Prediction with Confidence: A General and Unified Framework for Prediction

Jieli Shen

In collaboration with Regina Liu & Min-ge Xie
Rutgers, The State University of New Jersey

2016 Joint Statistical Meeting, Chicago
August 1, 2016

Research supported in part by grants from NSF and DHS

General setup

- Observed: $\mathbf{Y}_n \sim G_\theta(\cdot)$
- Future: $Y^* \sim F_\theta(\cdot)$

Inference goal

- Make prediction about the unknown Y^*
- Quantify the prediction uncertainty using a *distribution function*

Existing methods

- Plug-in predictive distribution: $F_{\hat{\theta}}(y^*)$
- Bayesian predictive distribution: $\int_\Theta F_\theta(y^*)p(\theta|\mathbf{y}_n)d\theta$.
- Others

Formal Definition

$\mathcal{Y}^* =$ sample space of Y^*; $\mathcal{Y}^n =$ sample space of $\mathbf{Y}_n$

Definition - Predictive distribution

A function $Q(\cdot;\cdot)$ on $\mathcal{Y}^* \times \mathcal{Y}^n \longrightarrow [0, 1]$ is called a predictive distribution for a new observation Y^*, if it follows the following two requirements:

R1) *For each given $\mathbf{Y}_n = \mathbf{y}_n \in \mathcal{Y}^n$, $Q(\cdot; \mathbf{y}_n)$ is a c.d.f on $\mathcal{Y}^*$;*

R2) *The joint distribution of $Q(Y^*; \mathbf{Y}_n)$, as a function of both random sample Y^* and $\mathbf{Y}_n$, satisfies the following equation:*

$$\mathbb{P}_{\mathbb{J}}(Q(Y^*; \mathbf{Y}_n) \leqslant \alpha) = \alpha, \quad \textit{for} \quad \forall\alpha \in (0, 1).$$

We propose constructing predictive distribution based on **confidence distributions** (CDs) with (1) sound theoretical developments and (2) simple simulation algorithm.

A Brief Introduction of Confidence Distribution

Parameter estimation

- Point estimate
- Interval estimate
- Distribution estimate (e.g., Bayesian posterior, bootstrap distribution, **confidence distribution**)

Example: $X_1, \ldots, X_n \sim N(\mu, 1)$

- Point estimate: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Interval estimate: $(\bar{X}_n - 1.96/\sqrt{n}, \bar{X}_n + 1.96/\sqrt{n})$
- Distribution estimate: $N(\bar{X}_n, 1/n)$

A **confidence distribution** *(CD) is a sample-dependent distribution function that can represent confidence intervals (regions) of all levels for an unknown parameter. [Xie & Singh, 2013; Schweder & Hjort, 2014]*

A CD-Based Formulation

Suppose $H(\cdot; \mathbf{y}_n)$ is a CD for θ based on the sample data $\mathbf{y}_n$, we propose to construct a distribution function for the prediction by

$$Q(y^*; \mathbf{y}_n) = \int_\Theta F_\xi(y^*) dH(\xi; \mathbf{y}_n).$$

Computing algorithm

1. Simulate a CD-random variable $\xi | \mathbf{y}_n \sim H(\cdot; \mathbf{y}_n)$ and then simulate $y^* | \xi \sim f_\xi(\cdot)$
2. Repeat the procedure for B times and obtain B copies of y^*
3. Use the B copies of y^* to construct a predictive distribution (intervals).

Main Results (in Plain Words)

Result 1 - Asymptotic frequentist coverage

$Q(Y^*; \mathbf{Y}_n) \to U[0, 1]$ *as* $n \to \infty$.

Result 2 - Exact frequentist coverage

Under a pivot-based construction, $Q(Y^*; \mathbf{Y}_n) \sim U[0, 1]$.

Result 3 - Comparison of CD-based predictive distributions

A more precise CD leads to a more precise CD-based predictive distributions.

Result 4 - Kullback-Leibler superiority

$Q(\cdot, \mathbf{Y}_n)$ *is better than* $F_{\hat{\theta}}(\cdot)$ *measured by the average Kullback-Leibler divergence to the true* $F_{\theta_0}(\cdot)$.