Prediction Regions

November 23, 2023

Summary

- We consider intervals based on observable random variables that contain an unobservable random variable with probability at least $1 - \alpha$.
- In the case of a normal sample of size $n + 1$ with only n variables observable, we construct the Student t prediction interval for the unobservable variable.
- For a sample of size $n + 1$ from a continuous distribution we show how the order statistics can be used to give a distribution-free prediction interval.
- The Bayesian formulation is based on the posterior predictive distribution which is the conditional distribution of the unobservable variable given the observable variables.
- The Bayesian prediction interval is derived for the normal model with a normal prior.

Definitions and Examples

- Prediction regions apply to situations in which we want to predict the value of a random Y that is not observable, i.e. evaluate a random interval I based on the observed variables such that

$$P(Y \in I) \geq 1 - \alpha$$

  i.e. I contains the unknown value Y with prescribed probability $1 - \alpha$.
- We define a level $1 - \alpha$ prediction interval as an interval $[\underline{Y}, \overline{Y}]$ based on the observable data X such that

$$P([\underline{Y} \leq Y \leq \overline{Y}]) \geq 1 - \alpha.$$

- Problem of finding prediction intervals is similar to finding confidence intervals using a pivot.
- Examples:
  - For instance, a doctor administering a treatment with delayed effect will give patients a time interval $[\underline{Y}, \overline{Y}]$ in which the treatment is likely to take effect.
  - Similarly, we may want an interval for the future GPA of a student or a future value of a portfolio.

A simple prediction problem

- Let $\mathbf{X} = (X_1, ..., X_n)$ be i.i.d. $N(\mu, \sigma^2)$.
- We want a prediction interval for $Y = X_{n+1}$.
- The predictor based $\mathbf{X}$, say $\hat{Y}(\mathbf{X})$, must be independent of Y.
- The mean square prediction error is then

$$MSPE = E(\hat{Y} - \mu)^2 + \sigma^2$$

We note that if $\hat{Y}$ is considered to be an estimate of μ then

$$MSPE = MSE + \sigma^2.$$

We call a predictor $\hat{Y}$ unbiased if $E\hat{Y} = EY$ and in the class of unbiased predictors the optimal one is $\hat{Y} = \bar{X}$.

The (Student) t Prediction Interval

- Prediction interval is based on a pivot constructed as follows

$$\hat{Y} - Y = \bar{X} - X_{n+1} \sim \mathcal{N}(0, [n^{-1} + 1]\sigma^2).$$

  Moreover, $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2$ is independent of $\bar{X}$ and independent of X_{n+1}. Thus

$$Z_p(Y) = \frac{\hat{Y} - Y}{\sqrt{n^{-1} + 1}\sigma}$$

  and

$$T_p(Y) = \frac{Z_p(Y)}{\sqrt{S^2/\sigma^2}} = \frac{\hat{Y} - Y}{\sqrt{n^{-1} + 1}S}$$

  has the t-student distribution with $n - 1$ degrees of freedom.
- This yields the prediction interval

$$Y = \bar{X} \pm \sqrt{n^{-1} + 1}S\, t_{n-1,1-\alpha/2}.$$

A general approximate approach

- Suppose that the joint distribution of $\mathbf{X}, Y$ is given by $p(\mathbf{x}, y|\theta)$.
- Assume that $\hat{\theta}$ is an estimator of θ based solely on $\mathbf{X}$.
- Consider the conditional distribution $p(y|\mathbf{x};\theta)$ and a center $\hat{Y}(x,\theta)$ of this distribution (mean or median or location). Take $\hat{Y}(x,\hat{\theta})$ as a predictor of Y.
- Consider and a connected set $A_{\theta,\mathbf{x}}$ containing $\hat{Y}(x,\theta)$ (in 1-D an interval) such that
$$P(A_{\theta,\mathbf{x}}|\mathbf{x};\theta) \geq 1-\alpha.$$
- take $A_{\hat{\theta},\mathbf{x}}$ as an approximate prediction region (interval) for Y.
- Take the previous example and explain why it is only an approximate interval.

Bayesian Predictive Distributions

- In the Bayesian setup, let $X_1, ..., X_n$ be observable and X_{n+1} is to be predicted.
- The posterior predictive distribution of X_{n+1} given $\mathbf{X} = \mathbf{x} = (x_1, \ldots, x_n)$, i.e.

$$q(x_{n+1}|\mathbf{x}) \propto E_\Theta \prod_{i=1}^{n+1} p(x_i|\Theta)$$

- Any interval $[\underline{Y}_B, \overline{Y}_B]$ is said to be a level $(1-\alpha)$ Bayesian prediction interval fo $Y = X_{n+1}$ if

$$Q(\underline{Y}_B \leq Y \leq \overline{Y}_B|\mathbf{x}) \geq 1-\alpha.$$

- As usually, the Bayesian approach can be interpreted as a fully specified stochastic models with an unobservable θ (to be averaged with respect to its distribution). Here we have also unobservable $Y = X_{n+1}$ but this variable is to be predicted not averaged.