_2$).

For the negative binomial models, the parameter r can be estimated using several methodologies, such as the method-of-moment [21] or the maximum quasi-likelihood methods [22]. Furthermore, cross-validation techniques can be used to estimate the dispersion parameter [23]. However, in our experiments, we deliberately used the exact value of the parameter r to circumvent potential biases and errors that might stem from the estimation of r .

4.1. Experiment I: p -value exploration

In our first experiment, we examined 19 different p values, ranging from 0.05 to 0.95 in increments of 0.05, using the negative binomial model with the ℓ_p quasi-norm. These values were applied to three separate images. As anticipated, a smaller dispersion parameter r resulted in noisier observations, leading to the highest RMSE values for the reconstructions with $r = 3$. Conversely, the lowest RMSEs correspond to the case where r is highest, namely $r = 1,000$. We note that the results are not monotonic as a function of p . This

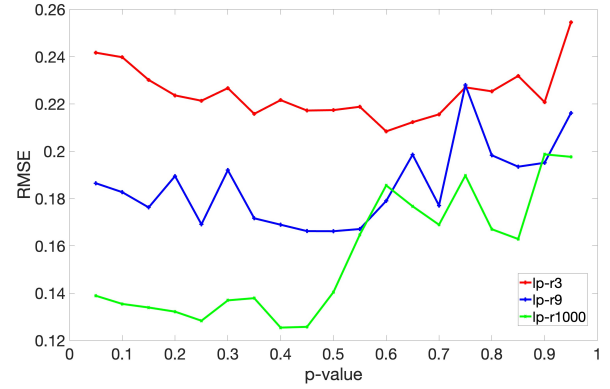


Fig. 2. Experiment I: RMSE of 2D data reconstruction using the negative binomial model with varying noisy level, i.e., $r = 3, 9$, and $1,000$ for different p values. Note that the lowest RMSE values generally occur around $p = 0.5$.

might be due to the nonconvexity of the problem and that the global minima are not achieved. Note however that our results are consistent with the observations in [13], which suggest that care must be employed when smaller of values of p is used.

4.2. Experiment II: Comparison between Poisson and negative binomial models with ℓ_p quasi-norm

In our second experiment, we juxtaposed the results obtained using the negative binomial and Poisson models at $r = 3, 9$ and $1,000$. Our findings indicate that for $r = 3$ and 9 , the negative binomial model consistently achieved lower RMSE values than the Poisson model (see Figs. 3(a) and (b)). Interestingly, we observed that as the value of r increased, the discrepancy between the two models progressively diminished. This observation is consistent with the theoretical understanding that the Poisson distribution is a special case of the negative binomial distribution when the dispersion parameter r tends towards infinity (see Fig. 3(c)).

4.3. Experiment III: Comparison between ℓ_p quasi-norm and ℓ_1 norm

In our third experiment, we undertook a comparative study between the ℓ_p quasi-norm and the ℓ_1 norm across three distinct images. Our findings suggest that, at the p values that yield the smallest RMSEs, the ℓ_p quasi-norm yields lower RMSE values compared to the ℓ_1 norm. However, it is noteworthy that the RMSE values pertaining to the ℓ_p quasi-norm consistently are lower than those of the ℓ_1 norm. We also observe that employing the ℓ_p quasi-norm narrows the gap between the error between the negative binomial and Poisson models (see Table 1 and Fig. 4).

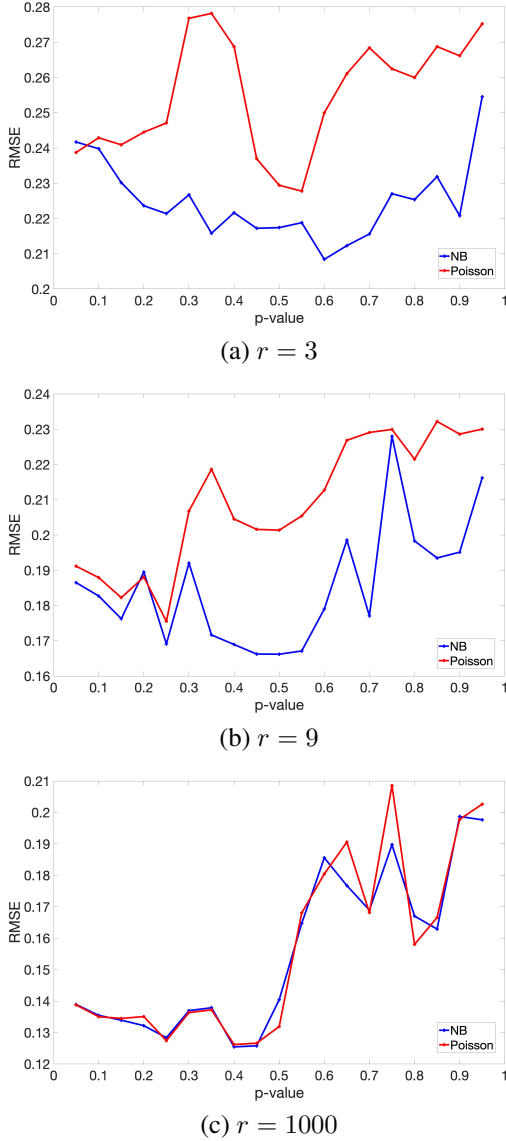


Fig. 3. Experiment II: RMSE of 2D data reconstructions using the negative binomial (NB) and Poisson models under various noise levels (e.g., for dispersion parameters $r = 3, 9$, and $1,000$). Note that for $r = 3$ and $r = 9$, the negative binomial models produce lower RMSEs. For the large dispersion parameter $r = 1,000$, the observed data is similar to Poisson measurements; consequently, the reconstruction RMSEs are similar for both negative binomial and Poisson models.

5. CONCLUSION

We investigated the use of the ℓ_p quasi-norm for signal reconstruction where the measurements are drawn from a negative binomial distribution. In this work, we contrasted the efficacy of the ℓ_p quasi-norm with the conventional ℓ_1 norm. In addition, we compared the reconstructions from using the neg-

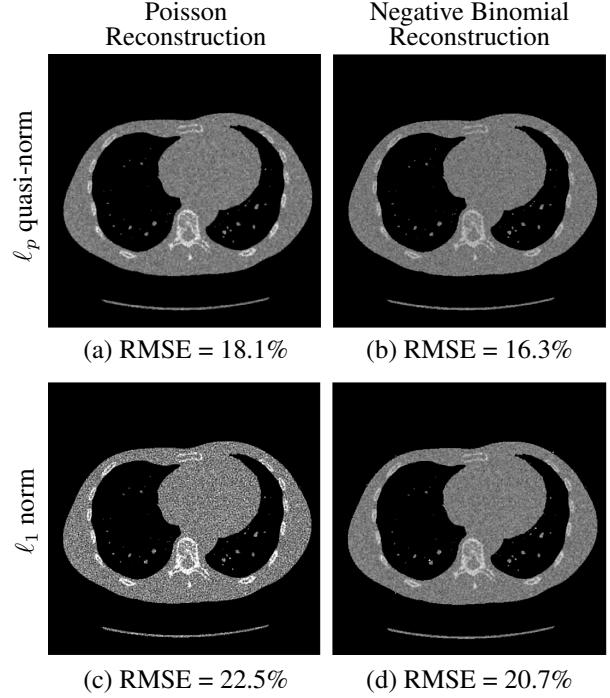


Fig. 4. Experiment III: 2D data drawn from a negative binomial distribution with $r = 9$. Note that the RMSE from using the negative binomial model is lower than that from the Poisson model when using an ℓ_1 norm. Similarly, when applying the ℓ_p quasi-norm, the RMSE of the negative binomial model is lower than the Poisson model. In both cases, the ℓ_p quasi-norm delivers superior results to the ℓ_1 norm.

Root-Mean-Square Error Table

r	NB(ℓ_p)	Poisson(ℓ_p)	NB(ℓ_1)
3	20.84%	22.77%	25.07%
9	16.62%	17.55%	20.95%
1,000	12.53%	12.62%	17.28%

Table 1. Experiment III: Root-mean-square error (RMSE) for three different dispersion parameters ($r = 3, 9$, and $1,000$) and three different approaches: negative binomial model with an ℓ_p regularization (NB(ℓ_p)), Poisson model with an ℓ_p quasi-norm (Poisson(ℓ_p)), and negative binomial model with an ℓ_1 norm (NB(ℓ_1)). The parameter p is chosen for each approach that results in the smallest RMSE value.

ative binomial and Poisson models. Our results show that the negative binomial model consistently outperforms the Poisson model in terms of reconstruction accuracy. Interestingly, in some instances, using a Poisson model with an ℓ_p quasi-norm regularization term yields lower RMSE values than a negative binomial model with an ℓ_1 norm regularization term, which suggests that the ℓ_p quasi-norm can help overcome deficiencies in statistical models and improve reconstruction outcomes.

6. REFERENCES

- [1] R. M. Willett and R. D. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 332–350, 2003.
- [2] C. Li, X. Qu, A. Gnanasambandam, O. A. Elgendy, J. Ma, and S. H. Chan, "Photon-limited object detection using non-local feature matching and knowledge distillation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3976–3987.
- [3] E. L. Frome and H. Checkoway, "Use of Poisson regression models in estimating incidence rates and ratios," *American Journal of Epidemiology*, vol. 121, no. 2, pp. 309–323, 1985.
- [4] G. Zou, "A modified Poisson regression approach to prospective studies with binary data," *American Journal of Epidemiology*, vol. 159, no. 7, pp. 702–706, 2004.
- [5] S. B. Slimane and T. Le-Ngoc, "A doubly stochastic Poisson model for self-similar traffic," in *Proceedings IEEE International Conference on Communications ICC '95*, 1995, vol. 1, pp. 456–460 vol.1.
- [6] E. del Arco, E. Morgado, M. I. Chidean, J. Ramiro-Bargueno, I. Mora-Jiménez, and A. J. Caamano, "Sparse vehicular sensor networks for traffic dynamics reconstruction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2826–2837, 2015.
- [7] O. Gouvert, T. Oberlin, and C. Févotte, "Negative binomial matrix factorization," *IEEE Signal Processing Letters*, vol. 27, pp. 815–819, 2020.
- [8] P. D. Allison and R. P. Waterman, "7. fixed-effects negative binomial regression models," *Sociological methodology*, vol. 32, no. 1, pp. 247–265, 2002.
- [9] Y. Lu and R. F. Marcia, "Negative binomial optimization for low-count overdispersed sparse signal reconstruction," in *Accepted*, 2023.
- [10] C. G. Taborda, F. Pérez-Cruz, and D. Guo, "New information-estimation results for Poisson, binomial and negative binomial models," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 2207–2211.
- [11] D. Guo, "On information-estimation relationships over binomial and negative binomial models," in *2013 IEEE International Symposium on Information Theory*. IEEE, 2013, pp. 459–463.
- [12] R. M. Willett, Z. T. Harmany, and R. F. Marcia, "Poisson image reconstruction with total variation regularization," in *2010 IEEE International Conference on Image Processing*. IEEE, 2010, pp. 4177–4180.
- [13] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.
- [14] J. K. Pant, W.-S. Lu, and A. Antoniou, "New improved algorithms for compressive sensing based on ℓ_p norm," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 61, no. 3, pp. 198–202, 2014.
- [15] P. Bohra and M. Unser, "Continuous-domain signal reconstruction using $l_{\{p\}}$ -norm regularization," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4543–4554, 2020.
- [16] L. Adhikari and R. F. Marcia, "Nonconvex relaxation for Poisson intensity reconstruction," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2015, pp. 1483–1487.
- [17] J. Barzilai and J. M. Borwein, "Two-Point Step Size Gradient Methods," *IMA Journal of Numerical Analysis*, vol. 8, no. 1, pp. 141–148, 01 1988.
- [18] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 217–224.
- [19] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson Intensity Reconstruction ALgorithms:Theory and Practice," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1084–1096, 2012.
- [20] A. Orkusyan, L. Adhikari, J. Valenzuela, and R. F. Marcia, "Analysis of p-norm regularized subproblem minimization for sparse photon-limited image recovery," in *2016 IEEE international conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 1407–1411.
- [21] S. J. Clark and J. N. Perry, "Estimation of the negative binomial parameter κ by maximum quasi-likelihood," *Biometrics*, pp. 309–316, 1989.
- [22] W. W. Piegorsch, "Maximum likelihood estimation for the negative binomial dispersion parameter," *Biometrics*, pp. 863–867, 1990.
- [23] C. Gu, "Cross-validating non-Gaussian data," *Journal of Computational and Graphical Statistics*, vol. 1, no. 2, pp. 169–179, 1992.