Presenting Aleatory and Epistemic Uncertainty in Regression Models

By AY and MW

Notations

Regression: $Y = g(x; \beta) + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$ is considered as the random residual error that consists with (unknown) different proportions of model error (error in model structure), measurement error in $y_1, \ldots, y_n$ and measurement error in $x_1, \ldots, x_n$.

Notes:

(1) The fourth component, unobserved heterogeneity, sometimes also exists in ε. A simple example of the unobserved heterogeneity is different metallurgical components in the pipe steels from different manufacturers.

(2) In my view, unless extra efforts are made, there is no way to identify and separate the three/ four components in ε. However, sensitivity analysis might be able to identify and justify which component(s) should be further reduced.

(3) The measurement error in x sometimes cause model error (model misspecification) and underestimate of parameter uncertainty.

Suppose a lifetime T is defined by the response Y and a threshold value ζ. Then $T = h(Y)$. Often ζ is a decision parameter and subject to no uncertainty. We omit this parameter in the following discussion.

Setting

Using the given data $(x_i, y_i; i = 1, \ldots, n)$, we have obtained the estimate $\hat{\beta}$ and $\hat{\sigma}^2$. Their parameter uncertainty is also quantified. For a linear regression model, the expressions are well established. That is, the estimator β follows a normal distribution and estimator σ^2 follow a χ^2 distribution. For nonlinear regression model, similar results can be obtained using the delta approach (originated from Taylor series expansion of function $g(x)$) and the large sample theory. Asymptotically, β also follows a normal distribution and σ^2 a χ^2 distribution.

In this model setting, ε is called aleatory uncertainty, and the parameter uncertainty is the epistemic uncertainty.

Presentation of Results

The prediction of Y can be presented in several ways.

(1) Y as a function of x. To represent the uncertainty, five curves are often included. They are, from top to bottom, 95 percentile upper bound for response $Y|x$, 95 percentile upper bound for the mean response $E[Y|x]$, mean response $E[Y|x]$, 5 percentile lower bound for the mean response $E[Y|x]$, and 5 percentile lower bound for the response $Y|x$.

(2) Probability distribution of $Y|x$. It is often represented by the cumulative distribution function (CDF), rather than the probability density function. There are also two approaches to present the CDF.

a. Integrated CDF: For linear regression, it is just a CDF of t distribution that considers both aleatory uncertainty ε and epistemic uncertainty in parameter.

b. Aleatory CDF plus epistemic bounds: A normal CDF for $Y|(x, \beta, \sigma)$ as the central curve, plus 95th- and 5th-percentile bounds of the CDF when the parameter uncertainty of β and σ is included. It can be shown that the average of those epistemic plots should equal the integrated CDF in (2a).

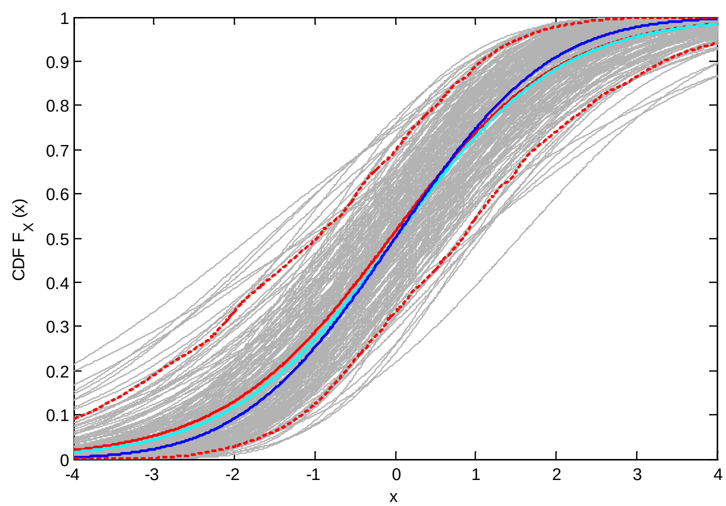

Figure 1: Blue curve is the aleatory CDF (i.e., normal distribution); the cyan curve is the integrated CDF (i.e., the t-distribution); the red curve of solid line is the average of the hair plots, which overlaps the cyan curve (i.e., a sample average approximate of t distribution); the red curves of broken line is the 5th and 95th epistemic bounds of the CDF.

c. Aleatory quantile plus epistemic bounds: A normal quantile for $Y|(x, \beta, \sigma)$ as the central curve, plus 95th- and 5th-percentile bounds of the quantile when the parameter uncertainty of β and σ is included. It can be shown that the average of those epistemic plots should equal the integrated quantile of the t distribution. Although a quantile is an inverse function of the CDF, the probability bounds of the quantile are not the inverse of the corresponding bounds of the CDF. Alternatively, the quantile curve with a specific percentage of confidence (one-sided bound) can be plotted.

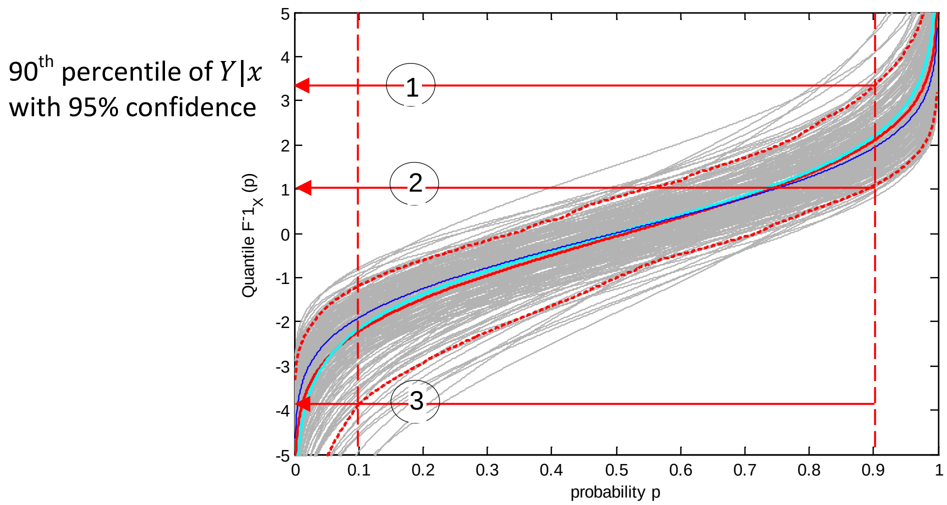

Figure 2: color convention is the same as in Figure 1.

(3) 90th percentile of $Y|x$ with 95% confidence, or pth percentile of $Y|x$ with $q\%$ confidence. This can be directly read from Figure 2, if the 95% one-sided bound are given.

It should be clear now that the prediction of the lifetime T is exactly the same as that of Y. The confusion arose probably because we often use (1) for Y and (2) for T.

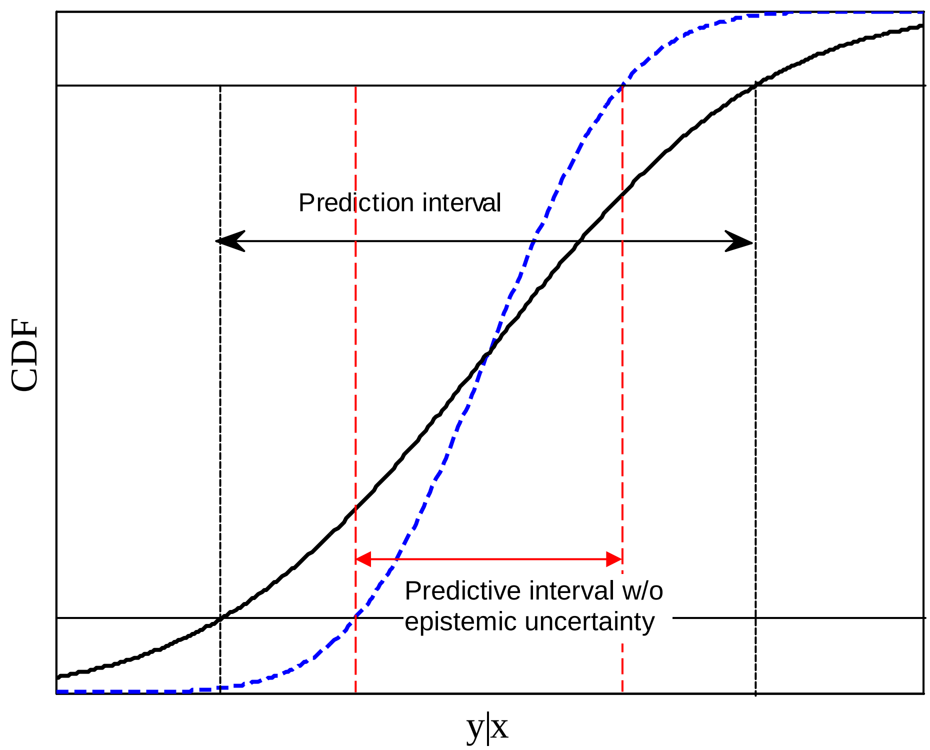

Figure 3: Can you add another curve for the mean of y|x with consideration of epistemic uncertainty? It should be even steeper than the blue line.

Interpretation

There are still several issues to be addressed.

(1) Since we decided to include model error into the residual error ε term, how do we explain the alteatory CDF of $Y|x$ in (2b)? Although Der Kiureghian and Ditlevsen suggested that categorization of the aleatory uncertainty and epistemic uncertainty be based on our answer to whether the uncertainty can be further reduced through efforts in a reasonable time, people often interpret the aleatory CDF as the objective probability, something like the property or propensity of the subject we analyze, similar to the mass of a coin. However, the inclusion of model error in ε contaminates this interpretation.

(2) Use of terminology. Regarding epistemic uncertainty, shall we use confidence interval or credibility interval? In my view, we use avoid using the frequentist terminology because confidence interval actually describes only the inference procedure, not the quantity of our interest.

(3) I believe that the pth percentile with $q\%$ confidence represents one kind of tolerance limit, it is a one-sided tolerance limit. According to Cambridge Dictionary of Statistics (Everitt, 2006, 3rd

ed) a tolerance interval is the statistical interval that contain at least a specified proportion of a population either on average, or else with a stated confidence value. If we look at Figure 2, we can actually read predictive interval (also refer to Figure 3), which is the distance (not shown in Figure 2) dictated by the cyan line. On the other hand, the y-axis value pointed by arrow 1 is 90th percentile with 95% confidence. The distance between 1 and 2 is the (two-sided) *confidence interval* for 90th percentile with 90% (not a mistake!) confidence. But the distance between 1 and 3 is the (two-sided) *tolerance interval* for 90th percentile with 90% confidence. However, I think the value at 1 also represents a one-sided upper tolerance limit if we read the definition of tolerance limit. Another thing we should be clear now is that the *predictive interval* represents only the tolerance limit with 50% confidence.

Appendix A: Interpretation from Bayesian Statistics

Take the univariate variable case for an example. Suppose $X \sim N(\mu, \sigma^2)$ and the distribution parameters μ and σ^2 or $\lambda = \sigma^{-2}$ (usually called precision) are to be estimated from a set of independent observations, collectively called data $\mathcal{D} = \{x_1, \dots, x_n\}$. In Bayesian statistics, the parameters are treated as unknown quantities and also modeled as two random variables with prior distributions $\pi(\mu)$ and $\pi(\lambda)$. Assume both priors to be noninformative, i.e., $\pi(\mu) \propto 1$ for $-\infty < \mu < \infty$ and $\pi(\lambda) \propto \lambda^{-1}$ for $\lambda > 0$. (This is also called the reference prior for μ and λ)

Denote the sample mean and sample variance of the data as $m = \frac{1}{n}\sum x_i$ and $s^2 = \frac{1}{n-1}\sum(x_i - m)^2$, respectively. Then the posterior distributions of the two parameters are

$$p(\mu, \lambda) \propto \pi(\mu)\pi(\lambda)\lambda^{\frac{n}{2}} \exp\left(-\frac{\lambda}{2}\sum(x_i - \mu)^2\right)$$

Note that

$$\sum(x_i - \mu)^2 = \sum[(\mu - m) - (x_i - m)]^2 = \sum(\mu - m)^2 + \sum(x_i - m)^2 = n(\mu - m)^2 + (n-1)s^2$$

Hence,

$$p(\mu, \lambda) \propto \left\{\lambda^{\frac{1}{2}} \exp\left(-\frac{1}{2}n\lambda(\mu - m)^2\right)\right\} \times \left\{\lambda^{\frac{n-1}{2}-1} \exp\left(-\frac{1}{2}n\lambda s^2\right)\right\}$$

This joint posterior can be expressed as

$$p(\mu, \lambda) = p(\lambda) \times p(\mu|\lambda)$$

where

$$p(\lambda) = Ga\left(\lambda\middle|\frac{n-1}{2}, \frac{n}{2}s^2\right)$$

and

$$p(\mu|\lambda) = N(\mu|m, n\lambda)$$

The marginal posterior for μ is actually a t distribution as

$$p(\mu) = t_{n-1}\left(\mu\middle|m, \frac{s^2}{n-1}\right)$$

The posterior predictive distribution is

$$p(x|D) = t_{n-1}\left(x\middle|m, \left(1+\frac{1}{n}\right)s^2\right)$$

These all are the same as the frequentist results.

Note that the posterior predictive is derived from the following integral

$$p(x|D) = \int \color{blue}{p(x|\mu, \lambda, D)}\color{red}{p(\mu, \lambda|D)}\mathrm{d}\mu\mathrm{d}\lambda$$

in which the blue colored $\color{blue}{p(x|\mu, \lambda, D)}$ is actually the aleatory uncertainty while the red colored $\color{red}{p(\mu, \lambda|D)}$ is the epistemic uncertainty.

In Figures 1 and 2, I used $m = 0, s = 1$ and $n = 10$ for the simulation. Nsim = 200.