Marginally Calibrated Deep Distributional Regression

Yeseul Jeon

May 21, 2021

Contents

1. History
2. Introduction
3. Copula Model
4. Marginal Calibration
5. Application to lieklihood-Free Inference

Studies for estimating Deep learning Uncertainty

- **Parameters estimation**
  $w \sim (\mu, \sigma)$ using Variational Inference
- **Predictive Distribution Estimation**
  Regard weight matrix of DNN as Gaussian Process
  Mc dropout from Bernoulli(p_i)
  Construct distribution of $\hat{y}_i$ using samples from N iteration of MC dropout
- **Confidence Interval for prediction**
  Construct Calibration for predictive distribution to calculate confidence interval using quantile regression

Preidctive distribution estimation

Marginally-calibrated deep distributional regression

Nadja Klein*†, David J. Nott‡ and Michael Stanley Smith§

Abstract

Deep neural network (DNN) regression models are widely used in applications requiring state-of-the-art predictive accuracy. However, until recently there has been little work on accurate uncertainty quantification for predictions from such models. We add to this literature by outlining an approach to constructing predictive distributions that are 'marginally calibrated'. This is where the long run average of the predictive distributions of the response variable matches the observed empirical margin. Our approach considers a DNN regression with a conditionally Gaussian prior for the final layer weights, from which an implicit copula process on the feature space is extracted. This copula process is combined with a non-parametrically estimated marginal distribution for the response. The end result is a scalable distributional DNN regression method with marginally calibrated predictions, and our work complements existing methods for probability calibration. The approach is first illustrated using two applications of dense layer feed-forward neural networks. However, our main motivating applications are in likelihood-free inference, where distributional deep regression is used to estimate marginal posterior distributions. In two complex ecological time series examples we employ the implicit copulas of convolutional networks, and show

*Nadja Klein is Assistant Professor of Applied Statistics at Humboldt-Universität zu Berlin

†Communicating Author: nadja.klein@hu-berlin.de. Routines required to estimate the DNNC for Section 4 are provided as part of the supplementary material.

‡David J. Nott is Associate Professor of Statistics and Applied Probability at National University of Singapore

§Michael Stanley Smith is Professor of Management (Econometrics) at Melbourne Business School, University of Melbourne.

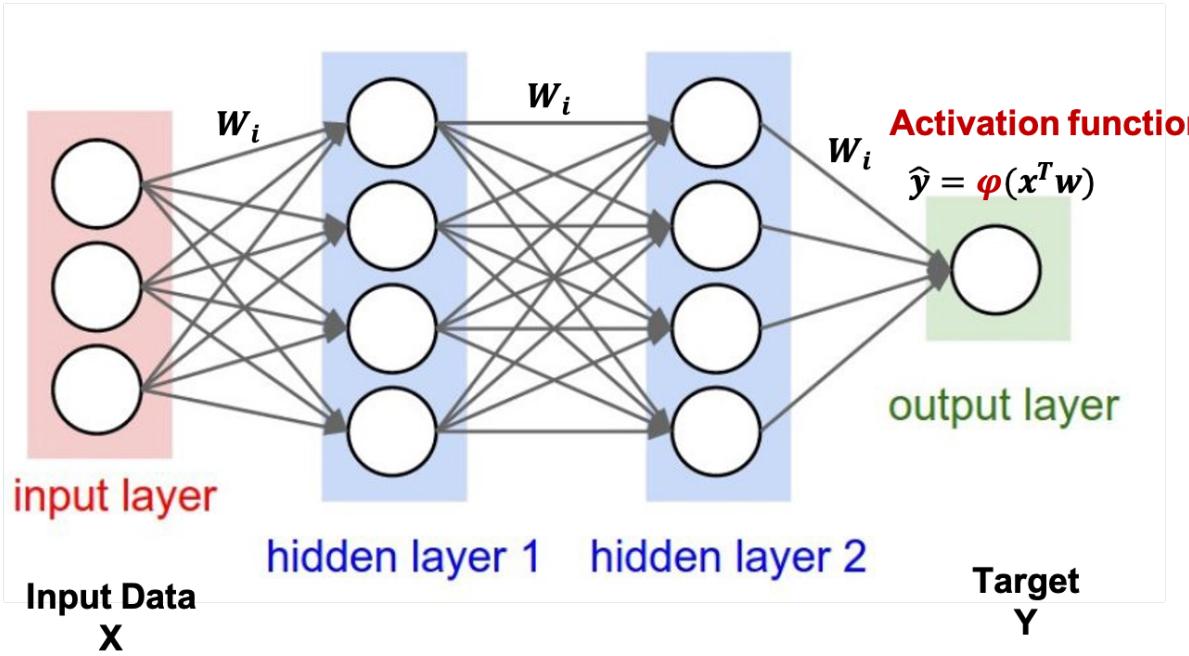
Goal

- To construct predictive distribution of y given x (x,y)

$$p(\theta|x) = p(x|\theta)p(\theta)$$

- Too complex to calculate $p(x|\theta)$

Deep Learning Model



- Outcome $y = \varphi(x^T w) = \varphi(x^T w - b)$
- $f_\eta(x_i) = \phi_\lambda(x_i)^T \beta + \beta_0$

where $\eta = (\beta_0, \beta, \lambda^T)$

Idea

- **Copula**
  Construct $p(x|\theta)$ using Copula
- **Regression Copula**
  Employ the copula of a pseudo-response vector from a DNN regression model

Quick Summary

- **Step1:**
  Estimate the marginal F_y using a nonparametric estimator using Kernel density.
  $$\Phi^{-1}(F_y(y)) = \tilde{z}$$
- **Step2:**
  Fit DNN with x_i and $\tilde{z}_i$
- **Step3:**
  Extract last layer of DNN, $B_w(x_i)$
- **Step4:**
  Fit linear model with $\tilde{z}_i$ and $B_w(x)$: $\tilde{z}_i = B_w(x)\beta + \epsilon$

Quick Summary

- **Step5:**
  Estimate β by constructing hierachical structure using MCMC.
- **Step6:**
  With β distribution of samples, derive predicitve distribution like bayes regression

Copula Model

It is a function which is used to build the dependence between two or more random variables

$$p(\boldsymbol{y}|\boldsymbol{x}) = c^{\dagger}(F(y_1|\boldsymbol{x}_1), \ldots, F(y_n|\boldsymbol{x}_n)|\boldsymbol{x}) \prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i)\,,$$

where $F(y_i|x_i)$ is the distribution function of $Y_i|x_i$ and $c^t(u|x)$ is n-dimensional copula density

Copula of DNN regression

Klein and Smith (2019) suggested calibrating the distribution of $Y_i|x_i$ to its invariant margin, so that density $p(y_i|x_i) = p_y(y_i)$ with distribution function F_Y estimated non-parametrically.

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) = c_{\text{DNN}}(F_Y(y_1), \ldots, F_Y(y_n)|\boldsymbol{x}, \boldsymbol{\theta}) \prod_{i=1}^{n} p_Y(y_i)\,.$$

Regression Copula

$$\tilde{\boldsymbol{Z}} = B_{\boldsymbol{\zeta}}(\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,,\ \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I)$$

where $B_\zeta = [\Psi_\zeta(x_1)| \cdots |\Psi_\zeta(x_n)]^T$ is $(n \times p)$ matrix.

Hierarchical Structure

$C_{\text{DNN}}(u|x,\theta)$ refers to Gaussian copula

$$\boldsymbol{\beta}|\boldsymbol{\theta},\sigma^2 \sim N(\mathbf{0},\sigma^2 P(\boldsymbol{\theta})^{-1})$$

$$\tilde{\boldsymbol{Z}}|\boldsymbol{x},\sigma^2,\boldsymbol{\theta} \sim N\left(0,\sigma^2\left(I + B_{\boldsymbol{\zeta}}(\boldsymbol{x})P(\boldsymbol{\theta})^{-1}B_{\boldsymbol{\zeta}}(\boldsymbol{x})^\top\right)\right)$$

$$c_{\text{DNN}}(\boldsymbol{u}|\boldsymbol{x},\boldsymbol{\theta}) = \frac{p(\boldsymbol{z}|\boldsymbol{x},\sigma^2,\boldsymbol{\theta})}{\prod_{i=1}^n p(z_i|\boldsymbol{x},\sigma^2,\boldsymbol{\theta})} = \frac{\phi_n(\boldsymbol{z};\mathbf{0},R(\boldsymbol{x},\boldsymbol{\theta}))}{\prod_{i=1}^n \phi_1(z_i)}$$

where

$$R(\boldsymbol{x},\boldsymbol{\theta}) = S(\boldsymbol{x},\boldsymbol{\theta})\left(I + B_{\boldsymbol{\zeta}}(\boldsymbol{x})P(\boldsymbol{\theta})^{-1}B_{\boldsymbol{\zeta}}(\boldsymbol{x})^\top\right)S(\boldsymbol{x},\boldsymbol{\theta}),$$

What makes it possible?

- Marginal Distribution Calibration

: Marginal calibration is where the average of nature's true distribution show match the average forecast distribution.

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} H_t = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} F_t,$$

In practice, while H_t is unknown, a draw from it is observed and H_t can be replaced with a point mass at this value, so that $\frac{1}{T}\sum_{t=1}^{T} H_t$ is the empirical distribution function.

- How to achieve marginal calibration? Use Copula!
- Marginal Calibration property holds for the regression

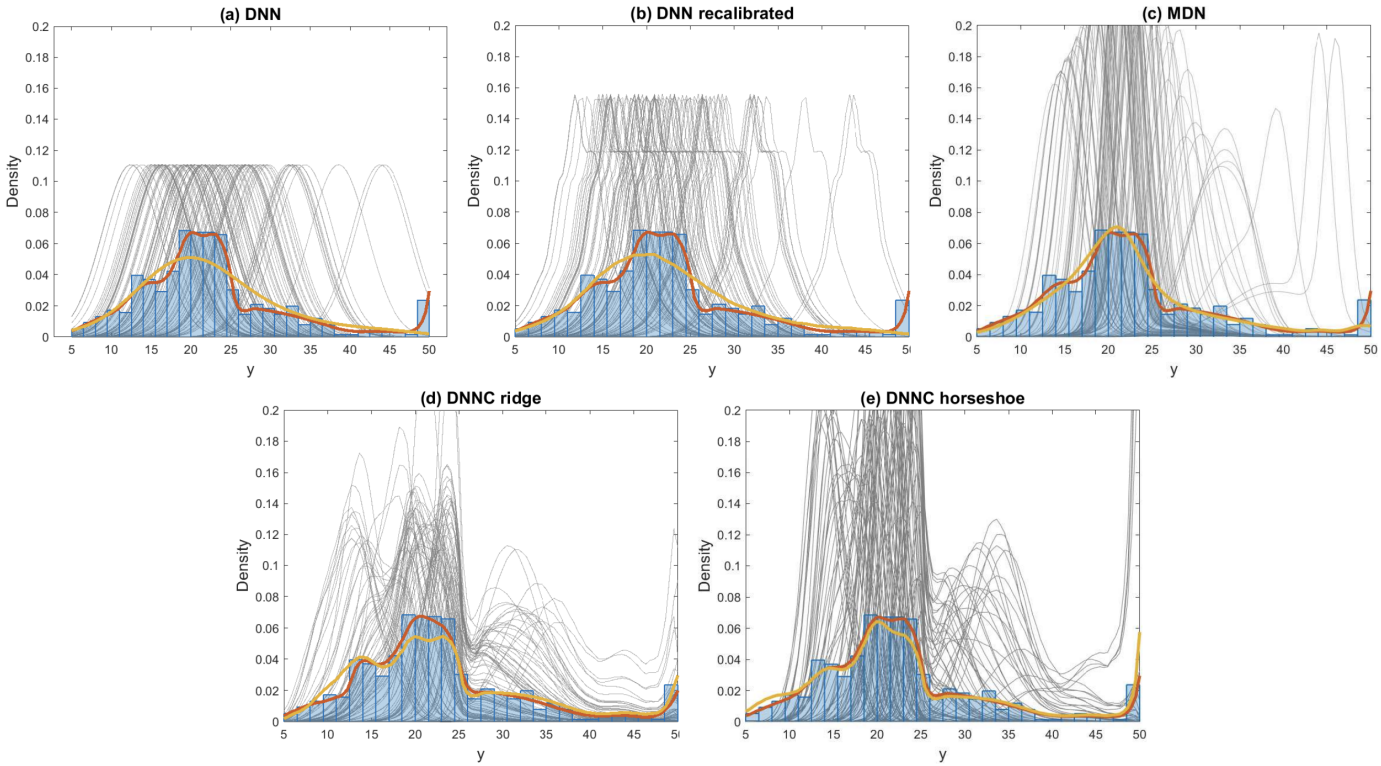
$$\frac{1}{n}\sum \hat{p}(\rho|x_i) \approx \int \hat{p}(\rho|x_i)p(x_i)dx = p(\rho)$$

Four Benchmarks

- **DNN**
  Feed-forward network with the same architecture above, but applied directly to the response data y with $B_w(x)$ with ridge prior to β
- **DNN recallibrated**
  Recalibration of the predictive densities from the DNN
- **MDN**
  Mixture density network implemented in the R-package CaDENCE

Application to Boston Housing Price

Predictive densities for the Boston housing data



A key strength of the regression copula modelling approach is that the entire predictive distribution—including higher order moments—can vary with feature values.