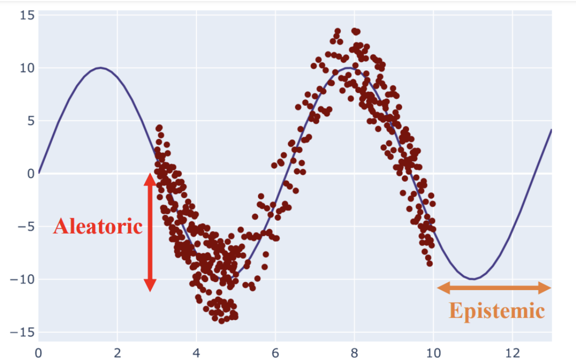
Decomposition of Uncertainty in Bayesian Deep Learning for Efficient and Risk-sensitive Learning

S. Depeweg, J.M. Hernandez-Lobato, F. Doshi-Velez & S. Udluft

Paris Dauphine - Université PSL
Nathaniel Cogneaux

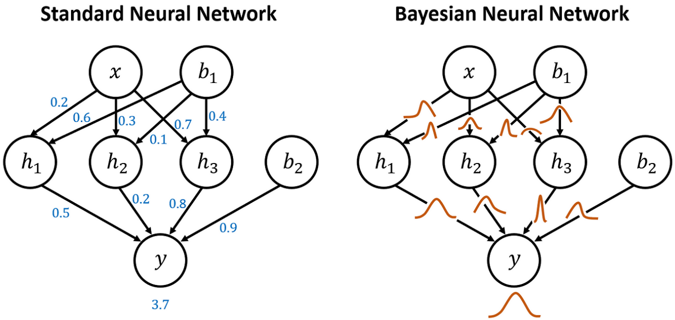
Main Concepts



Two types of uncertainty:

- **Aleatoric uncertainty:** Captures noise inherent in the observations.
- **Epistemic uncertainty:** Accounts for uncertainty in the model, which can be explained away with enough data.

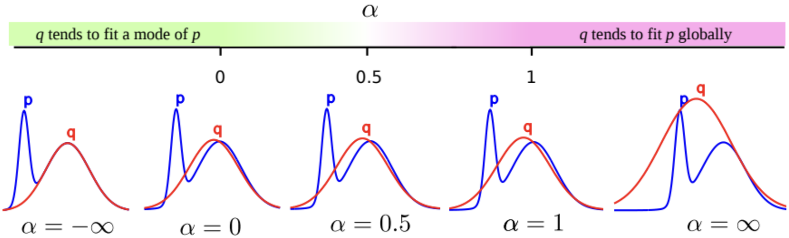
Bayesian Neural Networks + Latent Variables



$$p(y_*|x_*, D) = \int\int p(y_*|W, x_*, z_*)p(z_*)dz_*\, p(W, z|D)dWdz.$$

$$p(W, z|D) = \frac{p(y_*|W, x_*, z_*)p(z_*)p(W)}{p(Y|X)}$$

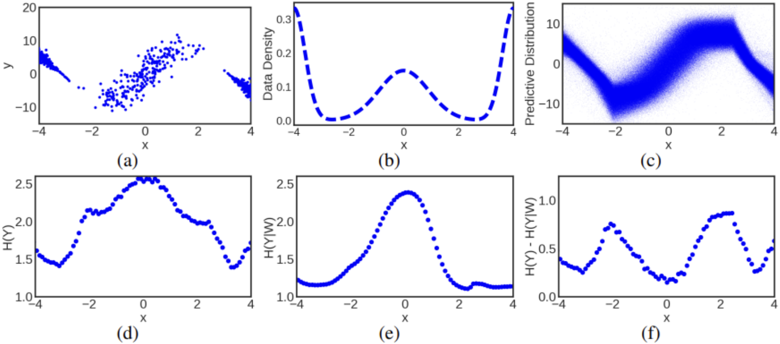
Bayesian Neural Networks + Latent Variables



$$q(W,Z) = \left[\prod_{l=1}^{L}\prod_{i=1}^{V_l}\prod_{j=1}^{V_{l-1}+1}\mathcal{N}(w_{ij,l}|m^w_{ij,l}, v^w_{ij,l})\right] \times \left[\prod_{n=1}^{N}\mathcal{N}(z_n|m^z_n, v^z_n)\right]$$

$$D_\alpha[p(W,z|D)\parallel q(W,z)] = \frac{1}{\alpha(\alpha-1)}\left(1 - \int p(W,z|D)^\alpha q(W,z)^{(1-\alpha)}\,dWdz\right)$$

A novel approach: Uncertainty Decomposition



(a) (b) (c)
(d) (e) (f)

$$H[y_*|x_*] - \mathbb{E}_{q(W)}[H(y_*|W, x_*)] = \mathcal{I}(y_*, W)$$

$$\sigma^2(y_*|x_*) = \sigma^2_{q(W)}(\mathbb{E}[y_*|W, x_*]) + \mathbb{E}_{q(W)}[\sigma^2(y_*|W, x_*)].$$

Risk-sensitive Reinforcement Learning

model-based RL:

- Learn model dynamics from Batch $\mathcal{D} = \{(s_t, a_t, s_{t+1})\}$ and cost function c
- Assumption: dynamical system can be expressed by an unknown neural network:
  $s_t = f_{\text{true}}(s_{t-1}, a_{t-1}, z_t; W_{\text{true}})$

Method:

- Given s_0, we sample $W \sim q$
- for T steps using the model $s_{t+1} = f(s_t, a_t, z_t; W) + \varepsilon_{t+1}$
- policy $a_t = \pi(s_t; W_\pi)$, input noise $z_t \sim \mathcal{N}(0, \gamma)$ and additive noise $\varepsilon_{t+1} \sim \mathcal{N}(0, \Sigma)$

Monte Carlo approximation of the expected cost given the initial state s_0:

$$J(W_\pi) = \mathbb{E}_{s_0 \sim \mathcal{D}}\left[\sum_{t=1}^{T} c_t\right]$$

Risk-Sensitive criterion

Usual balance between expected cost and risk:

$$J(W_\pi) = \mathbb{E}[C] + \beta\sigma(C)$$

With uncertainty decomposition:

$$\sigma(c_t) = \left(\sigma^2_{q(W)}(\mathbb{E}[c_t|W]) + \mathbb{E}_{q(W)}[\sigma^2(c_t|W)]\right)^{\frac{1}{2}}$$

summing the resulting expression for t = 1, . . . , T:

$$\sigma(\gamma, \beta) = \sum_{t=1}^{T} \left(\beta^2\sigma^2_{q(W)}(\mathbb{E}[c_t|W]) + \gamma^2\mathbb{E}_{q(W)}[\sigma^2(c_t|W)]\right)^{\frac{1}{2}}.$$

New risk-sensitive criterion:

$$J(W_\pi) = \mathbb{E}[C] + \sigma(\gamma, \beta),$$

Understanding Epistemic Risk in Model-Based RL

The epistemic risk reflects model-bias in model-based RL:

$$b(W_\pi) = \sum_{t=1}^{T} \left(\mathbb{E}_{\text{true}}[c_t] - \mathbb{E}[c_t]\right)^2 ,$$

$$\mathbb{E}[b(W_\pi)] = \mathbb{E}_{q(W_{\text{true}})}\left[\sum_{t=1}^{T} \left(\mathbb{E}[c_t | W_{\text{true}}] - \mathbb{E}[c_t]\right)^2\right] ,$$

$$= \sum_{t=1}^{T} \sigma^2_{q(W_{\text{true}})}(\mathbb{E}[c_t | W_{\text{true}}]).$$

Risk term guide the policy to operate in areas of state space where model-bias is expected to be low.

Experiments

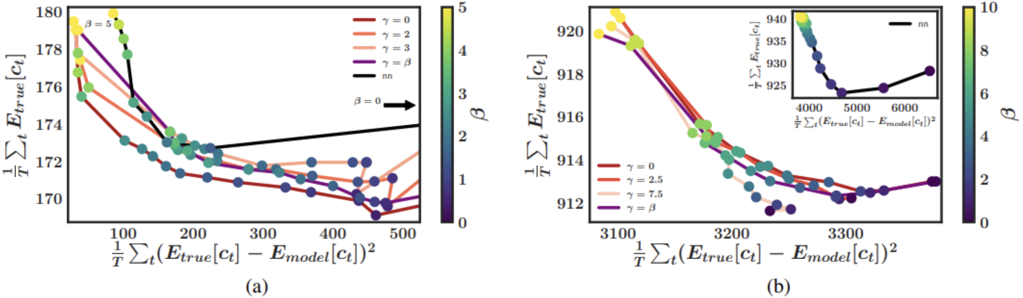


Figure 4. RL experiments. (a): results on industrial benchmark. (b): results on wind turbine simulator. Each curve shows average cost (y-axis) against model-bias (x-axis). Circle color corresponds to different values of β (epistemic risk weight) and curve color indicates different values of γ (aleatoric risk weight). The purple curve is the baseline $\gamma = \beta$. The black curve is nearest neighbor baseline.

Experiments

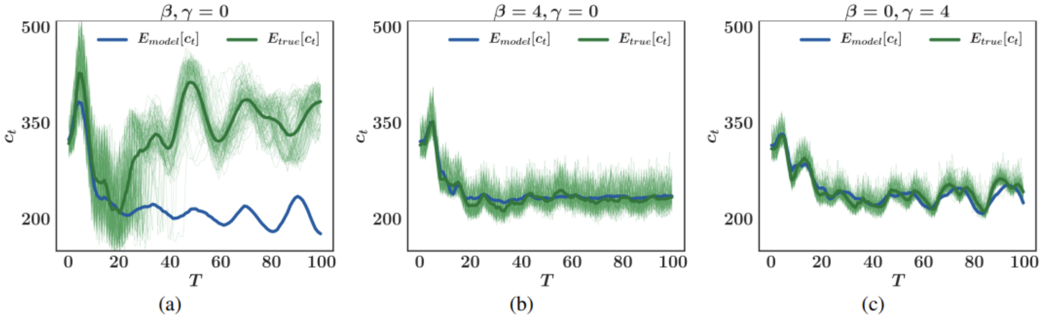


Figure 5. 100 roll-outs on the industrial benchmark ground truth system (light green), their average cost (dark green), and the average cost of corresponding roll-outs on the BNN+LV model (blue) for a fixed value of the initial state s_0. We show results for three policies with different epistemic and aleatoric risk trade-offs. Policies are optimized using (a): no risk penalty ($\beta, \gamma = 0$). (b): a penalty on the epistemic risk only ($\gamma = 0, \beta = 4$). (c): a penalty on the aleatoric risk only ($\gamma = 4, \beta = 0$).