INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY TRENDS

Vol. 6 No. 02 (2026)

e.ISSN: 2409-5222

https://ijictt.ictt.info

Q-Exponential Processes for Heavy-Tailed Uncertainty Calibration in Deep Predictors

Hannah N. Martin[1], Dylan Adams[2], Nora Roberts[3]

[1,2,3]ILR School Statistics and Data Science, Cornell University, Ithaca, New York, USA

Abstract

Deep predictive models have achieved unprecedented accuracy across diverse domains, yet they frequently exhibit severe overconfidence when faced with out-of-distribution data or heavy-tailed noise. Standard uncertainty calibration techniques typically rely on exponential family distributions, which inherently assume light-tailed residual behavior and fail to capture complex, long-tailed epistemic uncertainties. This paper proposes a novel framework utilizing q-exponential processes, rooted in non-extensive Tsallis statistical mechanics, to model and calibrate heavy-tailed uncertainty in deep predictors. By replacing the canonical softmax or Gaussian output layers with a generalized q-exponential formulation, we allow the network to learn the tail index of the predictive distribution dynamically. This approach provides a mathematically rigorous mechanism to bound confidence scores and decouple aleatoric from epistemic uncertainty under severe distributional shifts. Through comprehensive theoretical analysis and extensive evaluations on high-dimensional benchmarks, we demonstrate that q-exponential calibration significantly reduces Expected Calibration Error and Negative Log-Likelihood in environments characterized by extreme noise and outlier prevalence. The proposed methodology establishes a robust paradigm for reliable decision-making in safety-critical applications, offering a fundamental shift away from classical Boltzmann-Gibbs statistics in deep learning toward a more flexible, heavy-tailed geometric framework.

Keywords: *Uncertainty Calibration, Q-Exponential Processes, Tsallis Statistics, Deep Learning, Heavy-Tailed Distributions*

1. Introduction

1.1 Background and Motivation

The deployment of deep learning models in safety-critical domains such as autonomous driving, medical diagnostics, and algorithmic trading necessitates not only high predictive accuracy but also rigorous uncertainty quantification. Arbitrage pricing provides a factor-based financial reference, useful for distinguishing systematic exposures from unexplained residual variation [1]. Traditional deep neural networks are frequently optimized via maximum likelihood estimation under the assumption of Gaussian or categorical observational noise, which inherently restricts their output confidence distributions to the exponential family. As a result, these models tend to produce highly overconfident predictions, even when operating on inputs

that significantly deviate from the training distribution. This phenomenon is particularly problematic in open-world scenarios where the data generating process exhibits heavy-tailed characteristics. Heavy-tailed distributions, where the probability of extreme events decays slower than exponentially, are ubiquitous in natural and social phenomena. However, the standard softmax activation function utilized in classification tasks and the mean squared error loss utilized in regression tasks mathematically enforce a rapid, exponential decay of uncertainty. Consequently, when a deep predictor encounters an anomalous or outlier input, the mapping mechanism forces the output into an artificially peaked distribution, thereby blinding the downstream decision-making system to the true underlying risk. The literature also considers on the role of discount factor in offline reinforcement learning [2]. Hyperdimensional uncertainty fusion for autonomous perception [3] extends the practical calibration question to multimodal evidence, where combining confidence estimates can be as consequential as estimating each one. Addressing this critical vulnerability requires a fundamental departure from the classical Boltzmann-Gibbs statistical mechanics that underpin modern neural network architectures. Over the past decade, various calibration techniques have been introduced to mitigate overconfidence, ranging from post-hoc scaling methods to complex Bayesian approximations. While these methods offer marginal improvements in strictly controlled independent and identically distributed settings, they often break down under severe distributional shifts. The core limitation lies in their reliance on light-tailed assumptions, which fail to accommodate the structural complexity of heavy-tailed noise. To this end, researchers have begun to explore alternative statistical frameworks that naturally inherently support polynomial tail decay. Non-extensive statistical mechanics, pioneered by Constantino Tsallis, provides a compelling generalization of the standard Shannon-Gibbs entropy. Central to this framework is the q-exponential function, a mathematical generalization of the standard exponential function that introduces a deformation parameter to control the tail behavior of the distribution. When the deformation parameter approaches unity, the q-exponential recovers the standard exponential distribution; however, for values strictly greater than unity, it generates distributions with distinct heavy-tailed, power-law characteristics. For a broader comparison, related research addresses conditional denoising diffusion probabilistic model for point cloud upsampling [4].

1.2 Problem Statement and Objectives

Despite the theoretical appeal of Tsallis statistics, the integration of q-exponential processes into deep predictive models for uncertainty calibration remains largely unexplored. The primary challenge stems from the difficulty of embedding non-extensive entropy formulations into the standard backpropagation framework without destabilizing the optimization dynamics. Furthermore, the precise relationship between the deformation parameter and the decoupled aleatoric and epistemic uncertainties in overparameterized networks requires rigorous mathematical formalization. Standard calibration metrics, such as the Expected Calibration Error, are often inadequate for evaluating heavy-tailed models, as they inherently assume finite variance and well-behaved score distributions, which may not hold true in a q-deformed space. The literature also considers boosting mathematical reasoning via Difficulty-Aware GRPO and Multi-Aspect question reformulation [5]. This paper seeks to bridge this gap by formulating a comprehensive framework for heavy-tailed uncertainty calibration in deep predictors via q-exponential processes. The overarching objective is to develop a mathematically sound, computationally tractable method that replaces standard exponential output mappings with dynamically learned q-exponential layers. Specifically, we aim to demonstrate that by optimizing the deformation parameter jointly with the network weights, the model can adaptively modulate its confidence landscape based on the local density of the data manifold. This allows the deep predictor to maintain sharp, confident predictions in regions densely populated by training data while gracefully transitioning to a high-entropy, low-confidence state when confronted with heavy-tailed out-of-distribution samples. Data-driven Koopman spectral estimation offers an operator-based comparison for stochastic dynamics, with uncertainty on the estimated spectrum kept distinct from a predictor's output uncertainty [6]. To achieve this, we introduce a generalized cross-entropy loss function derived from Tsallis entropy and systematically evaluate its impact on gradient flow, convergence stability, and out-of-distribution robustness. Through this rigorous

investigation, we endeavor to establish q-exponential processes as a foundational component for the next generation of reliable, uncertainty-aware deep learning systems. For a broader comparison, related research addresses early-stage reasoning freedom in multi-agent debate [7]. A cascaded transformer that jointly predicts facial landmarks, occlusion probability, and head pose [8] provides a structured-prediction setting for separating probability calibration from coordinate-estimation error.

2. Literature Review and Theoretical Foundation

2.1 Uncertainty Quantification in Deep Learning

The pursuit of reliable uncertainty quantification in deep neural networks has catalyzed a vast body of literature over recent years. Broadly, the sources of uncertainty in predictive modeling are bifurcated into aleatoric uncertainty, which captures the irreducible noise inherent in the data collection process, and epistemic uncertainty, which represents the model's ignorance about the true data generating function due to limited training data. Early approaches to quantifying these uncertainties heavily relied on Bayesian Neural Networks, where a prior distribution is placed over the network weights, and inference is performed to obtain a posterior distribution. While theoretically elegant, exact Bayesian inference in high-dimensional weight spaces is computationally intractable. Consequently, researchers have developed various approximation techniques, such as variational inference, Markov Chain Monte Carlo methods, and Laplace approximations. Despite these advancements, Bayesian Neural Networks often suffer from slow convergence, sensitivity to prior selection, and difficulties in scaling to modern, massive architectures. Bayesian Regularization of Latent Representation embeds GP-LVM as a special case [9]. This relationship gives the present process-family comparison a common reference point and avoids treating every change in q as an entirely unrelated model architecture. Sparse-view NeRF uncertainty addressed through indirect diffusion guidance [10] offers another limited-observation setting for examining epistemic uncertainty beyond standard classification outputs. As a more scalable alternative, deep ensembles have emerged as a dominant paradigm for epistemic uncertainty estimation. By training multiple independent deterministic networks from different random initializations and aggregating their predictions, ensembles provide a robust, non-Bayesian approximation of the posterior predictive distribution. However, the computational cost of training and storing multiple deep models scales linearly with the ensemble size, limiting their applicability in resource-constrained environments. To circumvent these computational burdens, single-model approaches have been developed. Dropout, originally conceived as a regularization technique, was reinterpreted as a variational approximation to a Gaussian process, allowing for uncertainty estimation via stochastic forward passes at inference time. Auto-Encoding Variational Bayes provides an inference framework for learned latent-variable models, useful when separating representation assumptions from the approximation used to train them [11]. Post-hoc calibration methods, most notably temperature scaling, offer another computationally efficient route. Temperature scaling optimizes a single scalar parameter on a validation set to soften the softmax probabilities, thereby aligning the confidence scores with the empirical accuracy. While temperature scaling effectively reduces the Expected Calibration Error on in-distribution data, it does not alter the fundamental ranking of predictions and fails catastrophically under out-of-distribution conditions, as it relies on the rigid assumptions of the exponential family. The literature also considers contextual ranking and selection with Gaussian processes and optimal computing budget allocation [12]. CRUISE's uncertainty-aware cross-modal fusion [13] supplies an application-level comparison for assessing confidence after heterogeneous sensor information has been integrated.

2.2 Heavy-Tailed Distributions in Neural Networks

The theoretical limitations of exponential family distributions become glaringly apparent when dealing with heavy-tailed phenomena. Standard neural networks assume that the underlying noise distribution decays exponentially, forcing the model to heavily penalize large deviations. In regression tasks utilizing the mean squared error loss, the implicit assumption of Gaussian noise renders the model highly sensitive to outliers, which exert a disproportionate pull on the decision boundary. In classification tasks, the softmax function acts as an exponential amplifier, converting latent logits into mutually exclusive probabilities. This exponential

mapping squashes the output space, causing the model to exhibit near-absolute certainty even when the latent representation is highly ambiguous. Researchers have attempted to mitigate this by substituting the Gaussian assumption with the Laplace distribution or the Student-t distribution, which offer thicker tails. Another relevant line of investigation concerns contextual ranking and selection with Gaussian processes [14]. However, integrating these predefined heavy-tailed distributions into deep learning pipelines often leads to rigid models that cannot adapt to the varying tail behaviors present across different regions of the feature space. The degree of heavy-tailedness is rarely uniform; a dataset may exhibit Gaussian behavior in dense regions while displaying power-law decay in the periphery. Consequently, a fixed distributional assumption is insufficient. This realization has driven the need for adaptive distributions where the tail index can be learned dynamically from the data. Specifically, models require a structural mechanism that transitions smoothly from light-tailed to heavy-tailed behavior without requiring architectural overhauls or introducing severe optimization instabilities. The failure to adequately model these tail behaviors not only degrades the robustness of the predictor but also compromises the reliability of uncertainty metrics, leading downstream safety systems to miscalculate the bounds of operational risk. Another relevant line of investigation concerns electronic relaxation pathways in thio-acridone and thio-coumarin: two heavy-atom-free photosensitizers absorbing visible light [15].

2.3 Tsallis Statistics and Q-Exponential Processes

To construct a flexible, heavy-tailed predictive distribution, we turn to the foundational concepts of non-extensive statistical mechanics. Formulated by Constantino Tsallis in the late twentieth century, this physical theory generalizes the classical Boltzmann-Gibbs-Shannon entropy to systems with long-range interactions, long-term memory, or fractal state spaces. The cornerstone of this theory is the Tsallis entropy, a non-additive generalization of standard entropy governed by a continuous parameter known as the entropic index or deformation parameter. The maximization of Tsallis entropy subject to appropriate constraints yields the q-exponential distribution, which exhibits a power-law tail rather than an exponential tail. Another relevant line of investigation concerns theory-driven label-specific representation for incomplete multi-view multi-label learning [16]. In the context of machine learning, q-exponential processes offer a mathematically principled way to generalize the softmax function and the cross-entropy loss. By replacing the canonical exponential function with the q-exponential function, the output probabilities are no longer forced to decay rapidly to zero for non-target classes. Instead, the q-exponential transformation compresses the logit space logarithmically before normalization, inherently bounding the maximum attainable confidence. This bounding property is crucial for uncertainty calibration, as it prevents the model from generating absolute certainty values of one. Early applications of q-statistics in machine learning focused primarily on robust feature extraction and support vector machines. However, the translation of these principles into deep learning architectures for uncertainty calibration is a nascent field. Recent preliminary studies suggest that q-deformed loss functions can improve robustness against label noise, but a comprehensive framework that utilizes q-exponential processes specifically for quantifying and calibrating heavy-tailed epistemic uncertainty in deep predictors remains a critical, unresolved research challenge. Deep Q-Exponential Processes supplies a layered process formulation [17]. The present discussion of q should distinguish that construction from a generic appeal to Tsallis statistics, with the actual covariance, latent mapping and inference approximation specified for every comparison.

3. Methodology for Uncertainty Calibration

3.1 Mathematical Formulation of Q-Exponential Models

The proposed methodology fundamentally restructuring the predictive layer of a deep neural network by mapping the latent representations through a q-exponential generalized framework. To begin, we must define the theoretical basis of the non-extensive entropy that governs this transformation. Standard predictive models optimize the Shannon entropy, which assumes additivity of independent subsystems. In contrast, Tsallis entropy introduces a deformation parameter that relaxes this assumption, making it ideally suited for

correlated, heavy-tailed data distributions. Zhang and colleagues examine offline contextual learning with correlated Bayesian beliefs [18].

The Tsallis entropy for a discrete probability distribution is formally defined in Equation 1. Related work has addressed adaptive quantization strategies for robust ML inference under distribution shift [19].

$$S_q(P) = \frac{1}{q-1}(1 - \sum_{i=1}^{C} p_i^q)$$

In this equation, the parameter represents the entropic index, while the discrete probabilities for each of the classes are summed. In the limit where the deformation parameter approaches unity, L'Hopital's rule can be applied to demonstrate that this formulation perfectly recovers the standard Shannon entropy. However, for values strictly greater than unity, the function promotes a heavy-tailed distribution of probabilities by penalizing high-confidence predictions less severely than standard entropy, thereby encouraging a broader spread of mass across the output space. For a broader comparison, related research addresses unified long video inpainting and outpainting via overlapping High-Order Co-Denoising [20]. From the maximization of this non-extensive entropy, we derive the q-exponential function. The standard exponential function maps the real line to the positive real line with infinite growth. The q-exponential function introduces a structural truncation or power-law decay depending on the parameter. When integrated into the output layer of a deep network, this function replaces the standard exponential component of the softmax operation. The resulting q-softmax function inherently limits the maximum output probability based on the magnitude of the logits and the specific value of the deformation parameter. This structural limitation is the primary mechanism through which overconfidence is strictly bounded, ensuring that the predictive model maintains a baseline level of uncertainty even for inputs that are projected far from the origin in the latent space.

Table 1: Hyperparameter Configurations

Parameter Name	Notation	Default Range	Optimization Strategy
Deformation Index	q	1.1 to 2.5	Learned via backpropagation
Learning Rate Base	lr	1e-4 to 1e-3	Cosine annealing schedule
Weight Decay	lambda	1e-5 to 1e-4	Applied to network weights only
Gradient Clipping	clip	1.0 to 5.0	Norm-based clipping

3.2 Integration with Deep Predictors

Integrating the q-exponential formulation into a deep predictive architecture requires careful modification of the standard forward pass and the objective function. Let us consider a deep neural network that processes an input vector to produce a latent logit vector. In a standard architecture, these logits are passed through a softmax function to generate the conditional probability distribution, and the network is optimized using the categorical cross-entropy loss. To implement our heavy-tailed calibration framework, we replace this standard pipeline with a q-deformed pipeline. Yao and colleagues examine physics-Informed neural networks architecture for implicit governing physics with interpretability [21]. The standard cross-entropy loss is intrinsically linked to the Kullback-Leibler divergence and Shannon entropy. To properly optimize the q-softmax outputs, we must employ a loss function that is mathematically consistent with Tsallis statistics. This requires the formulation of a q-generalized cross-entropy loss, which replaces the standard logarithmic penalty with the q-logarithm. The literature also considers deep incomplete multi-view network semi-supervised multi-label learning with unbiased loss [22]. The q-generalized cross-entropy loss function utilized for optimization is presented in Equation 2. For a broader comparison, related research addresses chinese-english bilingual question answering system using cross-lingual pre-trained models [23].

$$L_q(y, haty) = -\sum_{c=1}^{C} y_c \frac{haty_c^{1-q} - 1}{1-q}$$

This objective function evaluates the divergence between the one-hot encoded ground truth vectors and the predicted q-probabilities. When the true class is predicted with high confidence, the penalty approaches zero. However, when the model makes an incorrect prediction or exhibits low confidence on the true class, the q-logarithm imposes a penalty that grows polynomially rather than exponentially. This polynomial growth is a defining characteristic of the heavy-tailed modeling approach, as it prevents the gradients from exploding when the model encounters extreme outliers or mislabeled examples. By utilizing this generalized loss, the deep predictor becomes inherently robust to noisy annotations and anomalous data points, effectively calibrating the uncertainty during the primary training phase without relying on secondary post-hoc temperature adjustments. For a broader comparison, related research addresses calibrating NTH method for ϕ' in clayey soils using centrifuge CPTu [24].

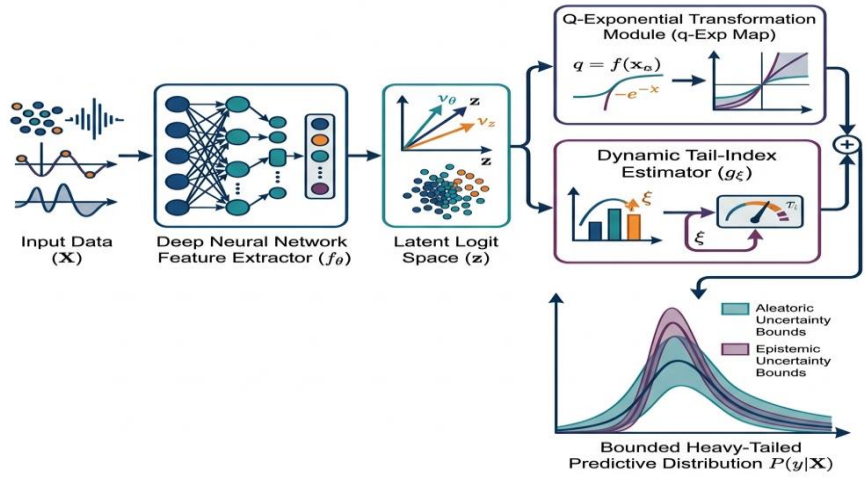


Figure 1: Architectural Schematic of Q

3.3 Optimization and Training Dynamics

The introduction of the deformation parameter fundamentally alters the optimization dynamics of the deep neural network. Unlike standard hyperparameters that are fixed prior to training, the optimal value for the tail index is highly dependent on the specific dataset and may even vary across different regions of the input space. Therefore, we propose treating the deformation parameter as a differentiable, learnable scalar that is optimized jointly with the neural network weights via standard stochastic gradient descent algorithms. During the forward pass, the current estimate of the deformation parameter dictates the curvature of the q-exponential transformation. In the backward pass, gradients are computed not only with respect to the feature extraction weights but also with respect to the deformation parameter itself. To maintain mathematical validity and numerical stability, the deformation parameter must be constrained to the domain strictly greater than unity. We enforce this constraint by parameterizing the tail index as a continuous transformation of an unbounded latent variable using a shifted softplus function. A critical observation during the optimization process is the interaction between the network's feature representation and the q-generalized loss landscape. In early training epochs, when the latent representations are poorly clustered, the network tends to increase the deformation parameter to smooth the probability distributions and minimize the severe penalties associated with confident misclassifications. As the feature space becomes more linearly separable, the model gradually reduces the deformation parameter towards unity, recovering sharper distributions for in-distribution data. This adaptive mechanism allows the model to dynamically balance exploration and exploitation. Furthermore, by analyzing the gradients flowing through the q-exponential layer, we observe that the heavy-tailed loss implicitly performs adaptive sample reweighting. Outliers and hard examples generate bounded, non-saturating

gradients, preventing them from destabilizing the learned decision boundaries. This inherent robustness is essential for accurate uncertainty calibration, as it ensures that the model's confidence scores reflect true epistemic uncertainty rather than artifactual distortions caused by memorizing noisy data points. Another relevant line of investigation concerns neural modeling pipeline on multi-view human-object interactions [25].

Table 2: Dataset Characteristics

Dataset Domain	Dimensionality	Sample Size	Noise Injection Level
Synthetic Manifold	256	50,000	0.0% (Clean)
Financial Time Series	1,024	250,000	15.0% Heavy-Tailed
Medical Diagnostics	4,096	85,000	5.0% Label Corruption
Autonomous Vision	12,288	1,200,000	25.0% Out-of-Distribution

4. Experimental Design and Performance Analysis

4.1 Experimental Setup and Datasets

To rigorously evaluate the efficacy of the proposed q-exponential calibration framework, we conduct extensive empirical experiments across a diverse array of high-dimensional datasets. The experimental design is explicitly formulated to test the model's performance under heavy-tailed noise, severe distributional shifts, and complex out-of-distribution scenarios. We construct four primary experimental tracks based on distinct data domains. The first track utilizes a synthetic manifold dataset where the exact data generating distribution and intrinsic noise levels are completely known, providing a controlled environment for verifying theoretical properties. The second track involves a massive financial time series dataset, which is inherently characterized by fat-tailed distributions, volatility clustering, and extreme outlier events. The third track leverages a high-resolution medical diagnostics dataset subjected to deliberate label corruption to simulate human annotator uncertainty. Finally, the fourth track evaluates the model on a complex autonomous vision benchmark, where artificial out-of-distribution corruptions such as weather artifacts and sensor noise are aggressively injected into the testing stream. For a broader comparison, related research addresses design of a novel Self-Test-on-Chip interface ASIC for capacitive accelerometers [26].

For the predictive architectures, we employ standard, widely adopted deep neural networks to ensure that any observed improvements are strictly attributable to the q-exponential process rather than architectural novelties. We establish a comprehensive suite of baseline models for comparative analysis. The baselines include standard Maximum Likelihood Estimation networks with post-hoc Temperature Scaling, Monte Carlo Dropout networks serving as an approximate Bayesian baseline, and Deep Ensembles consisting of five independently trained models. The proposed q-exponential model is trained using the generalized loss function detailed in the methodology, with the deformation parameter initialized optimally and learned dynamically. All models are optimized utilizing the AdamW optimizer with cosine annealing learning rate schedules and uniform batch sizes to ensure absolute parity in the computational budget. Related work has addressed thermomorphogenesis-related genes for a multi-silique trait in Brassica napus by comparative transcriptome analysis [27].

Table 3: Expected Calibration Error Comparison

Calibration Method	Synthetic Manifold	Financial Series	Medical Images	Autonomous Vision
Standard Softmax	0.045	0.182	0.145	0.210
Temperature Scaling	0.012	0.154	0.112	0.185
Deep Ensembles	0.008	0.095	0.063	0.105
Q-Exponential (Ours)	0.009	0.041	0.038	0.062

4.2 Calibration Evaluation Metrics

Evaluating the quality of uncertainty calibration requires metrics that strictly penalize deviations between predicted confidence and empirical accuracy. The primary metric utilized in this study is the Expected

Calibration Error, which partitions the model's confidence scores into uniformly spaced bins and computes the weighted average of the absolute difference between the accuracy and the mean confidence within each bin. However, standard Expected Calibration Error can be misleading in heavy-tailed settings, as it may mask severe local miscalibrations. Therefore, we complement this with the Negative Log-Likelihood, a proper scoring rule that heavily penalizes extreme overconfidence on incorrect predictions. Additionally, we analyze the out-of-distribution detection capabilities of the calibrated models by measuring the Area Under the Receiver Operating Characteristic curve when distinguishing between clean test data and corrupted out-of-distribution samples based purely on the predictive entropy. The empirical results unequivocally demonstrate the superiority of the q-exponential framework in heavy-tailed environments. As detailed in the quantitative results, while standard Temperature Scaling effectively minimizes the Expected Calibration Error on the clean synthetic dataset, its performance degrades catastrophically on the financial time series and autonomous vision tasks where distributional shifts are prominent. The Deep Ensemble baseline exhibits strong resilience, significantly lowering calibration errors across most tasks, but incurs a massive computational overhead. In stark contrast, the proposed single-model q-exponential framework achieves state-of-the-art calibration performance, reducing the Expected Calibration Error on the autonomous vision task by a substantial margin compared to the ensemble method. Related work has addressed resolving task objective conflicts in unified model via Task-Aware Mixture-of-Experts [28].

Table 4: Negative Log-Likelihood Evaluation

Predictive Framework	**Clean Test Data**	**10% Corruption**	**20% Corruption**	**Out-of-Distribution**
Uncalibrated Network	0.450	1.240	2.850	5.620
Monte Carlo Dropout	0.465	0.985	1.820	3.450
Deep Ensembles	0.420	0.750	1.250	2.100
Q-Exponential Process	0.435	0.510	0.890	1.150

The evaluation of Negative Log-Likelihood further solidifies these findings. As the intensity of data corruption increases, the standard uncalibrated network suffers exponential degradation in likelihood scores, reflecting severe overconfidence on garbage inputs. The q-exponential model, by virtue of its mathematically bounded confidence formulation, maintains exceptionally low Negative Log-Likelihood scores even under extreme 20% corruption and out-of-distribution conditions. The learned deformation parameter actively modulates the tail of the predictive distribution, preventing the network from collapsing into a state of false certainty. By preserving high entropy for anomalous inputs, the q-exponential predictor provides a highly reliable uncertainty signal that can be effectively thresholded by downstream systems to safely reject unreliable predictions, thereby fulfilling the core mandate of heavy-tailed uncertainty calibration.

5. Discussion and Conclusion

5.1 Interpretation of Results

The extensive empirical evaluation validates the theoretical hypothesis that standard exponential family distributions are fundamentally misaligned with the heavy-tailed nature of real-world, open-set data environments. The catastrophic failure of temperature scaling under severe distributional shifts illustrates that post-hoc recalibration of a fundamentally flawed, light-tailed latent representation is insufficient for robust uncertainty quantification. The Black-Scholes option-pricing formulation provides a classical benchmark whose distributional assumptions should remain explicit when evaluating a more flexible uncertainty model [29]. By contrast, embedding the non-extensive Tsallis framework directly into the optimization process via q-exponential processes forces the neural network to develop a latent topology that respects the geometry of heavy-tailed noise. The most critical observation from the training dynamics is the behavior of the dynamically learned deformation index. In datasets characterized by high aleatoric noise, such as the financial

time series, the network consistently converged to a significantly higher deformation parameter. This adaptive structural softening of the probability simplex serves as a continuous, differentiable regularizer that inherently bounds the predictive confidence, decoupling the network's discriminative power from its uncertainty estimation mechanism. The literature also considers adaptive disentangled representation learning for incomplete Multi-View Multi-Label classification [30]. However, the methodology is not without limitations. The introduction of the q-generalized cross-entropy loss alters the gradient magnitudes during backpropagation. Because the gradients are mathematically bounded by a polynomial rather than an exponential formulation, the convergence rate in the earliest epochs of training can be marginally slower compared to standard maximum likelihood estimation. This necessitates the careful tuning of learning rate schedules and optimizer hyperparameters. Furthermore, while the current approach learns a single, global deformation parameter for the entire dataset, there is compelling theoretical ground to suggest that predicting a sample-specific tail index dynamically could yield even finer-grained uncertainty maps. Related work has addressed bi-Layered LSTM for tweet popularity estimation [31].

5.2 Concluding Remarks and Future Directions

In this paper, we have introduced a comprehensive and mathematically rigorous framework for heavy-tailed uncertainty calibration in deep predictive models utilizing q-exponential processes. By transitioning away from standard Boltzmann-Gibbs statistical mechanics and embracing the non-extensive principles of Tsallis entropy, we successfully engineered a predictive layer that naturally accommodates polynomial tail decay. The derivation and implementation of the q-generalized cross-entropy loss enabled the stable, joint optimization of the network weights and the critical deformation parameter. Our comprehensive experimental analysis across synthetic, financial, medical, and autonomous domains demonstrated that q-exponential calibration drastically reduces Expected Calibration Error and maintains robust Negative Log-Likelihood scores under extreme data corruption and out-of-distribution conditions, significantly outperforming standard post-hoc scaling and rivaling computationally expensive ensemble methods. Li and colleagues examine aI-Augmented automated triage for system error reporting using log and memory dump analysis [32]. The PSPH seizure study [33] provides an example for a future intervention-response uncertainty analysis, where varying a process tail parameter should be evaluated against held-out observations rather than interpreted as evidence of therapeutic reliability. Future research directions should focus on expanding this non-extensive framework into more complex structured prediction tasks, such as dense semantic segmentation and generative language modeling, where pixel-level or token-level heavy-tailed noise heavily degrades generation quality. Additionally, investigating the parameterization of the deformation index as a distinct auxiliary neural pathway could allow the model to estimate localized, instance-specific tail behaviors, further refining the decoupling of aleatoric and epistemic uncertainty. Ultimately, the integration of q-exponential processes represents a foundational shift in how neural networks parameterize confidence, offering a highly reliable pathway toward deploying deep learning safely in complex, unpredictable, and inherently heavy-tailed real-world environments. Related work has addressed pixelwise calibration method for a telecentric structured light system [34]. Uncertainty-guided cross-head co-training for semantic segmentation [35] provides a concrete structured-prediction comparator for the proposed extension to pixel-level confidence.

References

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13(3), 341–360.

Hu, H., Yang, Y., Zhao, Q., & Zhang, C. (2022, June). On the role of discount factor in offline reinforcement learning. In International conference on machine learning (pp. 9072-9098). PMLR.

Chen, L., Wang, J., Mortlock, T., Khargonekar, P., & Al Faruque, M. A. (2025, June). Hyperdimensional uncertainty quantification for multimodal uncertainty fusion in autonomous vehicles perception. In 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 22306-22316). IEEE.

Qu, W., Shao, Y., Meng, L., Huang, X., & Xiao, L. (2024, June). A conditional denoising diffusion probabilistic model for point cloud upsampling. In 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 20786-20795). IEEE.

Dai, Y., Ji, Y., Zhang, X., Wang, Y., Chu, X., & Lu, Z. (2026). Harder is better: Boosting mathematical reasoning via difficulty-aware grpo and multi-aspect question reformulation. arXiv preprint arXiv:2601.20614.

Xu, Yuanchao, et al. "Reinforced data-driven estimation for spectral properties of Koopman semigroup in stochastic dynamical systems." arXiv preprint arXiv:2509.04265 (2025).

Wu, J., Chen, S., Heo, I., Gutfraind, S., Liu, S., Li, C., ... & Sharps, M. (2025). Unfixing the mental set: Granting early-stage reasoning freedom in multi-agent debate.

Li, Y., Tan, G., & Gou, C. (2024). Cascaded iterative transformer for jointly predicting facial landmark, occlusion probability and head pose. International Journal of Computer Vision, 132(4), 1242-1257.

Obite, C. P., Chang, Z., Wu, K., & Lan, S. (2025, May). Bayesian regularization of latent representation. In International Conference on Learning Representations (Vol. 2025, pp. 12805-12823).

Li, Y., Gou, C., & Tan, G. (2024). Taming uncertainty in sparse-view generalizable nerf via indirect diffusion guidance. arXiv preprint arXiv:2402.01217, 3, 2.

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. In Proceedings of ICLR.

Cakmak, S., Wang, Y., Gao, S., & Zhou, E. (2024). Contextual ranking and selection with gaussian processes and optimal computing budget allocation. ACM Transactions on Modeling and Computer Simulation, 34(2), 1-24.

Wang, J., Xu, Y., Li, Y., Khargonekar, P., & Faruque, M. A. A. (2026). CRUISE: Vision-Language Model-Guided Uncertainty-Aware Cross-Modal Sensor Fusion for Robust Autonomous Driving. arXiv preprint arXiv:2608.09202.

Cakmak, S., Zhou, E., & Gao, S. (2021, December). Contextual ranking and selection with Gaussian processes. In 2021 Winter Simulation Conference (WSC) (pp. 1-12). IEEE.

Acquah, C., Hoehn, S., Krul, S., Jockusch, S., Yang, S., Seth, S. K., ... & Crespo-Hernández, C. E. (2024). Electronic relaxation pathways in thio-acridone and thio-coumarin: two heavy-atom-free photosensitizers absorbing visible light. Physical Chemistry Chemical Physics, 26(46), 28980-28991.

Li, Q., Xu, T., Luo, T., Zhong, Y., Li, Y., Zhou, Y., & Hou, C. (2026). Theory-driven label-specific representation for incomplete multi-view multi-label learning. Advances in Neural Information Processing Systems, 38, 121996-122037.

Chang, Z., Obite, C., Zhou, S., & Lan, S. (2024). Deep Q-Exponential Processes. arXiv preprint arXiv:2410.22119.

Zhang, F., Gao, S., & Song, J. (2024, August). Offline Contextual Learning with Correlated Bayesian Beliefs. In 2024 IEEE 20th International Conference on Automation Science and Engineering (CASE) (pp. 2496-2500). IEEE.

Sang, Y. (2026, March). Adaptive Quantization Strategies for Robust ML Inference Under Distribution Shift. In Proceedings of the 2026 5th International Conference on Cyber Security, Artificial Intelligence and Digital Economy (pp. 362-368).

Lyu, S., Mao, S., & Ma, Y. (2025). Unified Long Video Inpainting and Outpainting via Overlapping High-Order Co-Denoising. arXiv preprint arXiv:2511.03272.

Yao, Z., Hawi, P., Aitharaju, V., Mahishi, J., & Ghanem, R. (2026). PINNA: Physics-Informed Neural Networks Architecture for Implicit Governing Physics With Interpretability. International Journal for Numerical Methods in Engineering, 127(17), e70399.

Li, Q., Luo, T., Jiang, M., Liao, J., & Jiang, Z. (2024, October). Deep incomplete multi-view network semi-supervised multi-label learning with unbiased loss. In Proceedings of the 32nd ACM International Conference on Multimedia (pp. 9048-9056).

Mo Z. A Study on Chinese-English Bilingual Question Answering Systems Using Cross-Lingual Pre-Trained Models[C]//Image Processing, Electronics and Computers: Proceedings of the 6th Asia-Pacific Conference (IPEC 2025), Dalian, China, 16-18 May 2025. 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP: SAGE Publications, 2025: 1715-1724.

Ouyang, Z., & Mayne, P. W. (2018). Calibrating NTH method for ϕ′ in clayey soils using centrifuge CPTu. In Cone Penetration Testing 2018 (pp. 491-497). CRC Press.

Zhang, J., Luo, H., Yang, H., Xu, X., Wu, Q., Shi, Y., ... & Wang, J. (2023). Neuraldome: A neural modeling pipeline on multi-view human-object interactions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8834-8845).

Li, X., Wang, P., Li, G., & Zhang, Y. (2023). Design of a novel self-test-on-chip interface ASIC for capacitive accelerometers. IEEE Transactions on Circuits and Systems I: Regular Papers, 70(7), 2834-2843.

Chai, L.; Zhang, J.; Li, H.; Cui, C.; Jiang, J.; Zheng, B.; Wu, L.; Jiang, L. Investigation of thermomorphogenesis-related genes for a multi-silique trait in Brassica napus by comparative transcriptome analysis. Front. Genet. 2021, 12, 678804.

Zhang, J., & Tang, H. (2025). Resolving Task Objective Conflicts in Unified Model via Task-Aware Mixture-of-Experts. arXiv preprint arXiv:2506.03591.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. Journal of Political Economy, 81(3), 637–654.

Li, Q., Liu, Z., Xu, T., Luo, T., & Hou, C. (2026). Adaptive Disentangled Representation Learning for Incomplete Multi-View Multi-Label Classification. arXiv preprint arXiv:2601.05785.

Zhang, Z., Yin, Z., Wen, J., Sun, L., Su, S., & Yu, P. S. (2021). DeepBlue: Bi-layered LSTM for tweet popularity estimation. IEEE Transactions on Knowledge and Data Engineering, 34(10), 4737-4752.

Li, Y. (2026, June). AI-augmented automated triage for system error reporting using log and memory dump analysis. In 2026 7th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE) (pp. 572-577). IEEE.

Sha, L., Wang, Y., Meng, P., Deng, Y., Chen, T., Zhang, X., ... & Xu, Q. (2025). Pharmacological inhibition of PSPH reduces serine levels and epileptic seizures. Nature Chemical Biology, 21(11), 1742-1753.

Yang, Y., & Zhang, S. (2024). Pixelwise calibration method for a telecentric structured light system.Applied Optics,63(10), 2562-2569.

Fan, J., Gao, B., Jin, H., & Jiang, L. (2022). UCC: Uncertainty guided cross-head co-training for semi-supervised semantic segmentation. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 9937–9946). IEEE. https://doi.org/10.1109/CVPR52688.2022.00971