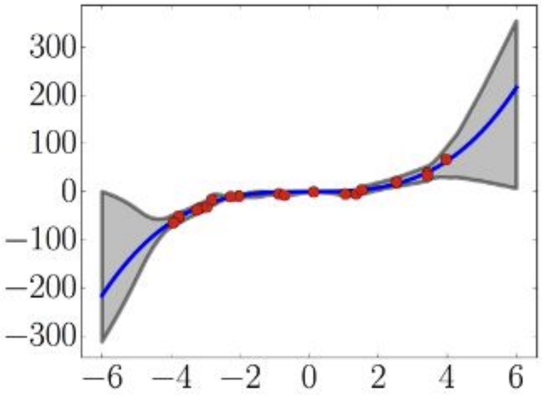
Fully Heteroscedastic Count Regression with Deep Double Poisson Networks

Spencer Young[*1], Porter Jenkins[*2], Longchao Da[3], Jeff Dotson[4], Hua Wei[3]

[1]Delicious AI, [2]Brigham Young University, [3]Arizona State University, [4]Ohio State University

*Equal contribution

Background / Motivation



Gaussian DE predictions on toy dataset.
Adapted from Figure 1 of [1].

Neural networks capable of representing their uncertainty are a crucial component of real-world ML systems.

In the *continuous* regression setting, deep ensembles (DEs) of Gaussian networks [1] have proven highly effective due to their flexibility and accuracy.

[1] Lakshminarayanan, B., Pritzel, A., & Blundell, C. Simple and scalable predictive uncertainty estimation using deep

ICML
International Conference
On Machine Learning

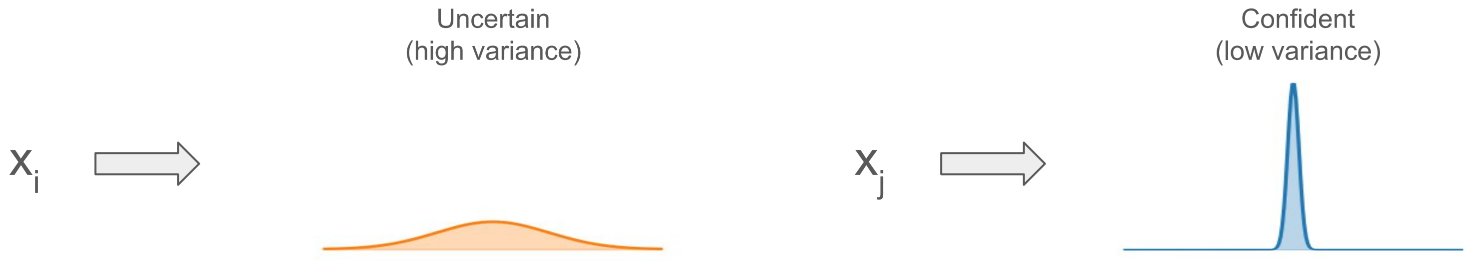
Background / Motivation

However, no analogous approach exists for *count* regression, an important subfield with many applications (estimating crowd size, inventory volume, traffic flow, etc.).

Background / Motivation

Uncertainty can be decomposed into two quantities: *aleatoric* (observation noise) and *epistemic* (model misspecification). Aleatoric uncertainty is commonly estimated via a neural network's predictive variance, while epistemic uncertainty can be estimated by deep ensembles.

In order for a model to properly represent its aleatoric uncertainty, it must be able to output arbitrarily high/low variance for any prediction. We call this property *full heteroscedasticity*.



Without accurate aleatoric uncertainty, downstream estimates of epistemic uncertainty can be corrupted [2].

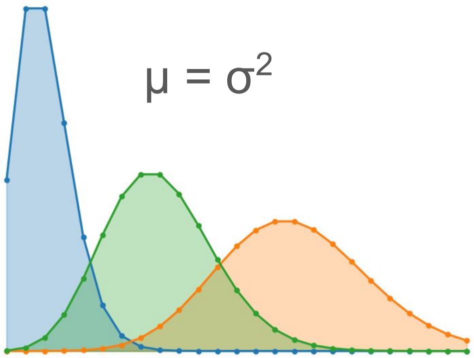
---

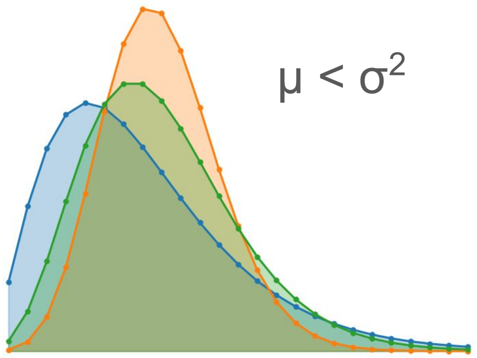
[2] Mucsányi, B., Kirchhof, M., & Oh, S. J. Benchmarking uncertainty disentanglement: Specialized uncertainties for specialized tasks. 2024

Background / Motivation

Existing approaches to deep count regression output the parameters of common distributions such as the Poisson or Negative Binomial.

These models are not fully heteroscedastic — they are constrained via *equidispersion* and *overdispersion* respectively. Thus, they often misrepresent aleatoric uncertainty.

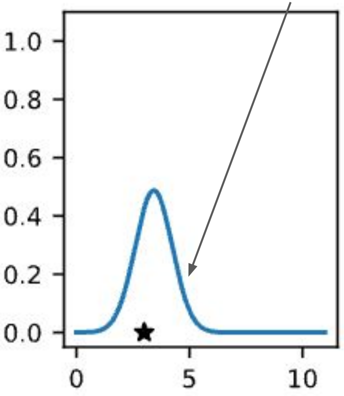


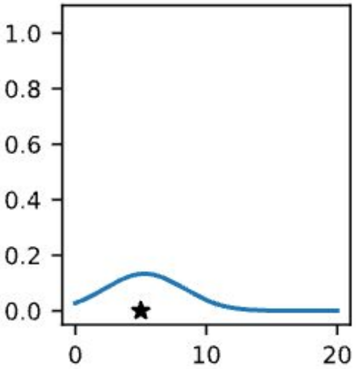


Background / Motivation

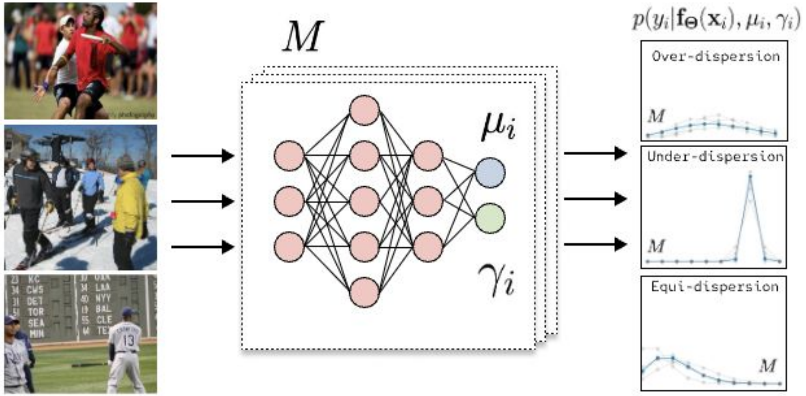
Meanwhile, Gaussian networks have no such restrictions on their predictive variance, but lack a key inductive bias for the counting setting — they model *discrete* probabilities with a *continuous* density function.

Gaussian models must assign probability density to infeasible values, like 4.9 (not a count)





Our Contribution



We propose to train neural networks to output the parameters of the Double Poisson [3] distribution. We call our model the *Deep Double Poisson Network* (DDPN).

DDPN is both discrete *and* fully heteroscedastic, thus it can effectively model integer-valued targets under various levels of aleatoric noise.

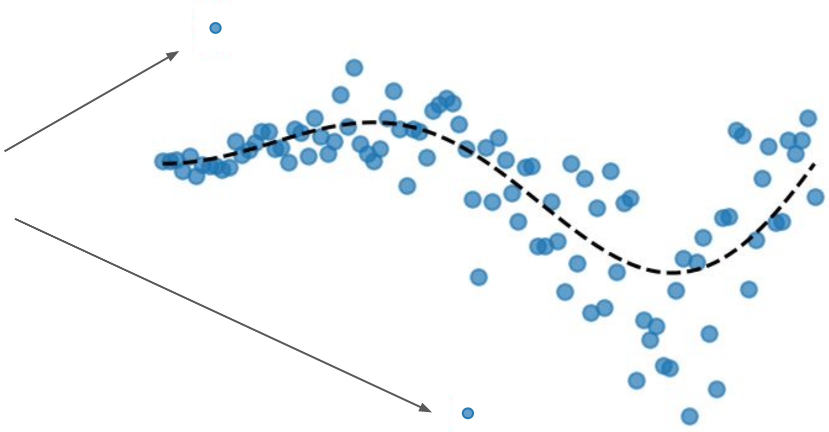
When ensembled, DDPN also provides calibrated estimates of epistemic uncertainty.

[3] Efron, Bradley. *Double exponential families and their use in generalized linear regression*, 1986

Our Contribution

We show that DDPN exhibits *learnable loss attenuation*, which allows it to discount the contribution of outliers / mislabels to its training objective.

Similar to a Gaussian model, DDPN can dynamically down-weight the contribution of individual inputs to the overall loss by assigning them high uncertainty, increasing its robustness to outliers.



Our Contribution

$$\mathscr{L}_i = \left(-\frac{\log \hat{\gamma}_i}{2} + \hat{\gamma}_i \hat{\mu}_i - \hat{\gamma}_i y_i (1 + \log \hat{\mu}_i - \log y_i) \right)$$

$$\mathscr{L}_i^{(\beta)} = \boxed{\left\lfloor \hat{\gamma}_i^{-\beta} \right\rfloor} \left(-\frac{\log \hat{\gamma}_i}{2} + \hat{\gamma}_i \hat{\mu}_i - \hat{\gamma}_i y_i (1 + \log \hat{\mu}_i - \log y_i) \right)$$

To avoid uncertainty collapse, DDPN's learnable loss attenuation can be controlled through a tunable modification to our proposed loss function.

Results

DDPN, both as an individual model and when ensembled, outperforms all baselines in terms of accuracy and calibration.

		Length of Stay		COCO-People		Inventory		Reviews	
		MAE (↓)	CRPS (↓)	MAE (↓)	CRPS (↓)	MAE (↓)	CRPS (↓)	MAE (↓)	CRPS (↓)
Aleatoric Only	Poisson DNN	0.664 (0.01)	0.553 (0.01)	1.099 (0.02)	0.851 (0.01)	1.023 (0.04)	0.706 (0.01)	0.818 (0.01)	0.559 (0.00)
	NB DNN	0.685 (0.00)	0.570 (0.00)	1.143 (0.05)	0.867 (0.01)	1.020 (0.04)	0.708 (0.01)	0.855 (0.01)	0.562 (0.00)
	Gaussian DNN	0.599 (0.01)	0.453 (0.02)	1.219 (0.12)	0.866 (0.07)	0.936 (0.01)	0.659 (0.00)	0.452 (0.01)	0.323 (0.00)
	Faithful Gaussian	0.582 (0.00)	0.436 (0.01)	1.082 (0.01)	0.879 (0.01)	0.959 (0.03)	0.688 (0.02)	0.428 (0.00)	0.428 (0.00)
	Natural Gaussian	0.597 (0.01)	0.439 (0.01)	1.157 (0.04)	0.848 (0.02)	0.958 (0.01)	0.675 (0.01)	0.428 (0.01)	0.312 (0.00)
	$\beta_{0.5}$-Gaussian	0.600 (0.01)	0.427 (0.01)	1.055 (0.01)	0.786 (0.00)	0.935 (0.01)	0.669 (0.01)	0.420 (0.00)	0.306 (0.00)
	$\beta_{1.0}$-Gaussian	0.646 (0.01)	0.462 (0.01)	1.085 (0.01)	0.809 (0.00)	0.923 (0.01)	0.653 (0.01)	0.458 (0.01)	0.327 (0.00)
	DDPN (ours)	**0.502** (0.01)	0.390 (0.04)	1.135 (0.08)	0.810 (0.03)	0.906 (0.01)	**0.632** (0.01)	0.392 (0.01)	0.277 (0.00)
	$\beta_{0.5}$-DDPN (ours)	0.516 (0.01)	**0.370** (0.01)	1.095 (0.03)	0.782 (0.02)	**0.905** (0.02)	0.635 (0.01)	**0.356** (0.01)	0.268 (0.00)
	$\beta_{1.0}$-DDPN (ours)	0.558 (0.01)	0.407 (0.01)	**1.006** (0.01)	**0.759** (0.01)	0.909 (0.01)	0.634 (0.01)	**0.356** (0.00)	**0.263** (0.00)
Aleatoric + Epistemic (DEs)	Poisson DNN	0.650	0.547	1.046	0.817	0.996	0.683	0.823	0.556
	NB DNN	0.681	0.567	1.066	0.824	0.982	0.686	0.857	0.560
	Gaussian DNN	0.590	0.450	1.148	0.815	0.902	0.634	0.447	0.319
	Faithful Gaussian	0.571	0.429	1.042	0.841	0.909	0.643	0.424	0.324
	Natural Gaussian	0.582	0.428	1.090	0.800	0.916	0.643	0.423	0.307
	$\beta_{0.5}$-Gaussian	0.591	0.420	1.019	0.740	0.879	0.619	0.414	0.302
	$\beta_{1.0}$-Gaussian	0.633	0.453	1.050	0.765	0.887	0.624	0.455	0.324
	DDPN (ours)	**0.485**	0.361	1.024	0.744	0.861	0.604	0.373	0.268
	$\beta_{0.5}$-DDPN (ours)	0.495	**0.359**	1.029	0.729	**0.840**	**0.590**	0.358	0.261
	$\beta_{1.0}$-DDPN (ours)	0.543	0.393	**0.959**	**0.712**	0.859	0.597	**0.344**	**0.257**

Results

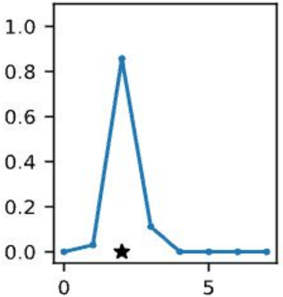
The total variance of DDPN DEs is a better OOD indicator than DEs of existing methods.

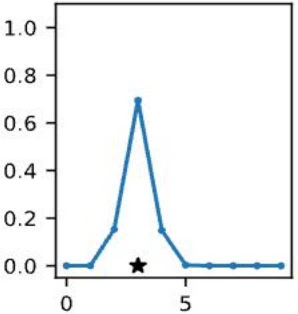
This is especially true in comparison to models that are not fully heteroscedastic (which consequently suffer from poorly-estimated aleatoric + epistemic uncertainty).

	AUROC (↑)	AUPR (↑)	FPR80 (↓)
Poisson DNN	0.330 (0.001)	0.413 (0.000)	0.793 (0.001)
NB DNN	0.280 (0.001)	0.397 (0.000)	0.819 (0.002)
Gaussian DNN	0.840 (0.001)	0.812 (0.005)	0.318 (0.002)
Faithful Gaussian	0.731 (0.001)	0.670 (0.001)	0.380 (0.002)
Natural Gaussian	0.836 (0.001)	0.827 (0.002)	0.317 (0.002)
$\beta_{0.5}$-Gaussian	0.829 (0.001)	0.797 (0.004)	0.323 (0.002)
$\beta_{1.0}$-Gaussian	0.817 (0.001)	0.806 (0.002)	0.338 (0.001)
DDPN (ours)	0.854 (0.001)	0.849 (0.003)	0.269 (0.002)
$\beta_{0.5}$-DDPN (ours)	**0.887** (0.001)	**0.875** (0.003)	**0.199** (0.001)
$\beta_{1.0}$-DDPN (ours)	<u>0.870</u> (0.001)	<u>0.851</u> (0.002)	<u>0.236</u> (0.002)

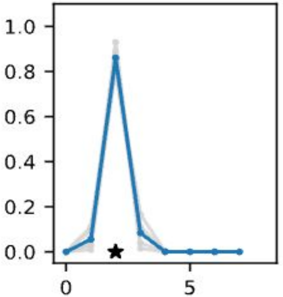
Example Predictions

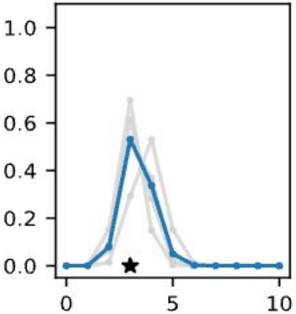
Single Network (DDPN)



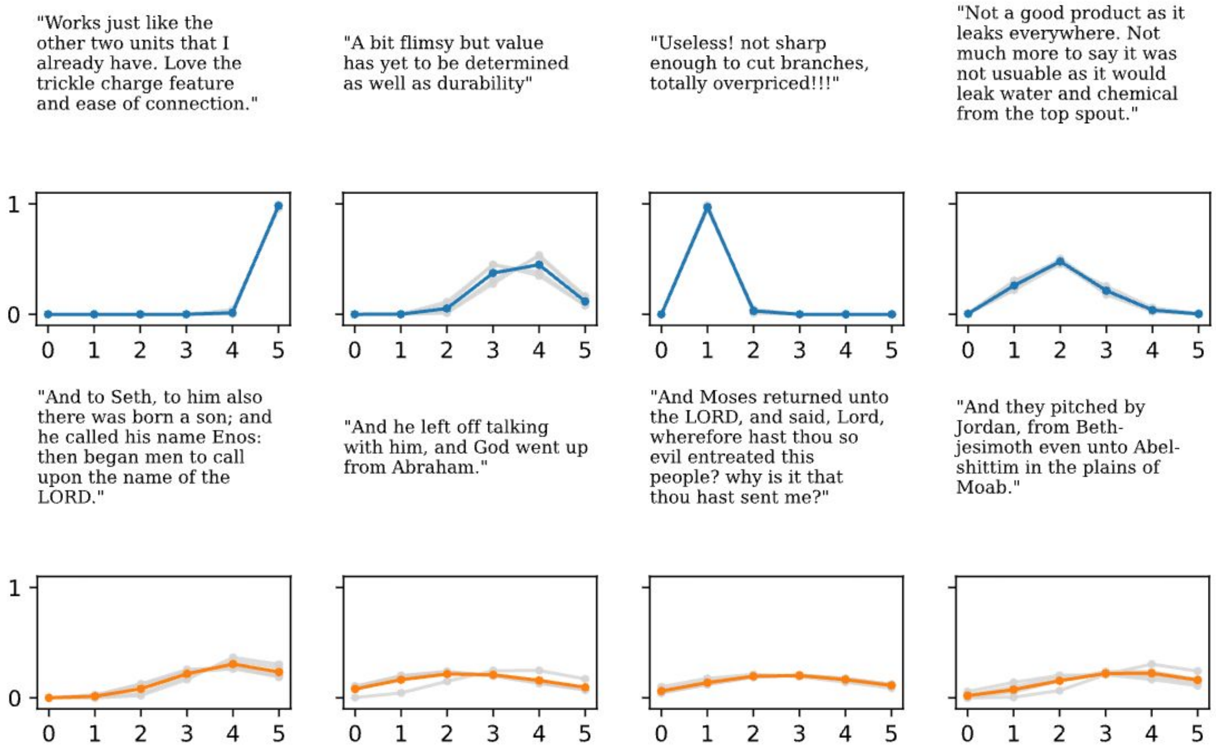


Deep Ensemble (DDPN DE)





In-distribution vs. Out-of-distribution



ICML
International Conference
On Machine Learning

Further Details

Paper

Code

Feel free to stop by our poster (ID: 46290) between 11:00 am and 1:30 pm on Thursday, July 17th.