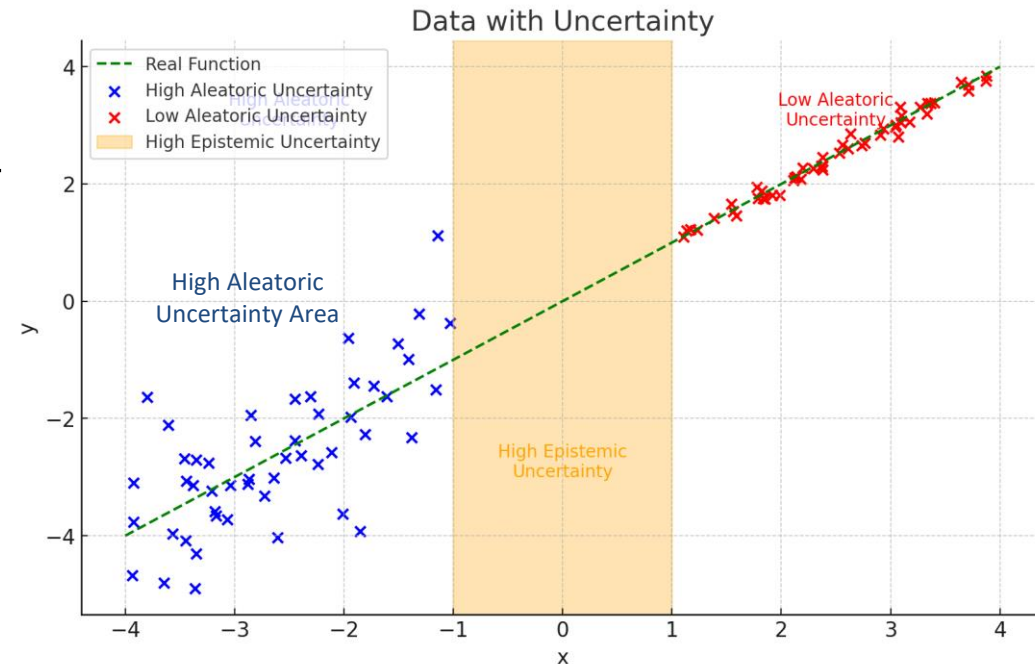


NATURAL POSTERIOR NETWORK: DEEP BAYESIAN UNCERTAINTY FOR EXPONENTIAL FAMILY DISTRIBUTIONS

Charpentier, B., Borchert, O., Zügner, D., Geisler, S., & Günnemann, S. (2021, October). Natural Posterior Network: Deep Bayesian Predictive Uncertainty for Exponential Family Distributions. In International Conference on Learning Representations.

2 Introduction

- Uncertainty Estimation은 모델이 얼마나 자신있는지, 예측이 얼마나 정확할 가능성이 있는지를 평가하는 지표
- Uncertainty estimation은 크게 두 가지 유형의 불확실성을 구분
 - Data uncertainty(Aleatoric Uncertainty) : 입력 데이터의 불확실성으로, 관측 데이터 자체의 잡음이나 불확실성을 의미(센서 데이터 측정 오류)
 - Knowledge uncertainty(Epistemic Uncertainty) : 모델의 불확실성으로, 훈련 데이터에 의해 충분히 학습되지 않은 모델이 불충분하게 이해하는 영역에서의 불확실성을 의미(모델이 더 많이 학습할수록 감소)
- 딥러닝 기반의 모델들은 overconfidence한 경향을 지님
- 신뢰할 수 있는 모델을 얻기 위해, **모르는 것을 모른다고 출력**하는 (calibration error 가 낮은)모델이 필요

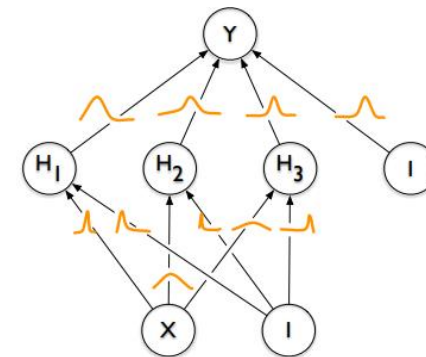
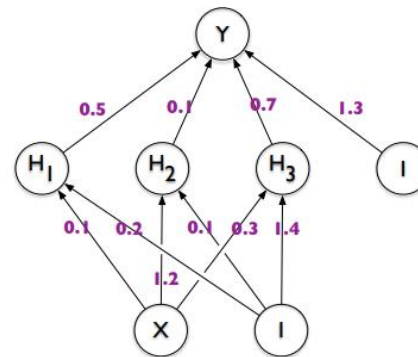


3 Introduction

- Overconfidence가 완화된 신뢰할 수 있는 딥모델을 얻기 위해 conjugate prior distribution 의 매개변수를 직접 예측하여 다양한 불확실성 수준을 반영
- 하지만 대부분의 연구는 분류에 초점을 맞추고 있으며, Normal Inverse-Gamma (NIG)의 하이퍼파라미터를 예측하는데 집중
- 또한, 여러 연구들이 학습 단계에서 분포 외(OOD) 데이터를 필요로 하기에, 실제로 사용할 수 없는 단점을 지님
- 본 논문은 범용성과 효율성을 갖춘 불확실성 추정을 위해 exponential family distributions로 표현이 가능한 모든 downstream task에 대해 OOD 데이터 필요없이, conjugate prior distribution을 매개변수화 하는 새로운 모델인 Natural Posterior Network (NatPN)제안

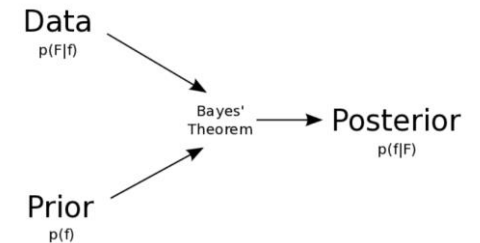
Sampling-based Uncertainty Method

- 같은 데이터가 입력되어도 여러 개의 서로 다른 출력 값을 얻게끔 학습
- 여러 번의 inference를 통해 얻은 output들의 변화량을 uncertainty로 활용
- 대표적인 방법론들
 1. Bayesian neural networks : 모델 weight의 분포($=p(w|D)$)를 학습
 2. Dropout 활용 : test phase 할 때에도 dropout을 활성화
 3. Ensemble 활용 : 서로 다른 여러 개의 모델을 학습
- 한계점
 - 메모리 비효율 : 다른 uncertainty 측정 방법에 비해 더 많은 파라미터를 학습해야 함
 - 시간 비효율 : 여러 번의 inference를 수행하기에, 다른 uncertainty 측정 방법에 비해 더 많은 시간이 소요



5 Related work

- Frequentism은 확률에 대해 '사건이 일어나는 장기적인 확률'로써 오로지 경험적 사실만을 통해 이야기할 수 있다는 객관적인 입장
 - 1000번을 던지면 주사위를 던져서 3이 횡수는 166번이다.
- Bayesianism은 '지식이나 판단의 정도를 나타내는 수단'으로써, 주관적인 입장
 - 주사위를 던져서 3이 나오는 확률이 1/6이기에, 주사위를 던지면 3이 나온다고 1/6(16.66%) 확신할 수 있다.
 - 사전지식을 활용해서 종속적(의존적) 관계에 놓인 사건들을 기반으로 경험적으로 얻기 힘든 사건에 대해 확률을 추정
- 베이지안 추론(Bayesian Inference)
 - 관측되지 않은 데이터에 대해 알고 있는 것으로 추론 대상에 대해 밝히는 것
 - 확률밀도함수 $\pi(\theta)$ 의 parameter θ 라는 추론 대상에 대해 데이터 x 를 보고 추론하는 상황에서 $f(x)$ 를 밀도함수라고 정의했을 때, 베이즈 정리에 따라 추론 대상에 대한 사전분포를 통해 사후분포를 추론할 수 있음



$$f(\theta|x) = \frac{\pi(\theta)f(x|\theta)}{f(x)} = \frac{1}{f(x)} \pi(\theta)f(x|\theta)$$

$$f(x) = \int f(x|\theta)\pi(\theta)d\theta, \text{ where } \iint f(x|\theta)\pi(\theta)dx d\theta = 1$$

- Exponential Family

확률변수 X 의 확률밀도함수(또는 확률질량함수)가 다음을 만족하면, X 는 지수족이라고 한다.

$$p(x|\eta) = \exp(T(x)^\top \eta - \mathbf{1}^\top A(\eta) - B(x))$$

이때 $T(x)$ 는 충분통계량, η 는 natural parameter, $A(\eta)$ 는 log-partition function, $B(x)$ 는 base measure 이다.

- 충분통계량(sufficient statistic) : 통계학에서 주어진 샘플로부터 얻은 정보를 요약하는 데 사용 (Ex, 베르누이 시행에서의 성공 확률 p 를 추정하기 위해, n 번의 독립 시행에서의 성공 횟수 X 는 p 에 대한 충분통계량)
- log-partition function : 분포의 매개변수에 대한 함수
- base measure : 입력 변수의 범위나 공간을 고려한 조정값
- 대표적인 종류 : 베르누이 분포, 정규 분포, 포아송 분포 등

- Conjugate Prior Distribution : 베이지안 확률에서 사후 확률이 사전 확률 분포와 같은 분포 계열에 속한 경우
 - Closed Form의 사후 확률을 얻을 수 있어서 수치 적분이 아닌, 해석적으로 적분 가능
 - 사전 확률의 분포 형태가 사후확률에서도 그대로 그 형태를 유지하되 파라미터의 값만 다르게 나타남
 - 이때, 하이퍼파라미터의 차원은 본래 모델의 차원에서 1을 더한 개수 보다 크게 나타남

- Dirichlet Distribution : 다변수 확률 분포로, 여러 개의 확률 변수가 있을 때 이들 변수의 합이 1이 되는 조건을 만족하는 연속 확률 분포

- Dirichlet 분포는 단일 시행에서 여러 범주에 대한 확률 분포의 집합을 모델링하는 데 사용

- x_i 는 i 번째 범주에 대한 확률이며, α_i 는 해당 범주의 사전 매개변수

- 모든 x_i 에 대해, $0 \leq x_i \leq 1$ 이고 $\sum_{i=1}^K x_i = 1$

$$f(\mathbf{x}; \alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^K x_i^{\alpha_i - 1} \quad B(\alpha) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^K \alpha_i\right)}$$

- $B(\alpha)$ 는 Dirichlet 분포의 정규화 상수로, 다변수 베타 함수(multivariate Beta function)를 사용하여 계산

- $\Gamma(\cdot)$ 는 감마 함수(Gamma function)로, $\Gamma(n) = (n-1)!$ 을 만족하는 연속 함수의 일반화

$$f(\mu, \sigma^2 | \mu_0, \lambda, \alpha, \beta) = \text{Normal}(\mu | \mu_0, \frac{\sigma^2}{\lambda}) \times \text{Inverse-Gamma}(\sigma^2 | \alpha, \beta)$$

- Normal Inverse-Gamma (NIG) Distribution : 두 개의 파라미터, 즉 평균 μ 와 분산 σ^2 을 갖는 정규 분포의 평균이 어떤 정규 분포를 따르고, 분산이 역 감마(Inverse-Gamma) 분포를 따를 때 그 결합 분포를 의미

- 베이زي안 통계에서 평균과 분산이 모두 불확실한 정규 분포를 가진 데이터에 대한 모델링에 유용

- Conjugate prior로 작용할 수 있어, 사후 분포를 쉽게 계산 가능

- 데이터를 통해 업데이트된 평균과 분산의 사후 분포를 얻을 수 있으며, 이를 통해 예측의 불확실성을 보다 정확하게 추정

Second-Order Predictor / Evidential Deep Learning^[1]

- Label 분포(Likelihood= $p(y|x;\theta)$)의 모수(θ)에 대한 분포(Conjugate prior= $Q(\theta|x; \alpha)$)를 학습하는 베이지안 기반 방식 -> 불확실성을 직접적으로 모델링하여, 모델이 만든 예측의 신뢰도를 평가할 수 있음
 - Classification : Categorical 분포에 대한 conjugate prior는 Dirichlet 분포 / $y \sim \text{Cat}(y;\theta) \sim \text{Dir}(\theta; \alpha)$
 - Regression : Likelihood는 normal 분포이며 conjugate prior의 평균은 normal & 분산은 NIG / $y \sim N(y; \pi=(\mu, \sigma)) \sim \mu - N(\mu; \mu_0, \sigma/\lambda), \sigma - \text{InvGamma}(\sigma; \alpha, \beta)$
- 학습된 모델은 **second-order distribution의 모수** 출력
- 특징
 - **단 한 번의 forward-pass**만으로 uncertainty estimation 가능 -> Closed-form 분포(수학적으로 명확한 해를 가진 분포)를 가정하고 있으므로, analytic solution(분산, 엔트로피 등) 계산 가능
 - 다른 방법론들(앙상블 러닝 등)에 비해 **학습 과정이 비교적 수월** -> second-order distribution에 대해 maximum likelihood / maximum a posteriori 추정하는 것만으로 충분
- 단점
 - 학습 데이터와 OOD 데이터가 서로 다른 Conjugate prior를 갖도록 학습시키기 위해 학습 시 OOD example이 필요

Contribution

- 제안하는 방법인 Natural Posterior Network은 특정 task에 국한되었던 기존 방법론들을 통합 및 일반화 한 Evidential Deep Learning 방법
- 학습 시 OOD sample이 필요하지 않도록 모델링 진행
- Classification 및 Regression의 label 분포 및 label 분포의 모수 분포 들은 아래의 수식을 모두 만족
 - $Q^{\text{post}}(\theta; \mathcal{X}^{\text{post}}, n^{\text{post}}) \propto \prod_{i=1}^N p(y_i; \theta) * p(\theta; \mathcal{X}^{\text{prior}}, n^{\text{prior}})$ $Q(\theta | \mathcal{X}^{\text{post}}, n^{\text{post}}) \propto \exp(n^{\text{post}} \theta^T \mathcal{X}^{\text{post}} - n^{\text{post}} A(\theta))$
 - $\mathcal{X}^{\text{post}} = \frac{n^{\text{prior}} \mathcal{X}^{\text{prior}} + \sum_{i=1}^N u(y_i)}{n^{\text{prior}} + N}$ (u는 Likelihood의 평균을 내는 개념과 유사)
 - $n^{\text{post}} = n^{\text{prior}} + N$ (N은 관측된 데이터의 개수)
 - 이런 분포가 있을 때 베이지안 업데이트에 의해서 데이터가 n개 들어왔을 때, $Q(\theta | x; \alpha)$ 은 다음과 같이 업데이트 진행

Likelihood \mathbb{P}	Conjugate Prior \mathbb{Q}	Parametrization Mapping m	Bayesian Loss (Eq. 5)
$y \sim \text{Cat}(\mathbf{p})$	$\mathbf{p} \sim \text{Dir}(\boldsymbol{\alpha})$	$\mathcal{X} = \frac{\boldsymbol{\alpha}}{n}$ $n = \sum_c \alpha_c$	(i) $= \psi(\alpha_{y^*}^{(i)}) - \psi(\alpha_0^{(i)})$ (ii) $= \log B(\boldsymbol{\alpha}^{(i)}) + (\alpha_0^{(i)} - C) \psi(\alpha_0^{(i)}) - \sum_c (\alpha_c^{(i)} - 1) \psi(\alpha_c^{(i)})$
$y \sim \mathcal{N}(\mu, \sigma)$	$\mu, \sigma \sim \mathcal{NT}^{-1}(\mu_0, \lambda, \alpha, \beta)$	$\mathcal{X} = \begin{pmatrix} \mu_0 \\ \mu_0^2 + \frac{2\beta}{n} \end{pmatrix}$ $n = \lambda = 2\alpha$	(i) $= \frac{1}{2} \left(-\frac{\alpha}{\beta} (y - \mu_0)^2 - \frac{1}{\lambda} + \psi(\alpha) - \log \beta - \log 2\pi \right)$ (ii) $= \frac{1}{2} + \log \left((2\pi)^{\frac{1}{2}} \beta^{\frac{3}{2}} \Gamma(\alpha) \right) - \frac{1}{2} \log \lambda + \alpha - \left(\alpha + \frac{3}{2} \right) \psi(\alpha)$

- 하지만 이러한 수식들은 x 에 dependent하지 않음 \rightarrow 이를 위해 Input-Dependent Bayesian Update 방식을 제안
 - 입력한 샘플 개수마다 개별로 사후 분포를 예측하여 single exponential family distribution을 업데이트 수행
 - $$\chi_{\text{post},x} = \frac{n^{\text{prior}} \chi^{\text{prior}} + n^{(x)} \chi^{(x)}}{n^{\text{prior}} + n^{(x)}} \text{ (업데이트된 부분들을 } x \text{에 dependent 하도록 변형)}$$
 - $N_{\text{post}}(x) = n^{\text{prior}} + n(x) \rightarrow$ epistemic uncertainty estimation에 활용
 - N^{prior} 와 χ^{prior} 는 고정된 값으로, 모든 x 가 공유하는 값
 - 두 개의 모델을 통해 각각 $\chi(x)$ 와 $n(x)$ 를 출력하도록 학습
 - 관측값이 많을수록 N_{post} 커지고 χ_{post} 도 정확해지기에 좀 더 certain 하면서 정확한 분포 예측 가능
 - Encoder: $z = f_{\theta}(x), f_{\theta}: \mathbb{R}^D \rightarrow \mathbb{R}^H (H < D)$ / 차원 축소에 사용되는 기존의 일반 인코더
 - Linear decoder for $\chi: \chi(x) = g_{\psi}(z)$ / mlp 같은 일반적인 linear decoder
 - Normalized density estimator for $n: n(x) = N_H p(z|w)$ (N_H : 하이퍼파라미터)
 - z 에 대한 분포의 density를 먼저 구한 후, 학습한 분포의 density를 구해 N_H 를 곱해 계산
 - 데이터가 많이 관측되는 area에 대해 높은 density를 주도록 학습
 - OOD에선 자연스레 낮은 density를 주게 됨(contribution 2)
 - 해당 논문에선 Normalizing flow 활용^[2]

- 사전에 관측한 0, 6 이미지 인스턴스($x^{(1)}, x^{(2)}$)에 대해서 density가 높게 설정되어 있어 좀 더 certain한 분포로 예측
- 하지만 처음 보는 OOD 인스턴스($x^{(3)}$)에 대해서 density가 낮게 설정되어 있어 certain 하지 않은 더 넓은 density 예측

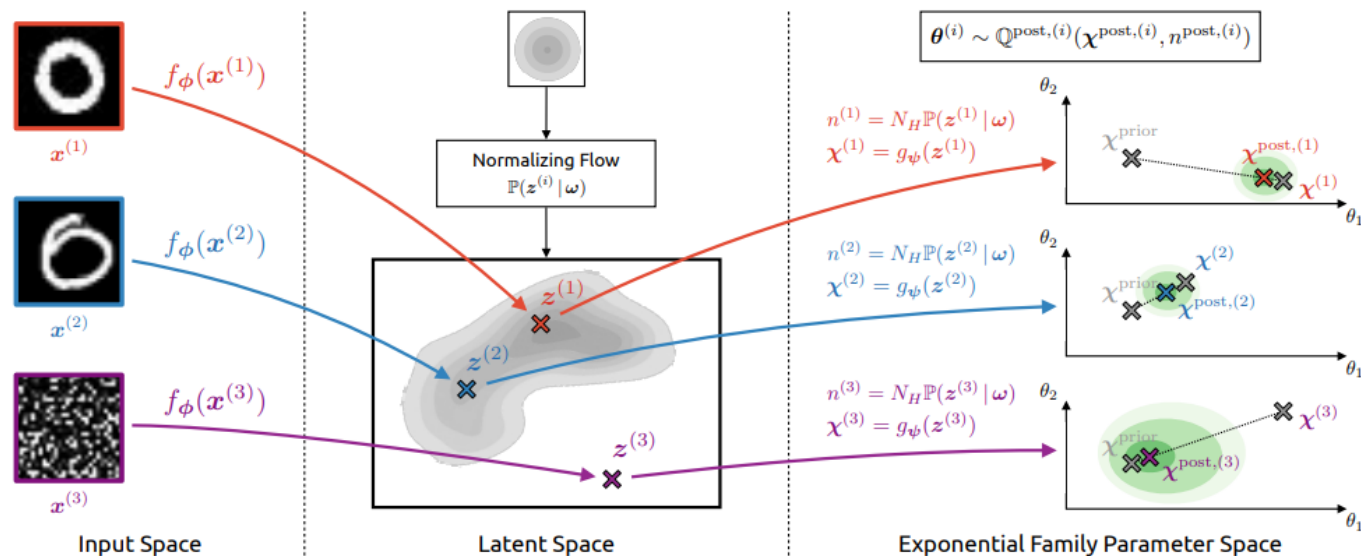


Figure 2: Overview of Natural Posterior Network. Inputs $x^{(i)}$ are first mapped to a low-dimensional latent representation $z^{(i)}$ by the encoder f_ϕ . From $z^{(i)}$, the decoder g_ψ derives the parameter update $\chi^{(i)}$ while a normalizing flow \mathbb{P}_ω yields the evidence update $n^{(i)}$. Posterior parameters are obtained from a weighted combination of prior and update parameters according to $n^{\text{post},(i)}$.

Optimization

- 기존의 Bayesian loss function을 이용해 최적화함으로써 모델 학습 진행[3]

$$\mathcal{L}^{(i)} = \underbrace{-\mathbb{E}_{\boldsymbol{\theta}^{(i)} \sim Q^{\text{post},(i)}} [\log \mathbb{P}(y^{(i)} | \boldsymbol{\theta}^{(i)})]}_{\text{(i)}} - \underbrace{\mathbb{H}[Q^{\text{post},(i)}]}_{\text{(ii)}}$$

(i) 모수가 posterior 분포를 따르도록 estimate 하는 과정에서 엔트로피를 최대화
 (ii) Entropy of $Q^{\text{post}}(.|i)$

- 본 논문의 분포들은 single exponential family distribution을 따르기에 다음과 같이 변형 가능 -> analytic form으로 계산

$$\mathcal{L}_{\lambda}^{(i)} \propto \mathbb{E}[\boldsymbol{\theta}]^T \mathbf{u}(y^{(i)}) - \mathbb{E}[A(\boldsymbol{\theta})] - \lambda \mathbb{H}[Q^{\text{post},(i)}]$$

- 해당 loss function을 이용해 encoder, linear decoder, normalized density estimator를 동시에 학습

- 체크포인트
 - Baseline과 비교하여 비슷하거나 더 높은 예측 정확도를 보이는지
 - Baseline에 비해 더 나은 uncertainty estimation을 보여주는지
 - Calibration Error
 - Estimated uncertainty를 활용한 OOD Detection(AUROC)
- 데이터셋
 - 분류 : CIFAR-10, Sensorless Drive
 - 회귀 : UCI dataset, NYU Depth v2

- Sampling based가 비sampling based 모델보다 좋은 성능을 보이고 있음
- 제안하는 방법이 여러 데이터셋에 대해 좋은 accuracy를 보이고 있지 않지만, Evidential 딥러닝 쪽에서는 좋은 성능을 보이고 있음
- OOD 체크 관련해서는 가장 좋은 성능을 보이고 있음

Table 2: Classification results on Sensorless Drive with Categorical target distribution. Best scores among all single-pass models are in bold. Best scores among all models are starred.

	Accuracy	Brier	9/10 Alea.	9/10 Epist.	OODom Alea.	OODom Epist.
Dropout	98.62 ± 0.11	3.79 ± 0.29	30.20 ± 0.85	32.57 ± 1.45	27.03 ± 0.51	95.30 ± 1.66
Ensemble	98.83 ± 0.17	3.00 ± 0.54	30.79 ± 0.74	32.61 ± 1.06	27.16 ± 0.59	99.97 ± 0.01
NatPE	*99.66 ± 0.03	*0.68 ± 0.05	77.05 ± 1.93	83.73 ± 1.89	99.99 ± 0.00	*100.00 ± 0.00
R-PriorNet	98.85 ± 0.25	2.01 ± 0.47	40.13 ± 2.99	30.07 ± 0.81	*100.00 ± 0.00	23.59 ± 0.00
EnD ²	93.95 ± 2.35	28.09 ± 6.40	26.35 ± 0.60	24.85 ± 0.43	84.43 ± 15.21	23.58 ± 0.00
PostNet	99.64 ± 0.02	0.75 ± 0.08	80.60 ± 1.68	*92.57 ± 1.41	*100.00 ± 0.00	*100.00 ± 0.00
NatPN	99.61 ± 0.05	1.04 ± 0.29	*81.43 ± 1.89	79.54 ± 2.62	99.98 ± 0.00	*100.00 ± 0.00

OOD 체크

Table 3: Classification results on CIFAR-10 with Categorical target distribution. Best scores among all single-pass models are in bold. Best scores among all models are starred. Gray numbers indicate that R-PriorNet has seen samples from the SVHN dataset during training.

	Accuracy	Brier	SVHN Alea.	SVHN Epist.	CelebA Alea.	CelebA Epist.	OODom Alea.	OODom Epist.
Dropout	88.15 ± 0.20	19.59 ± 0.41	80.63 ± 1.59	73.09 ± 1.51	71.84 ± 4.28	71.04 ± 3.92	18.42 ± 1.11	49.69 ± 9.10
Ensemble	*89.95 ± 0.11	17.33 ± 0.17	85.26 ± 0.84	82.51 ± 0.63	76.20 ± 0.87	74.23 ± 0.78	25.30 ± 4.02	89.21 ± 7.55
NatPE	89.21 ± 0.09	17.41 ± 0.12	85.66 ± 0.34	*83.16 ± 0.67	*78.95 ± 1.15	*82.06 ± 1.30	87.27 ± 1.79	*98.88 ± 0.26
R-PriorNet	88.94 ± 0.23	*15.99 ± 0.32	99.87 ± 0.02	99.94 ± 0.01	67.74 ± 4.86	59.55 ± 7.90	42.21 ± 8.77	38.25 ± 9.82
EnD ²	84.03 ± 0.25	40.84 ± 0.36	*86.47 ± 0.66	81.84 ± 0.92	75.54 ± 1.79	75.94 ± 1.82	42.19 ± 8.77	15.79 ± 0.27
PostNet	87.95 ± 0.20	20.19 ± 0.40	82.35 ± 0.68	79.24 ± 1.49	72.96 ± 2.33	75.84 ± 1.61	85.89 ± 4.10	92.30 ± 2.18
NatPN	87.90 ± 0.16	19.99 ± 0.46	82.29 ± 1.11	77.83 ± 1.22	76.01 ± 1.18	76.87 ± 3.38	*93.67 ± 3.03	94.90 ± 3.09

OOD 체크

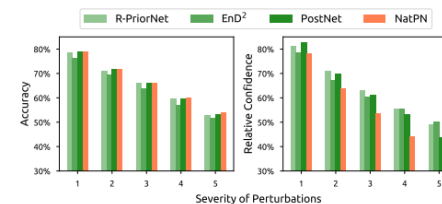


Figure 3: Averaged accuracy and confidence under 15 dataset shifts on CIFAR-10 (Hendrycks & Dietterich, 2019). On more severe perturbations (i.e. data further away from data distribution), NatPN maintains a competitive accuracy while assigning higher epistemic uncertainty as desired. Baselines provide a slower relative confidence decay.

- Sampling based가 비sampling based 모델보다 좋은 성능을 보이고 있음
- Regression은 좋은 RMSE를 보이며 또한 Evidential 딥러닝 쪽에서는 좋은 성능을 보이고 있음
- OOD 체크 관련해서는 가장 좋은 성능을 보이고 있음

Table 5: Regression results on models trained on different UCI datasets with Normal target distribution. The upper half displays models trained on Kin8nm, the lower half shows models trained on Concrete Compressive Strength.

	RMSE	Calibration	Energy Alea.	Energy Epist.	Concrete Alea.	Concrete Epist.	Kin8nm Alea.	Kin8nm Epist.
Dropout	0.09 ± 0.00	3.13 ± 0.43	90.18 ± 6.00	99.94 ± 0.06	*100.00 ± 0.00	*100.00 ± 0.00	in-distribution	
Ensemble	*0.07 ± 0.00	2.69 ± 0.49	*100.00 ± 0.00	*100.00 ± 0.00	*100.00 ± 0.00	*100.00 ± 0.00		
NatPE	0.08 ± 0.00	5.49 ± 0.30	*100.00 ± 0.00	*100.00 ± 0.00	*100.00 ± 0.00	*100.00 ± 0.00		
EvReg	0.09 ± 0.00	3.74 ± 0.53	88.06 ± 11.94	88.06 ± 11.94	*100.00 ± 0.00	86.84 ± 13.16	in-distribution	
NatPN	0.08 ± 0.00	*2.04 ± 0.45	*100.00 ± 0.00	*100.00 ± 0.00	*100.00 ± 0.00	*100.00 ± 0.00		
Dropout	5.67 ± 0.07	*3.03 ± 0.40	9.33 ± 0.36	93.53 ± 2.41	in-distribution		1.09 ± 0.13	64.30 ± 7.14
Ensemble	5.69 ± 0.20	3.81 ± 0.67	54.19 ± 18.93	*100.00 ± 0.00			72.57 ± 19.32	*100.00 ± 0.00
NatPE	*4.78 ± 0.20	5.58 ± 1.27	*100.00 ± 0.00	*100.00 ± 0.00			*100.00 ± 0.00	*100.00 ± 0.00
EvReg	6.04 ± 0.18	7.36 ± 1.04	8.93 ± 0.02	51.39 ± 18.56	in-distribution		0.93 ± 0.00	34.44 ± 20.95
NatPN	5.83 ± 0.23	5.41 ± 1.33	*100.00 ± 0.00	*100.00 ± 0.00			*100.00 ± 0.00	*100.00 ± 0.00

Table 6: Regression results on NYU Depth v2 with Normal target distribution. RMSE is in cm. OOD scores on LSUN are reported on the held-out classes 'classrooms' (left) and 'churches' (right).

	RMSE	Calibration	LSUN Alea.	LSUN Epist.	KITTI Alea.	KITTI Epist.	OODom Alea.	OODom Epist.
Dropout	46.95	4.03	*95.29 / 97.74	83.89 / 83.22	98.07	84.90	74.40	*100.00
EvReg	*28.88	*1.05	58.70 / 56.71	70.19 / 64.02	56.60	62.67	75.43	56.39
NatPN	29.72	1.14	94.13 / *98.67	*89.08 / *90.56	*98.93	*93.15	*100.00	*100.00

← OOD 체크 →

- 다른 evidential learning 모델에 비해서 inference 속도가 매우 크게 향상된 모습을 보이고 있음

Table 7: Batched Inference Time (in ms),
NVIDIA GTX 1080 Ti

	CIFAR-10 (batch size 4,096)	NYU Depth v2 (batch size 4)
Dropout	407.91 ± 5.65	650.96 ± 0.22
Ensemble	361.61 ± 5.41	649.78 ± 0.18
R-PriorNet	61.83 ± 2.57	—
EnD²	61.83 ± 2.57	—
PostNet	88.56 ± 0.06	—
EvReg	—	129.88 ± 0.75
NatPN	75.64 ± 0.04	137.13 ± 0.18
NatPE	370.17 ± 0.09	676.74 ± 0.38

- 기존 방법론들을 통합하면서, 동시에 evidential deep learning의 한계점을 극복
 - Exponential family 라는 특성을 활용해 특정 task에 국한되었던 기존 방법론들을 통합
 - Posterior update 방식을 활용해 학습 시 OOD sample이 필요하지 않도록 모델링
 - 해당 모델이 OOD 데이터에 대해 높은 uncertainty를 출력한다는 것을 이론적으로 보장
(Theoretically guaranteed, under the mild assumptions)

Reference

1. Sensoy, M., Kaplan, L., & Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. Advances in neural information processing systems, 31.
2. Rezende, D., & Mohamed, S. (2015, June). Variational inference with normalizing flows. In International conference on machine learning (pp. 1530-1538). PMLR.
3. Charpentier, B., Zügner, D., & Günnemann, S. (2020). Posterior network: Uncertainty estimation without ood samples via density-based pseudo-counts. Advances in neural information processing systems, 33, 1356-1367.
4. Biloš, M., Charpentier, B., & Günnemann, S. (2019). Uncertainty on asynchronous time event prediction. Advances in Neural Information Processing Systems, 32.
5. <https://jryoungwdl.notion.site/Bayesian-Deep-Learning-Yarin-Gal-30dbc30097e547fcb018aef6c1185f13#7fe97e4945cf465db022d399fc7b0f8c>
6. <https://shinminyong.tistory.com/37>