Regression recalibration

Apostolos Psaros

1 Probabilistic calibration

Consider input-output random variables following a distribution $X, Y \sim \mathbb{P}$ and T i.i.d. realizations $\{x_t, y_t\}_{t=1}^T$. The true data-generating process is defined by $G_t(y) = P(Y_t \leq y)$ for every t. Consider also a probabilistic model H which accepts x_t as input and outputs a CDF $F_t(y)$. The model is probabilistically calibrated if

$$\frac{1}{T}\sum_{t=1}^{T} G_t \circ F_t^{-1}(p) \to p \tag{1}$$

for all $p \in (0,1)$ and for $T \to \infty$. Eq. (1) is equivalent to

$$\frac{1}{T}\sum_{t=1}^{T} P(Y_t \leq F_t^{-1}(p)) \to p \tag{2}$$

and to

$$P(Y \leq F_X^{-1}(p)) = p \tag{3}$$

Also, because $T \to \infty$

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(y_t \leq F_t^{-1}(p)) \to p \tag{4}$$

where the random variable Y_t has been replaced by the realization y_t. Eq. (4) can also be written as

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(F_t(y_t) \leq p) \to p \tag{5}$$

2 Theoretical recalibration

Suppose that the model is miscalibrated and $P(Y \leq F_X^{-1}(p))$ in Eq. (3) is not equal to p, but to some $Q(p)$, i.e.,

$$Q(p) = P(Y \leq F_X^{-1}(p)) \neq p \tag{6}$$

Then consider $p = Q^{-1}(p')$ for arbitrary $p' \in (0,1)$ and

$$p' = Q \circ Q^{-1}(p') = P(Y \leq F_X^{-1} \circ Q^{-1}(p')) \tag{7}$$

Then by using the property $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$

$$p' = P(Y \leq (Q \circ F_X)^{-1}(p')) \tag{8}$$

Therefore, $Q \circ F_X$ yields a perfectly calibrated model. In other words, for recalibrating our model we train an auxiliary model R that approximates Q. Then we apply R to any output CDF from our model.

3 Approximate recalibration

Given the realizations $\{x_t, y_t\}_{t=1}^{T}$ and the CDFs $F_t(y)$ for every t, we can construct an estimate of $Q(p)$ as

$$Q_T(p) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(F_t(y_t) \leq p) \tag{9}$$

The aim of recalibration is to fit an auxiliary model R to the estimate $Q_T(p)$.

What is the training data? How can we combine the realizations and the CDFs in order to obtain useful data for training R? The input should be some confidence level p_t and the output should be $Q_T(p_t)$. Our training set is therefore

$$\mathcal{D}_R = \{p_t = F_t(y_t), Q_T(p_t)\} \tag{10}$$

where, for clarity,

$$Q_T(p_t) = \frac{1}{T}\sum_{t'=1}^{T} \mathbf{1}(F'_t(y'_t) \leq p_t) \tag{11}$$

We can use an isotonic regression model for R. See Fig. 1 for an example of isotonic regression in a context other than recalibration. See also Fig. 2 for a recalibration example from Kuleshov et al. (2018).

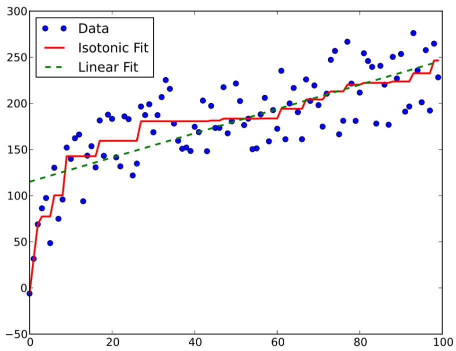

Fig. 1. Figure from Wikipedia. Example of isotonic regression.

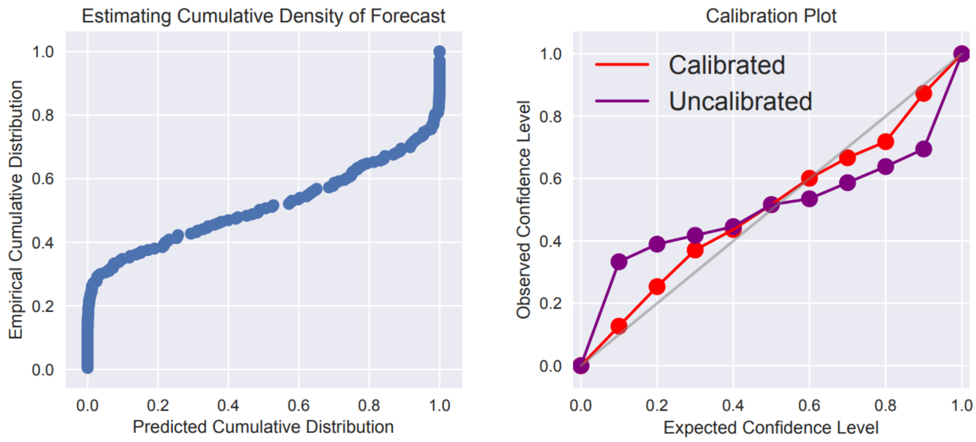

Fig. 2. Figure 3 from Kuleshov et al. (2018). Left side: p_t values on the x axis and $Q_T(p_t)$ values on the y axis. Right side: Calibration diagram after recalibration with R.

References

Kuleshov, V., Fenner, N., and Ermon, S. (2018). "Accurate Uncertainties for Deep Learning Using Calibrated Regression". *arXiv preprint arXiv:1807.00263.*