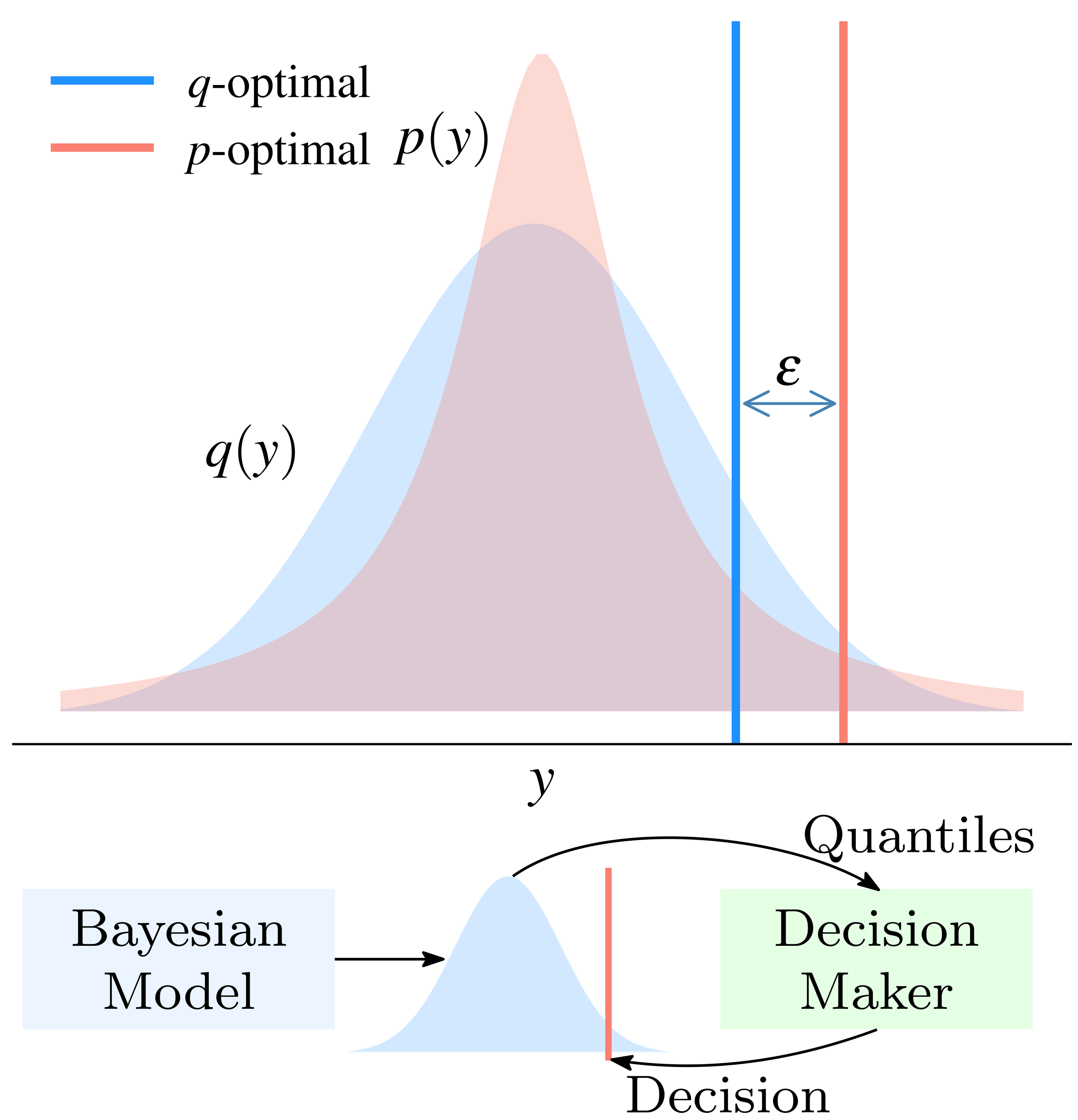
# Correcting Predictions for Approximate Bayesian Inference

Tomasz Kuśmierczyk Joseph Sakaya Arto Klami

Helsinki Institute for Information Technology HIIT
Department of Computer Science, University of Helsinki




- **Context:** Making predictions with Bayesian models.
- **Problem:** Approximate posteriors lead to sub-optimal decisions (=predictions) in down-stream tasks.
- **Solution:** Correct predictions based on the (approximate) posterior predictive distributions.
- **Advantages:** Retain speed of approximate inference and interpretability of Bayesian models.

## Our solution: Neural Network mapping predictive distributions to decisions $h$ to minimize Empirical Risk

1. Train Bayesian model and obtain quantiles from (approximate) predictive distributions $q(y)$
2. Train parametric *Decision Maker* (DM) with $q(y)$:

$$h = f(q(y), \omega)$$

3. Deploy Decision Maker along with Bayesian Model

## Traditional Bayesian Decision Theory

### Predictions from Bayesian Models

1. Find posterior $p(\theta|\mathcal{D})$
2. Make decisions $h$ minimizing the expected loss, i.e., *risk* [1]:

$$\mathcal{R}(h) = \int p(y)\ell(y, h)dy$$

where $p(y) := \int p(\theta|\mathcal{D})p(y|\theta, \mathcal{D})d\theta$ and $\ell$ is a *loss* of choice.
*For example,* squared loss implies $h =$ mean of $p(y)$.

### Challenges

- Approximation $q(\theta) \approx p(\theta|\mathcal{D})$ is often used instead
- Decisions made with $q(y)$ instead of $p(y)$ are sub-optimal
- Sub-optimal decisions do not minimize the true risk

## Predictive Decision Theory

- *Decision belief distribution* [2] of an individual data point $y$:

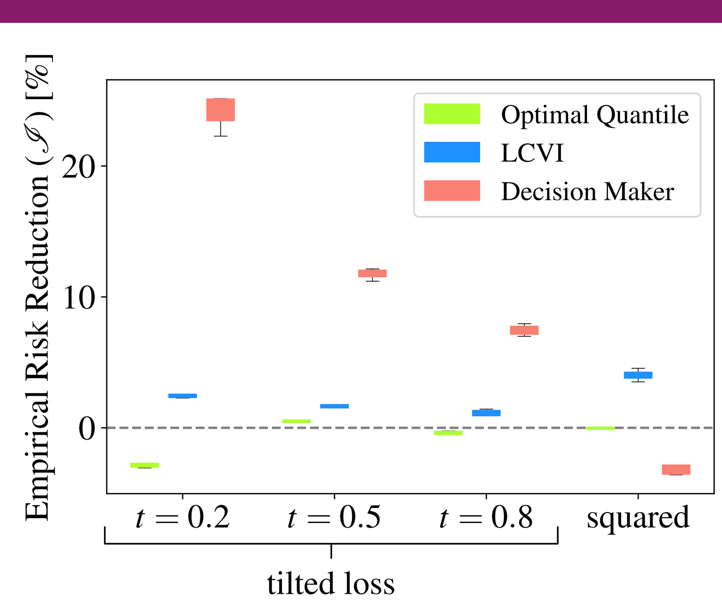
$$p(h|y) \propto e^{-\ell(h,y)}p(h)^{\lambda_1}$$

- Decisions for different $y$ can be *tied* via *Decision Maker* to obtain Bayesian update rule for belief distribution of $\omega$: $p(\omega|y) \propto e^{-\ell'(\omega,y)}p(\omega)^{\lambda_1}$
- Bayesian update rule for a collection of $N$ data instances leads to the log-posterior (=*training objective*):

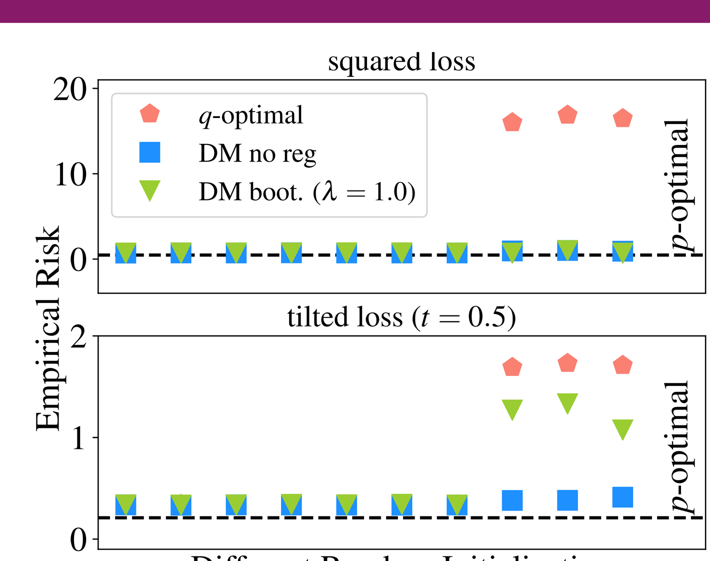
$$\log p(\omega|\mathcal{D}) \propto \underbrace{-\frac{1}{N}\sum_{n=1}^{N}\ell(\, f(q(y_n), \omega),\; y_n)}_{\text{empirical risk}} + \underbrace{\lambda \log p(\omega)}_{\text{regularization}} + C$$

- Prior $p(\omega)$ keeps predictions close to decisions obtained with $q(y)$.

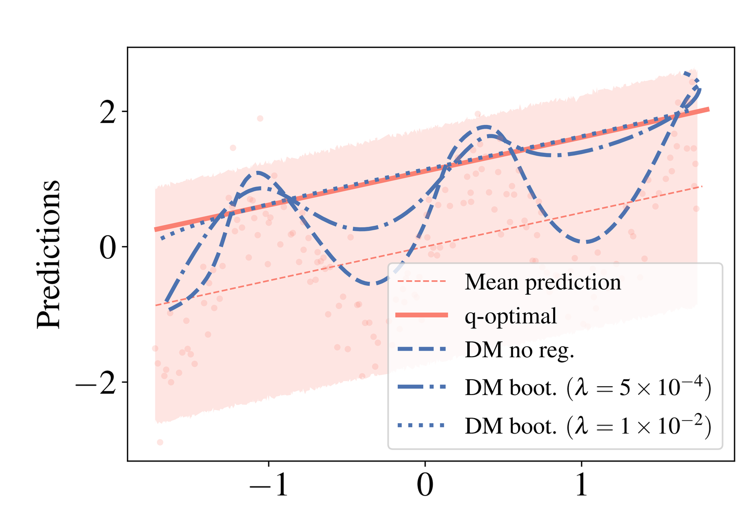
## Experiments



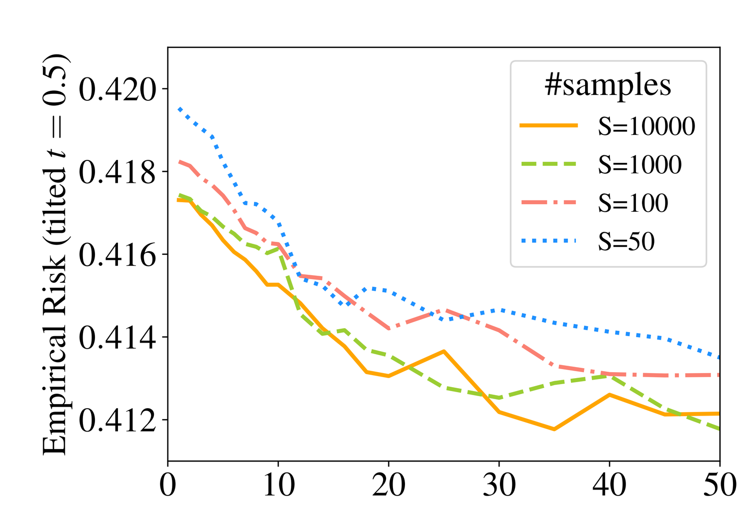

DM improves over the q-optimal baseline (black line) and competing approaches [3].




DM corrects prediction errors due to failure of posterior fitting.




Regularization prevents DM from ignoring predictions of the underlying linear model.




Worse representation of $q(y)$ (fewer quantiles) affects prediction quality.

## References

[1] Berger, J. O. (2013). Statistical decision theory and Bayesian analysis. Springer Science & Business Media.

[2] Bissiri, P. G., Holmes, C. C., & Walker, S. G. (2016). A general framework for updating belief distributions. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 78(5), 1103-1130.

[3] Kuśmierczyk, T., Sakaya, J., & Klami, A. (2019). Variational Bayesian Decision-making for Continuous Utilities. Advances in Neural Information Processing Systems 32 (NeurIPS 2019).