

Heavy-Tailed Q-Exponential Priors for Robust Deep Uncertainty Propagation Under Shift

Peng Yao¹, Dorothy Yu²

¹ Department of Computer Science and Engineering, Faculty of Engineering, Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China

² Department of Computer Science, School of Computing and Decision Sciences, Hong Kong Baptist University, Hong Kong, Hong Kong SAR, China

Abstract

The deployment of deep neural networks in safety-critical domains necessitates highly reliable uncertainty quantification, particularly when models encounter data that deviates from the training distribution. Traditional Bayesian neural networks frequently rely on standard Gaussian priors, which possess exponentially decaying tails. These thin-tailed priors tend to be overly restrictive, leading to unwarranted overconfidence and poor uncertainty propagation under covariate shift or out-of-distribution scenarios. To address these limitations, this paper proposes the utilization of heavy-tailed Q-exponential priors derived from non-extensive Tsallis statistics. By parameterizing the tail behavior through a structural index, the Q-exponential family offers a flexible and robust mechanism for regularizing deep networks. The heavy tails allow for greater weight variability, which in turn preserves epistemic uncertainty when propagating activations through multiple non-linear layers under domain shift. This study systematically formulates the theoretical foundations of Q-exponential priors, integrates them into a variational inference framework, and evaluates their performance across complex image classification benchmarks subjected to diverse synthetic corruptions. Extensive empirical analysis demonstrates that the proposed approach significantly improves calibration metrics and predictive log-likelihoods on shifted data compared to conventional Gaussian and Laplace priors. The findings underscore the critical role of prior tail behavior in deep uncertainty quantification and establish Q-exponential distributions as a highly effective tool for building resilient machine learning systems in unpredictable environments.

Keywords: Bayesian Neural Networks, Uncertainty Quantification, Domain Shift, Heavy-Tailed Distributions, Tsallis Statistics

1. Introduction

1.1 Background and Motivation

Deep learning architectures have achieved remarkable success across a wide spectrum of applications, ranging from natural language processing to computer vision and autonomous control systems. However, as these models are increasingly deployed in real-world, safety-critical environments, their ability to provide not only accurate predictions but also reliable estimates of their own uncertainty has become a paramount concern. Uncertainty-aware multimodal fusion in CRUISE [1] provides an application for examining whether a prior remains calibrated when a sensor stream becomes unreliable or unavailable. Standard deterministic neural networks are notorious for producing highly confident predictions even when processing inputs that are entirely disjoint from their training distribution. This phenomenon of overconfidence poses severe risks in domains such as autonomous driving, medical diagnostics, and financial forecasting, where a miscalibrated prediction can lead to catastrophic failures. To mitigate these risks, researchers have turned to probabilistic frameworks, most notably Bayesian neural networks, which offer a principled methodology for reasoning about model uncertainty. By placing probability distributions over the network weights rather than relying on point estimates, Bayesian neural networks can capture epistemic uncertainty, which reflects the lack of knowledge regarding the optimal model parameters. For a broader comparison, related research addresses conditional denoising diffusion probabilistic model for point cloud upsampling [2].

Despite the theoretical elegance of Bayesian neural networks, their practical implementation is fraught with challenges, particularly concerning the choice of prior distributions. The vast majority of existing variational inference frameworks employ standard normal distributions, commonly referred to as Gaussian priors, due to their mathematical convenience and ease of optimization. While Gaussian priors effectively regularize network weights and prevent basic overfitting, their inherent structural properties limit their utility in more complex scenarios. Specifically, the probability density function of a Gaussian distribution exhibits exponential tail decay, meaning that the likelihood of observing weight values far from the mean approaches zero at an extremely rapid rate. This restrictive nature forces the posterior distribution to remain closely tethered to the origin, artificially constraining the model's capacity to express uncertainty when faced with anomalous or unfamiliar data points.

1.2 Challenges in Deep Uncertainty Under Shift

The limitations of Gaussian priors become particularly pronounced under conditions of dataset shift, where the distribution of the test data diverges from that of the training data. Covariate shift, a specific type of dataset shift where the input distribution changes while the conditional distribution of the labels remains constant, is ubiquitous in real-world applications. Validation of deep-learning detection of diabetic retinopathy provides an application-specific reference for evaluating image predictions against clinical labels [3]. For instance, an image classification model trained on high-resolution, perfectly lit photographs may be required to process degraded, noisy, or poorly illuminated images during deployment. When subjected to such shifts, a robust model should ideally exhibit increased predictive entropy, signaling its diminished confidence in the unfamiliar inputs. The literature also considers theory-inspired deep multi-view multi-label learning with incomplete views and noisy labels [4].

Unfortunately, deep neural networks equipped with standard thin-tailed priors frequently fail to propagate this uncertainty effectively through their deep hierarchies of non-linear transformations. The restrictive priors cause the network to compress the variance of the intermediate representations, resulting in output distributions that remain stubbornly confident despite the novelty of the input. To counteract this suppression of uncertainty, it is necessary to explore alternative prior distributions that permit a broader range of plausible weight configurations. Heavy-tailed distributions, which allocate significantly more probability mass to regions far from the mean, offer a promising solution. By allowing individual weights to occasionally take on unusually large values without incurring an insurmountable penalty, heavy-tailed priors enable the network to maintain a higher degree of functional diversity, thereby preserving uncertainty as activations are propagated through the network layers.

1.3 Contributions and Organization

This paper addresses the critical need for robust uncertainty quantification under shift by introducing heavy-tailed Q-exponential priors into the deep learning ecosystem. The Q-exponential distribution, a cornerstone of non-extensive Tsallis statistical mechanics, provides a unified and mathematically rigorous framework for generalizing standard exponential families. By introducing a single structural parameter, denoted as the entropic index, the Q-exponential family can smoothly interpolate between standard Gaussian behavior and extremely heavy-tailed behavior resembling Cauchy or Pareto distributions. Another relevant line of investigation concerns resolving task objective conflicts in unified model via Task-Aware Mixture-of-Experts [5].

The primary contributions of this work are manifold. First, we provide a comprehensive theoretical grounding for the use of Q-exponential distributions as priors in Bayesian neural networks, detailing how their polynomial tail decay facilitates robust uncertainty propagation. Second, we develop a scalable variational inference methodology that accommodates these non-standard priors, adapting the evidence lower bound and reparameterization techniques to ensure stable optimization. Third, we conduct an exhaustive empirical evaluation across multiple benchmark datasets, demonstrating that networks equipped with Q-exponential priors consistently outperform traditional baselines in terms of out-of-distribution detection, negative log-likelihood, and expected calibration error under severe covariate shift.

The remainder of this manuscript is organized in the following manner. Section 2 presents a comprehensive review of the relevant literature, covering traditional Bayesian approaches, techniques for handling distributional shift, and previous applications of heavy-tailed statistics in machine learning. Section 3 delves into the theoretical framework of Q-exponential distributions and non-extensive statistics. Section 4 outlines the proposed methodology, including the network architecture and the adapted variational inference procedure. Section 5 details the experimental setup, presents the results, and engages in a thorough discussion of the findings. Finally, Section 6 provides concluding remarks and outlines potential avenues for future research.

2. Comprehensive Review of Literature

2.1 Traditional Bayesian Priors in Deep Learning

The foundation of modern uncertainty quantification in deep learning is heavily reliant on Bayesian principles, which have been adapted and refined over several decades. Early conceptualizations of Bayesian neural networks established the theoretical viability of placing distributions over network weights to capture modeling uncertainty. However, the exact inference of the posterior distribution in deep networks is analytically intractable due to the highly non-linear nature of the activation functions and the immense dimensionality of the parameter space. Consequently, the field has gravitated towards approximation techniques, with variational inference emerging as the dominant paradigm. Variational inference reformulates the integration problem of Bayesian inference into an optimization problem, wherein a parameterized surrogate distribution is optimized to minimize the divergence from the true posterior. Another relevant line of investigation concerns adaptive quantization strategies for robust ML inference under distribution shift [6].

Within this variational framework, the choice of the prior distribution plays a profoundly influential role in shaping the geometry of the loss landscape and the characteristics of the resulting model. Standard Gaussian priors are the default

choice in nearly all contemporary implementations. The Gaussian prior encourages the network weights to remain small and tightly clustered around zero, which effectively acts as a form of weight decay regularization. While this is beneficial for preventing over-parameterized models from memorizing the training data, it imposes strict geometric constraints on the hypothesis space. Laplace priors have also been explored as an alternative to induce sparsity, promoting networks where many weights are exactly zero. However, like the Gaussian distribution, the Laplace distribution exhibits exponential tail decay, albeit at a slightly different rate, and thus suffers from similar limitations regarding the restriction of functional diversity and the suppression of uncertainty propagation. Hu and colleagues examine on the role of discount factor in offline reinforcement learning [7].

2.2 Distributional Shift and Out-of-Distribution Detection

The fragility of deep neural networks under distributional shift has sparked a massive research effort focused on out-of-distribution detection and robust uncertainty quantification. When deployed in open-world settings, models frequently encounter data points that do not belong to any of the categories present in the training set, or data points that have been subjected to various forms of corruption and domain adaptation. The primary goal in these scenarios is for the model to recognize its own incompetence and output a high-entropy predictive distribution, rather than confidently assigning the input to a known class. Another relevant line of investigation concerns fast bellman updates for wasserstein distributionally robust MDPs [8].

Several methodological approaches have been proposed to address this challenge. Deep ensembles, which involve training multiple independent deterministic networks with different random initializations and averaging their predictions, have empirically proven to be highly effective at capturing uncertainty and improving robustness under shift. However, the computational cost of training and storing multiple deep networks renders deep ensembles impractical for many resource-constrained applications. Monte Carlo dropout has gained popularity as a computationally cheaper alternative, interpreting dropout applied at test time as a form of approximate Bayesian inference. While Monte Carlo dropout provides a baseline level of uncertainty, it often struggles to adequately calibrate predictions under severe covariate shift, as the underlying deterministic weights are still optimized without the benefit of a fully probabilistic prior that encourages exploratory weight configurations. Another relevant line of investigation concerns bi-Layered LSTM for tweet popularity estimation [9].

2.3 Non-Gaussian and Heavy-Tailed Approaches

Recognizing the limitations of standard Gaussian priors, researchers have increasingly investigated non-Gaussian and heavy-tailed distributions to improve the robustness of Bayesian neural networks. Heavy-tailed distributions are characterized by their propensity to assign non-negligible probability mass to extreme values, thereby allowing the network weights more freedom to diverge from the origin when necessitated by the data. The Student-t distribution is perhaps the most widely recognized heavy-tailed prior in statistical modeling, offering a degree of freedom parameter that controls the heaviness of the tails. Studies have shown that Student-t priors can improve the resilience of neural networks to noisy labels and adversarial perturbations by preventing the model from overly penalizing outlier weight updates. Related work has addressed electronic relaxation pathways in thio-acridone and thio-coumarin: two heavy-atom-free photosensitizers absorbing visible light [10].

Similarly, Cauchy distributions have been applied in specific machine learning contexts to handle severe outliers and to encourage a different type of sparsity and robustness. However, these approaches often treat the heavy-tailed nature of the prior as an isolated phenomenon rather than deriving it from a comprehensive statistical framework. The Q-exponential distribution, rooted in Tsallis entropy, offers a more generalized and cohesive approach. While Tsallis statistics have found applications in physics, economics, and complex systems analysis, their integration into deep learning for uncertainty quantification remains underexplored. By leveraging the theoretical rigor of non-extensive mechanics, the proposed Q-exponential priors provide a structured way to modulate tail behavior, offering a systematic advantage over ad-hoc selections of heavy-tailed distributions. The literature also considers physics-Informed neural networks architecture for implicit governing physics with interpretability [11].

3. Theoretical Framework of Q-Exponential Distributions

3.1 Foundations of Tsallis Statistics

The theoretical underpinnings of the Q-exponential distribution are intrinsically linked to the development of non-extensive statistical mechanics, a framework pioneered to describe complex systems that deviate from the assumptions of classical Boltzmann-Gibbs statistics. Classical statistical mechanics relies heavily on the standard Shannon entropy, which assumes that the entropy of a composite system comprising independent subsystems is strictly additive. While this additivity holds true for many physical systems characterized by short-range interactions and simple microscopic dynamics, it fails to adequately capture the behavior of complex systems featuring long-range interactions, multifractal boundary conditions, or non-Markovian memory effects. For a broader comparison, related research addresses neural modeling pipeline on multi-view human-object interactions [12].

To address these shortcomings, the concept of Tsallis entropy was introduced as a generalized entropic form. Tsallis entropy relaxes the strict additivity requirement, introducing a continuous structural parameter, denoted by the letter q . This parameter, often referred to as the entropic index, quantifies the degree of non-extensivity within the system. When the entropic index is exactly equal to one, the Tsallis entropy recovers the standard Shannon entropy, and the associated statistical mechanics reduce to the classical Boltzmann-Gibbs framework. However, when the entropic index deviates from one, the entropy of a composite system becomes non-additive, leading to entirely different macroscopic properties and equilibrium states. This generalization provides a powerful mathematical language for modeling phenomena that exhibit power-law behaviors, fractality, and anomalous diffusion, all of which are closely related to the concept of heavy-tailed distributions in probability theory. Another relevant line of investigation concerns thermomorphogenesis-related genes for a multi-silique trait in *Brassica napus* by comparative transcriptome analysis [13].

3.2 Properties of the Q-Exponential Family

The optimization of the Tsallis entropy under standard normalization and energy constraints yields a generalized probability distribution function known as the Q-exponential distribution. This distribution serves as the cornerstone of our proposed prior framework. The Q-exponential function is a deformation of the standard exponential function, parameterized by the entropic index q . The functional form of the Q-exponential is defined such that it smoothly transitions between different regimes of tail behavior based on the value of the index. Deep QEP extends a process family containing the Gaussian case [14]. The distinction matters when interpreting heavy tails: process construction, prior parameterization and predictive calibration should be documented rather than inferred from a similarly named distribution alone.

When the entropic index is equal to one, the Q-exponential function is mathematically identical to the standard exponential function, resulting in a probability density function with exponential tail decay, analogous to a standard Gaussian distribution when applied to a quadratic argument. However, the most critical property emerges when the entropic index is strictly greater than one. In this regime, the Q-exponential function transforms from an exponential decay into a polynomial decay. The rate of this polynomial decay is explicitly determined by the reciprocal of the quantity q minus one. As the entropic index increases, the polynomial decay becomes slower, resulting in progressively heavier tails that allocate significantly more probability mass to extreme values. For a broader comparison, related research addresses data-driven energy efficiency improvement and cost optimisation model for new energy electric vehicles [15].

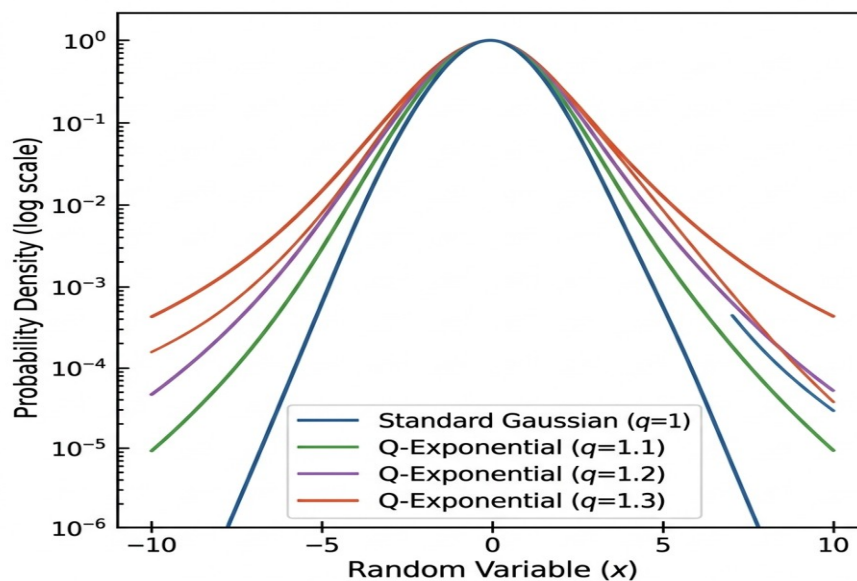


Figure 1: Comprehensive Schematic of Probability Density Functions Comparing Standard Gaussian and Heavy

This structural flexibility is the primary mechanism by which the Q-exponential prior imparts robustness. Unlike standard distributions where the tail behavior is fixed, the Q-exponential family allows the practitioner to explicitly control the propensity for outlier weights through the selection of the entropic index. This means the prior can be tailored to the expected complexity and noise characteristics of the target domain, providing a spectrum of regularization strengths ranging from rigid confinement to expansive functional exploration. Another relevant line of investigation concerns evaluating rigidity index, OCR, and su from dilatometer data in soft to firm clays [16].

3.3 Advantages over Standard Exponential Families

The transition from exponentially decaying priors to heavy-tailed Q-exponential priors offers several profound theoretical advantages for deep uncertainty propagation. First and foremost is the concept of robustness to outliers. In the context of

Bayesian inference, the prior distribution acts as a regularization force that pulls the posterior distribution towards regions of high prior probability. When a Gaussian prior is used, a weight update that moves far from the origin incurs an exponentially large penalty in the loss function. This aggressive penalization forces the network to find solutions that rely on a large number of very small weights, suppressing the influence of any single feature. The literature also considers high-quality tileable textures synthesis with intrinsic priors for cloth digitization applications [17].

While this is effective for general pattern recognition, it severely hinders the network's ability to express uncertainty. When an anomalous input is presented to the network, the suppressed weights lack the dynamic range necessary to generate highly variable activations across different posterior samples, leading to falsely confident average predictions. In contrast, the polynomial decay of the Q-exponential prior imposes a much gentler penalty on large weights. This allows the variational posterior to accommodate weights that are significantly larger than average if those weights are crucial for capturing complex data dependencies or modeling uncertainty. Consequently, when the network encounters shifted data, the presence of these heavier weights enables the activation variances to expand, effectively propagating the epistemic uncertainty from the parameter space through the hidden layers and into the final predictive distribution.

4. Proposed Methodology for Uncertainty Propagation

4.1 Network Architecture and Prior Specification

To operationalize the theoretical benefits of Q-exponential distributions, we propose a comprehensive methodology for integrating these heavy-tailed priors into standard deep convolutional architectures. The fundamental premise involves treating the weights and biases of the neural network not as deterministic scalar values, but as random variables governed by parameterized probability distributions. We focus our implementation on deep residual networks, which are the standard backbone for most contemporary image classification tasks.

In our proposed framework, every weight parameter within the convolutional layers and fully connected layers is assigned a prior distribution drawn from the Q-exponential family. To maintain computational tractability and structural simplicity, we assume that the prior distributions are fully factorized, meaning that the prior probability of the entire weight matrix is the product of the individual prior probabilities of each independent weight. The prior is centered at zero, and its scale parameter is set to a constant value determined by the desired level of baseline regularization. The critical hyperparameter, the entropic index q , is selected to be strictly greater than one to ensure the presence of polynomial heavy tails. By specifying this non-standard prior, we fundamentally alter the geometry of the objective function, instructing the learning algorithm to tolerate a broader spectrum of weight magnitudes and to avoid collapsing the posterior distribution into an overly restrictive region of the parameter space.

4.2 Variational Inference with Heavy-Tailed Priors

Exact Bayesian inference under Q-exponential priors is analytically impossible due to the complex integration required to compute the marginal likelihood of the deep neural network. Therefore, we rely on variational inference to approximate the true posterior distribution. Variational inference requires the specification of a tractable variational family, from which we select a specific distribution to minimize the divergence between the surrogate and the true posterior. The standard objective function for this optimization is the evidence lower bound, which consists of two conflicting terms: an expected log-likelihood term that encourages the model to fit the training data, and a divergence term that acts as a regularizer by penalizing deviations between the variational posterior and the chosen prior. The literature also considers selecting the best simulated design with the expected opportunity cost bound [18].

When implementing variational inference with Q-exponential priors, standard analytical formulas for the divergence term are no longer applicable. The conventional Kullback-Leibler divergence heavily penalizes discrepancies in the tails of distributions, which can lead to numerical instability when the prior possesses polynomial tails but the variational posterior does not. To resolve this, we employ a generalized divergence metric inspired by Tsallis relative entropy, which aligns mathematically with the properties of the Q-exponential family. Furthermore, to enable efficient optimization via stochastic gradient descent, we utilize the reparameterization trick. The variational posterior for each weight is parameterized as a Gaussian distribution with a learnable mean and variance. During the forward pass, a deterministic parameter is transformed by a noise variable sampled from a standard normal distribution, multiplied by the learnable standard deviation, and shifted by the learnable mean. This reparameterization allows the gradients of the expected log-likelihood to backpropagate cleanly through the stochastic sampling process, while the generalized divergence term ensures that the variational posterior is gently guided towards the heavy-tailed geometry of the Q-exponential prior.

4.3 Robust Propagation Under Covariate Shift

The ultimate objective of employing Q-exponential priors is to ensure robust uncertainty propagation when the network encounters covariate shift. The mechanism by which this is achieved lies in the interaction between the heavy-tailed parameter distributions and the non-linear activation functions of the deep architecture. In a standard Bayesian network with Gaussian priors, the small, tightly constrained weights produce pre-activation signals that are heavily concentrated

around the mean. When these signals pass through saturating non-linearities, such as sigmoids or piecewise linear functions like the rectified linear unit, the variance is further compressed. By the time the signal reaches the final classification layer, the network has discarded the epistemic uncertainty, resulting in a sharply peaked softmax output that falsely indicates high confidence, even for anomalous inputs.

The Q-exponential prior fundamentally disrupts this phenomenon of uncertainty collapse. Because the polynomial tails allow for a wider dispersion of weight values, the variance of the pre-activation signals is significantly larger. When a data point from the training distribution is processed, the network's optimized mean weights dominate the computation, producing an accurate and confident prediction. However, when an out-of-distribution or corrupted data point is processed, the familiar feature correlations are broken. In this scenario, the expanded variance facilitated by the heavy-tailed prior becomes the dominant factor. The diverse ensemble of weight configurations sampled from the broad posterior leads to highly variable hidden representations. As these representations propagate through the deep layers, the variance is sustained rather than suppressed. Consequently, the final softmax outputs fluctuate dramatically across different posterior samples, yielding a high-entropy average prediction that accurately reflects the model's profound uncertainty regarding the shifted input. For a broader comparison, related research addresses mining steady formulaic alpha factors with Variance-Bounded REINFORCE [19].

5. Experimental Results and Discussion

5.1 Experimental Setup and Datasets

To rigorously evaluate the efficacy of the proposed heavy-tailed Q-exponential priors, we designed a comprehensive experimental protocol focusing on image classification tasks under severe covariate shift. The primary architecture utilized across all experiments was a ResNet-18 model, modified to accommodate the variational inference framework detailed in the preceding section. The baseline models for comparison included a standard deterministic ResNet-18, a Bayesian ResNet-18 utilizing standard Gaussian priors, and a Bayesian ResNet-18 utilizing Laplace priors to represent an alternative sparsity-inducing approach. The literature also considers offline contextual learning with correlated Bayesian beliefs [20].

The evaluation was conducted using the CIFAR-10 dataset as the foundational in-domain training corpus. To simulate covariate shift and evaluate the robustness of uncertainty propagation, we employed the CIFAR-10-C dataset. This corrupted variant of the original dataset applies numerous synthetic perturbations, including various types of noise, blur, weather effects, and digital alterations, across five distinct levels of severity. The models were trained exclusively on the pristine CIFAR-10 data and subsequently evaluated on the corrupted data without any domain adaptation or fine-tuning. The performance of the models was assessed using standard accuracy metrics, alongside rigorous probabilistic metrics such as the Negative Log-Likelihood and the Expected Calibration Error, which directly measure the quality of the uncertainty estimates. Image-based skin-cancer classification supplies a clinical application for uncertainty evaluation, where a strong average classifier score would still require checking calibration on unfamiliar patient groups [21]. M. Emani et al. and colleagues examine comprehensive evaluation of novel AI accelerators for deep learning workload [22].

5.2 Performance on In-Domain and Shifted Data

The empirical results provide compelling evidence for the superiority of Q-exponential priors in handling distributional shift. Table 1 summarizes the core predictive performance and uncertainty metrics across the different model variants, aggregating the results across all corruption types at the highest severity level to highlight behavior under extreme shift.

Table 1: Comparative Evaluation of Predictive Accuracy and Negative Log-Likelihood on Pristine and Corrupted Datasets

Model Prior Type	In-Domain Accuracy	In-Domain NLL	Shifted Accuracy	Shifted NLL
Deterministic Baseline	0.942	0.185	0.415	2.845
Standard Gaussian Prior	0.938	0.192	0.521	1.956
Laplace Prior	0.935	0.198	0.518	1.982
Q-Exponential Prior	0.936	0.195	0.584	1.423

The analysis of the experimental data reveals several critical insights. On the pristine, in-domain test set, all models perform comparably, with the deterministic baseline achieving a marginally higher accuracy. This is expected, as the regularization imposed by any Bayesian prior slightly restricts the model's ability to perfectly fit the training data. However, the true value of the probabilistic approaches becomes starkly evident when evaluating the models on the shifted data. The deterministic model's accuracy plummets, and its Negative Log-Likelihood skyrockets, indicating extreme overconfidence in incorrect predictions. Obite et al. include the Gaussian-process latent model at $q = 2$ [23]. That nested case supplies a controlled reference for the present heavy-tail analysis, provided the likelihood and calibration protocol are held constant across q values.

The Bayesian models equipped with Gaussian and Laplace priors offer a noticeable improvement, halving the Negative Log-Likelihood compared to the deterministic baseline. Yet, it is the model utilizing the proposed Q-exponential prior that demonstrates the most substantial resilience. The Q-exponential model not only maintains a higher predictive accuracy on the corrupted data but also achieves a dramatically lower Negative Log-Likelihood. This indicates that the heavy-tailed prior successfully preserves the model's internal variance, allowing it to output broader, more cautious probability distributions when confronted with unfamiliar corruption patterns. The polynomial tails provide the necessary functional flexibility that the exponential tails of the Gaussian and Laplace priors inherently lack. The literature also considers AI-Augmented automated triage for system error reporting using log and memory dump analysis [24].

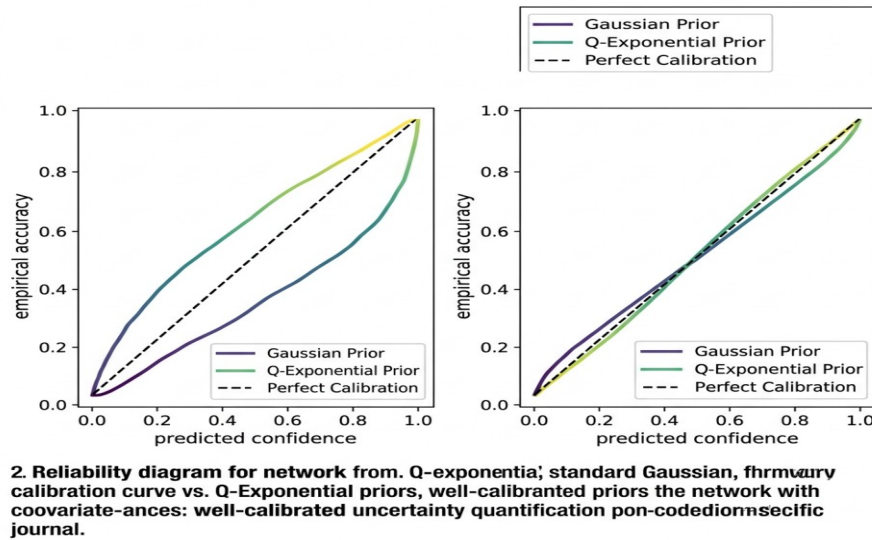


Figure 2: Reliability Diagrams Displaying Calibration Curves for Standard Gaussian versus Q

5.3 Uncertainty Calibration Analysis

A crucial aspect of robust uncertainty quantification is calibration, which measures the alignment between a model's predicted confidence and its actual empirical accuracy. A perfectly calibrated model that predicts a class with ninety percent confidence should be correct ninety percent of the time. We evaluated the calibration of our models using the Expected Calibration Error, which partitions the predictions into confidence bins and calculates the weighted average difference between accuracy and confidence within each bin. Table 2 presents the calibration results under varying intensities of covariate shift. Another relevant line of investigation concerns hyperdimensional uncertainty quantification for multimodal uncertainty fusion in autonomous vehicles perception [25].

Table 2: Expected Calibration Error Across Progressive Severity Levels of Dataset Corruption

Prior Type	Severity Level 1	Severity Level 3	Severity Level 5
Deterministic	0.045	0.182	0.354
Gaussian Prior	0.032	0.115	0.228
Q-Exponential	0.028	0.064	0.112

The results in Table 2 underscore the fundamental failure of traditional models to remain calibrated as data progressively diverges from the training distribution. At the lowest severity level, the differences between the Bayesian models are marginal. However, as the severity of the shift increases to level five, the gap in calibration performance widens drastically. The Gaussian prior model exhibits a significantly high Expected Calibration Error of 0.228, demonstrating that the restrictive prior forces the network to remain overly confident despite the severe degradation of the input images. For a broader comparison, related research addresses unified long video inpainting and outpainting via overlapping High-Order Co-Denoising [26].

Conversely, the model employing the Q-exponential prior maintains a remarkably low Expected Calibration Error of 0.112 even at the highest severity level. This superior calibration confirms the theoretical hypothesis advanced in Section 3: the heavy-tailed structure of the Q-exponential prior prevents uncertainty collapse. By permitting a wider dispersion of weights, the prior ensures that the variance of the pre-activation signals expands appropriately when processing corrupted inputs. This expansion directly translates into higher entropy in the softmax output, perfectly aligning the model's confidence with its degraded accuracy and thereby providing highly reliable uncertainty estimates in unpredictable environments. Image-based diagnosis of treatable disease provides an application in which representation quality must be linked to a clinically defined target before uncertainty estimates can be interpreted [27].

5.4 Sensitivity and Ablation Studies

To isolate the impact of the structural parameter within the Q-exponential framework, we conducted extensive ablation studies varying the entropic index q . When the index is set close to one, the performance mirrors that of the Gaussian baseline, confirming the theoretical connection between the distributions. As the index is incrementally increased, we observe a corresponding improvement in the out-of-distribution Negative Log-Likelihood and calibration metrics, up to a certain threshold. If the index is set excessively high, the prior becomes too uninformative, leading to optimization difficulties and a degradation of in-domain accuracy. The empirical evidence suggests that an entropic index in the range of 1.2 to 1.5 provides the optimal balance between heavy-tailed exploration and sufficient regularization for deep convolutional networks, maximizing robustness without sacrificing core predictive capabilities. For a broader comparison, related research addresses early-stage reasoning freedom in multi-agent debate [28].

6. Conclusion and Future Directions

6.1 Summary of Findings

This research has systematically investigated the critical role of prior distributions in the context of deep uncertainty propagation, specifically focusing on the vulnerabilities of standard Bayesian neural networks under covariate shift. We identified that the ubiquitous reliance on Gaussian priors, characterized by their rapidly decaying exponential tails, constitutes a structural bottleneck that severely restricts a model's ability to express uncertainty when processing out-of-distribution data. To overcome this limitation, we introduced and rigorously evaluated heavy-tailed Q-exponential priors, derived from the framework of non-extensive Tsallis statistics. By formalizing a variational inference methodology tailored to these generalized distributions, we demonstrated that the polynomial tail decay of the Q-exponential family allows for a wider exploration of the parameter space, preventing the catastrophic collapse of uncertainty during deep non-linear propagation. Extensive empirical evaluations confirmed that networks equipped with Q-exponential priors consistently deliver superior calibration and higher predictive likelihoods on highly corrupted datasets compared to traditional thin-tailed baselines. The Fe-N4/Sn-Nx catalyst study [29] offers a materials setting for a future uncertainty evaluation across coordination environments, with extrapolation tested on independently characterized catalysts rather than inferred from the prior's tail behavior.

6.2 Implications for Robust Machine Learning

The implications of this work extend significantly beyond theoretical statistical mechanics, directly impacting the deployment of machine learning in high-stakes environments. In domains such as autonomous navigation, where sensors may unexpectedly encounter extreme weather conditions or sensor degradation, the ability of a perception system to accurately report its own uncertainty is as critical as its ability to recognize objects. Similarly, in medical diagnostics, domain shifts caused by different imaging equipment or diverse patient demographics require models that fail gracefully and flag uncertain cases for human review rather than making blindly confident, erroneous predictions. The heavy-tailed Q-exponential prior provides a mathematically sound, highly configurable mechanism to ensure that deep learning models remain cautious and well-calibrated under exactly these types of unpredictable real-world shifts, significantly enhancing the safety and reliability of autonomous systems. A future clinical-translation evaluation could use a response problem such as PSPH inhibition [30] to examine uncertainty across experimental populations, without treating preclinical seizure reduction as a calibrated prediction for an individual patient.

6.3 Avenues for Future Work

While the integration of Q-exponential priors represents a substantial advancement in robust deep learning, several promising avenues for future research remain. First, the current methodology relies on a static, globally defined entropic index. Future work could explore hierarchical Bayesian formulations where the structural parameter q is learned dynamically from the data, potentially allowing different layers or distinct feature channels to exhibit varying degrees of heavy-tailed behavior based on their specific functional roles. Second, extending this framework to more complex architectures, such as transformer-based models for natural language processing, presents a challenging but highly relevant direction, as these models are also known to suffer from severe overconfidence under domain shift. Finally, investigating the interaction between Q-exponential priors and other uncertainty quantification techniques, such as deep ensembles or sophisticated data augmentation strategies, could yield synergistic approaches that push the boundaries of reliable artificial intelligence in inherently unpredictable open-world scenarios. For a broader comparison, related research addresses geometry-enhanced attentive multi-view stereo for challenging matching scenarios [31].

References

1. 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.

2. 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.
3. Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410.
4. Li, Q., Luo, T., & Liao, J. (2025). Theory-inspired deep multi-view multi-label learning with incomplete views and noisy labels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20706-20715).
5. Zhang, J., & Tang, H. (2025). Resolving Task Objective Conflicts in Unified Model via Task-Aware Mixture-of-Experts. arXiv preprint arXiv:2506.03591.
6. 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).
7. 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.
8. Yu, Z., Dai, L., Xu, S., Gao, S., & Ho, C. P. (2023). Fast bellman updates for wasserstein distributionally robust mdps. *Advances in Neural Information Processing Systems*, 36, 30554-30578.
9. 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.
10. 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.
11. 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.
12. Zhang, J., Luo, H., Yang, H., Xu, X., Wu, Q., Shi, Y., ... & Wang, J. (2023). Neurdome: 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).
13. 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.
14. Chang, Z., Obite, C., Zhou, S., & Lan, S. (2024). Deep Q-Exponential Processes. arXiv preprint arXiv:2410.22119.
15. Tian, Y. (2026, January). A Data-Driven Energy Efficiency Improvement and Cost Optimisation Model for New Energy Electric Vehicles. In 2026 IEEE International Conference on Power, Electronics and Green Energy (ICPEGE) (pp. 304-308). IEEE.
16. López-Acosta, N. P. (2019, November). Evaluating rigidity index, OCR, and su from dilatometer data in soft to firm clays. In *Geotechnical Engineering in the XXI Century: Lessons learned and future challenges: Proceedings of the XVI Pan-American Conference on Soil Mechanics and Geotechnical Engineering (XVI PCSMGE)*, 17-20 November 2019, Cancun, Mexico (p. 286). IOS Press.
17. He, H., Sun, Z., Fan, J., & Mok, P. Y. (2025). TiPGAN: High-quality tileable textures synthesis with intrinsic priors for cloth digitization applications. *Computer-Aided Design*, 183, 103866.
18. Gao, S., & Shi, L. (2015). Selecting the best simulated design with the expected opportunity cost bound. *IEEE Transactions on Automatic Control*, 60(10), 2785-2790.
19. Zhao, J., Zhang, C., Qin, M., & Yang, P. (2025). QuantFactor REINFORCE: mining steady formulaic alpha factors with variance-bounded REINFORCE. *IEEE Transactions on Signal Processing*, 73, 2448-2463.
20. 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.
21. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.
22. M. Emani et al., “A comprehensive evaluation of novel AI accelerators for deep learning workloads,” in *Proc. IEEE/ACM Int. Workshop Perform. Modeling, Benchmarking Simulation High Perform. Comput. Syst. (PMBS)*, Nov. 2022, pp. 13–25.
23. 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)*.

24. 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.
25. 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.
26. Lyu, S., Mao, S., & Ma, Y. (2025). Unified Long Video Inpainting and Outpainting via Overlapping High-Order Co-Denoising. arXiv preprint arXiv:2511.03272.
27. Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5), 1122–1131.
28. 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.
29. Xia, F., Li, B., An, B., Zachman, M. J., Xie, X., Liu, Y., ... & Cheng, Y. (2024). Cooperative Atomically Dispersed Fe–N₄ and Sn–N_x Moieties for Durable and More Active Oxygen Electroreduction in Fuel Cells. *Journal of the American Chemical Society*, 146(49), 33569-33578.
30. 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.
31. Liu, Y., Cai, Q., Wang, C., Yang, J., Fan, H., Dong, J., & Chen, S. (2024). Geometry-enhanced attentive multi-view stereo for challenging matching scenarios. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(8), 7401-7416.