Bayesian Analysis of a NIST Dataset: Spectrophotometer Calibration With Between-Run Variability

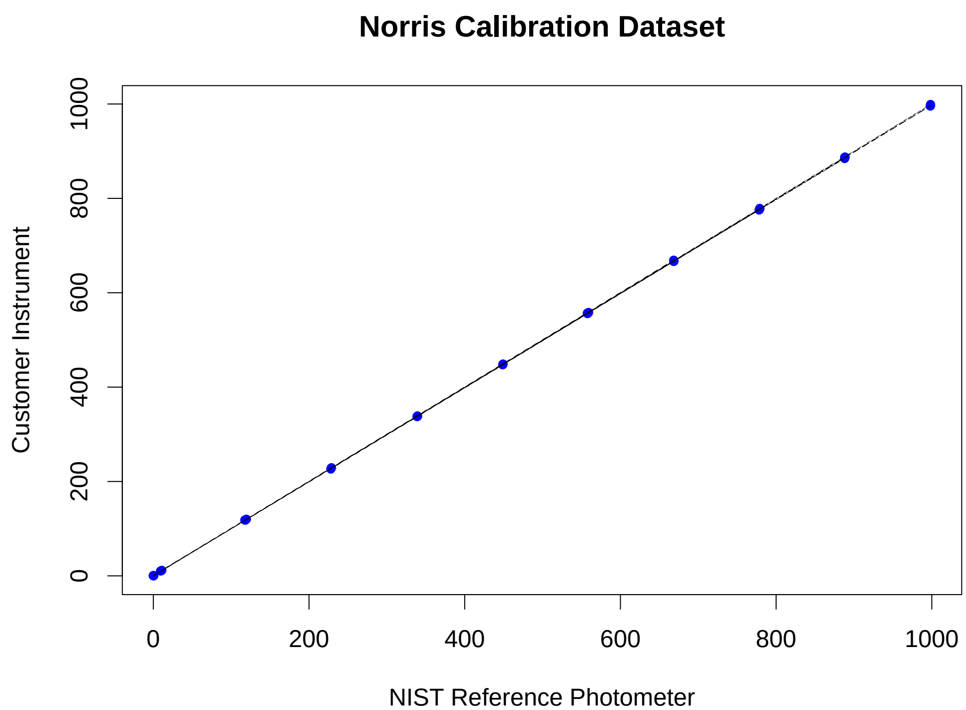
Bayesian Metrology Project
Statistical Engineering Division
National Institute of Standards and Technology
Gaithersburg, MD

Introduction

Statistical calibration is central to the standards mission of NIST. Calibration problems arise when there are two ways of measuring a quantity. One method is highly precise; in the example to be considered below it involves a unique reference instrument that is a national resource. The other method is somewhat less precise, but commonly available. Based on data obtained with these two instruments together, the goal is to predict from future measurements, using the less precise instrument, what the corresponding value would be for the reference instrument.

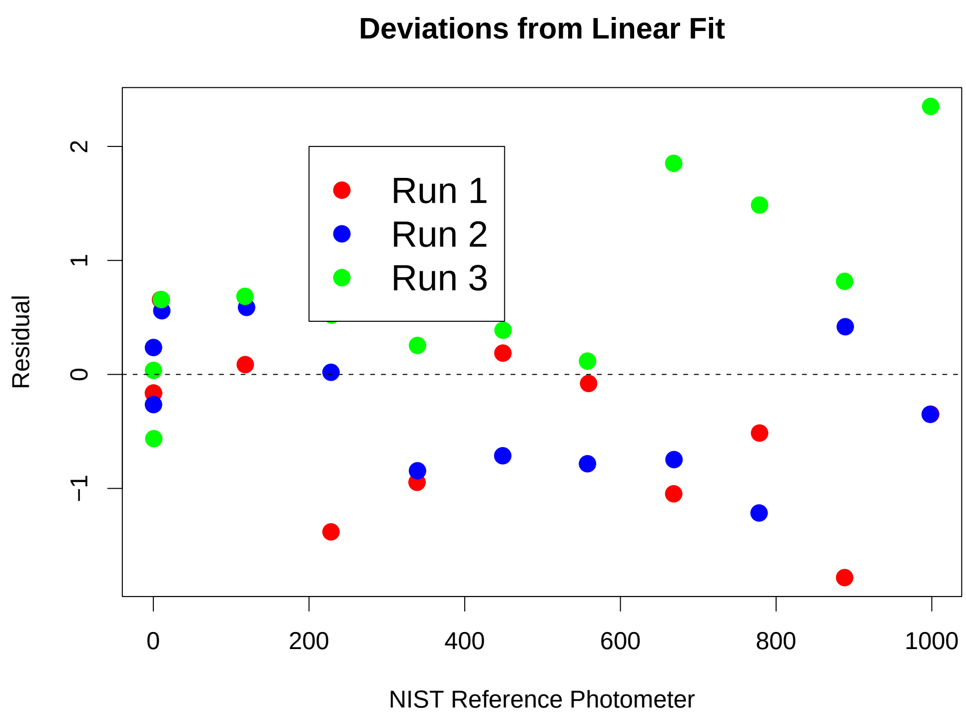
Although this problem is natural to consider from a Bayesian point of view, standard frequentist approaches are conceptually awkward, and limited in scope. For example, the between-run variability in this spectrophotometer data can not be handled naturally in the standard calibration framework.

Comparison of the Standard Reference Photometer With a Customer Instrument



The above plot illustrates an actual dataset comparing the NIST Standard Reference Photometer with an instrument which was sent by an industrial customer to NIST Calibration Services. Note the nearly precise linear fit. Still, the fit is not exact, as the next slide will show.

Deviations from Linear Fit, Illustrating Between-Run Variability



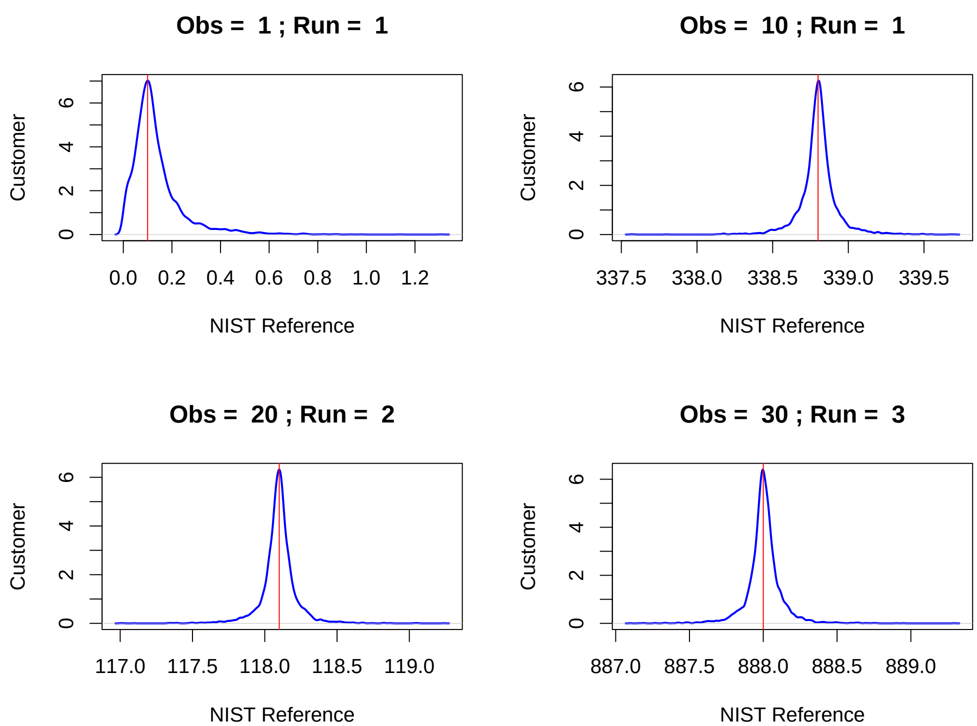
The above plot illustrates the sources of variability in this precision measurement problem. The measurements were made in three runs, and from the colored symbols one can see that there are systematic shifts in the uncertainty due to run-to-run variability. We will propose and fit a Bayesian model which takes this source of variability into account.

Bayesian Hierarchical Model

$$
\begin{aligned}
p(y_i|\mu_i, \sigma_w^2) &\sim \mathsf{N}\left(\mu_\mathsf{i}, \sigma_\mathsf{w}^2\right) \\
\mu_i &= \alpha + \beta x_i + r_i \\
p(r_i|\sigma_r^2) &\sim \mathsf{N}\left(0, \sigma_\mathsf{r}^2\right) \\
p(\alpha) &= 1 \\
p(\beta) &= 1 \\
p(x_i|a, b) &\sim \mathsf{U}\left(a, b\right) \\
p(\sigma_r) &\sim 1 \\
p(\sigma_w) &\sim 1/\sigma_w
\end{aligned}
$$

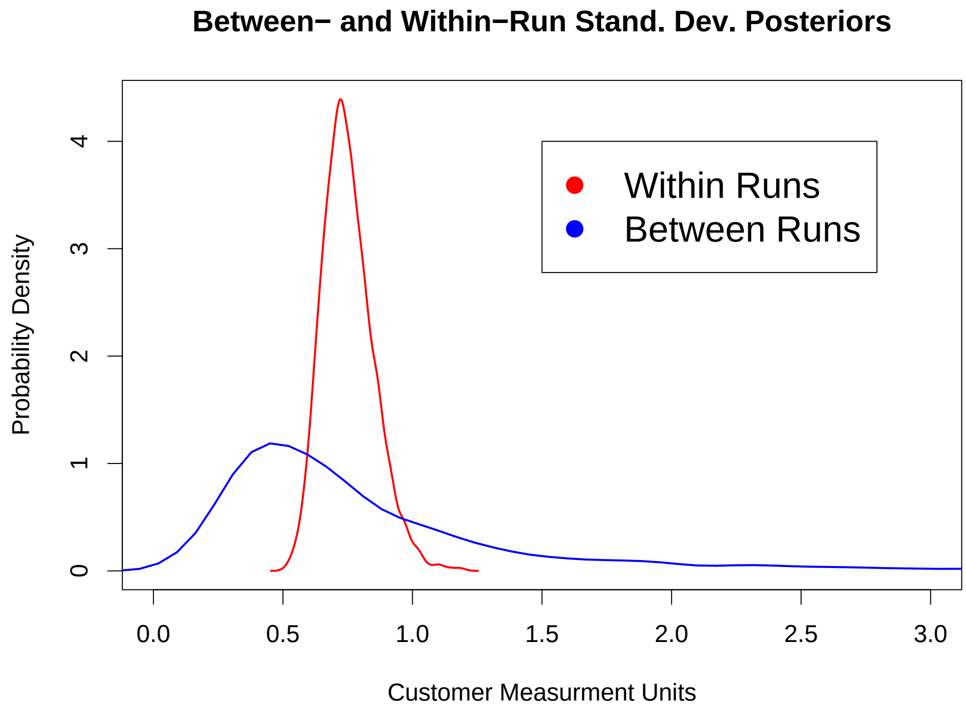
For the most part, this is a standard hierarchical linear model, with noninformative priors. An informative prior on the reference measurements x_i is necessary in order ensure that the posterior is proper; we take this prior to be uniform over the range of xs from the calibration experiment. The r_i represent the variability between the three runs. Given a future customer measurement, say y_*, the uncertainty in the corresponding (unobserved) reference photometer measurement will be obtained from its posterior predictive distribution.

Example Posterior Distributions for x_*s Corresponding to Future Customer y_*s.



The above posterior densities provide the uncertainties for four typical x_* reference photometer measurements, corresponding to four future customer y_*s. For convenience, these hypothetical future y_* values have been chosen to coincide with data values y_i, so that we can compare the predictive distributions of x_* with the actual observed reference photometer measurements. These actual x_i values are indicated by the vertical lines.

Posterior Comparison of Sources of Variability



Above are displays of the posterior densities for the between-measurement and between-run components of uncertainty. Note that though the mode of the between-run standard deviation is lower than that of the measurement uncertainty, there is considerable variability between runs. We only observed three runs in this dataset, so this potentially important source is variability is poorly estimated; hence the rather diffuse density.

Between-Run Uncertainty, Concluded

We could reduce our uncertainty in future x_*s by using prior information on between-run variability. This should be easy to do. For example, one might base an informative prior on historical data with multiple runs on similar instruments.