Estimating Continuous Distributions by Quantifying Errors in Probability Judgments for Fixed Values

Asa Palley, Saurabh Bansal

Estimating Continuous Distributions by Quantifying Errors in Probability Judgments for Fixed Values

KEYWORDS: Demand and Yield Forecasting, Judgment Errors, Estimating Distributions.

INTRODUCTION: In a number of operational situations, historical data is not available or may not be directly relevant for obtaining probability distributions of key uncertainties. This is especially true in today's market where firms frequently launch new products, often with very short life cycles. For example, the demands for new electronic products or seasonal fashion goods must be estimated in order to make inventory decisions, but prior data is usually not available for this task (Gaur et al. 2007, Akcay et al. 2011). Similarly, production yield distributions for new items may be unknown due to added product features or novel production technology, yet an accurate determination of this uncertainty is necessary for making production lot decisions (Bansal et al. 2014).

In each of these settings, a decision maker (DM) must rely on expert assessments (which may be subjective, driven by a formal quantitative model, or both) to determine the distribution of a random variable. Experts possess domain-specific knowledge and experience that they can use to assess various aspects of a distribution. Although a DM could ask an expert for direct estimates of the mean and standard deviation, existing literature suggests that he should instead elicit her judgments for a specific discrete set of points on the distribution and then use these points to deduce its moments. For example, consider an operations manager who asks a market researcher to specify the probability that sales of a product in the next month will be less than or equal to 90K units, 110K units, and 130K units and she states probabilities of 0.15, 0.55, and 0.80. How should the operations manager estimate the mean and standard deviation of the underlying distribution using these judgments?

To answer this question, we focus on a primary issue: *the expert's judgments may be prone to errors.* These errors may result from a variety of sources, such as an expert's limited understanding of the physical or market system they are forecasting, limited experience or data availability for forming a judgment, inherent uncertainty in the estimation of the distribution, or imperfect translation of a mental model into probabilities. Bansal et al. (2014) show with quantile judgments collected during an industry application that the estimates of the mean and standard deviation obtained when ignoring errors are only half as reliable as estimates obtained when the errors are incorporated appropriately. Akcay et al. (2011) establish that this loss in reliability of parameter estimates for demand distributions can increase the cost of an inventory system by 8 to 10%. Furthermore, ignoring these errors leads to an implicit assumption that all judgments are equally skillful even though it may be known *a priori* that some are more accurate than others.

RESEARCH FOCUS: We develop an approach to obtain estimates of distribution parameters from cumulative probability function (CDF) judgments that may contain judgmental errors. Specifically, we estimate the mean and standard deviation as a weighted linear combination of the fixed values that correspond to these cumulative probabilities, *where the weights are specific to the probabilities provided for these variable values and the expert's judgmental errors.* We show how the structure of these errors can be quantified with calibration data using a scale-free model of judgmental errors, and how a DM should optimally weight the expert's judgments as a function of this structure. We believe that this is the first attempt to solve this problem in an analytical fashion.

METHOD: Our approach is comprised of two steps:

In **Step 1**, the DM quantifies the expert's judgmental errors using her responses on a set of the calibration distributions. These distributions are obtained using past data on the realizations of random variables that have a long recorded history, such as products that have been sold or produced repeatedly in the past. For each calibration distribution, the expert provides her CDF judgments for a set of fixed points. By comparing these probability judgments with the true probabilities for the fixed points available from the historical data, the DM can quantify the expert's judgmental errors. These judgmental errors are decomposed into two parts: bias and residual noise.

The bias measures the expert's average under- or over-estimation of probabilities at different parts of the distribution. Prior literature has shown that experts tend to overestimate probabilities in the left tail and underestimate probabilities in the right tail. To account for these potential biases, we regress the true probabilities on a natural cubic spline of the stated probabilities. The resulting spline uniquely transforms the stated probability into an unbiased estimate of the true probability. The DM then compares these unbiased estimates with the true probabilities to obtain the residual judgment errors, which are used to estimate a variance-covariance matrix of the standardized residual errors.

In **Step 2**, the DM uses the structure of the judgmental errors quantified in step 1 to estimate future distributions. First, the DM obtains the expert's CDF judgments at a set of fixed values on a new distribution. Next, using the spline identified in Step 1, he computes the unbiased probability judgments. The variance-covariance matrix of residual errors and the unbiased probability judgments are then fed into an optimization model. The optimization problem estimates the mean and standard deviation of the distribution as a weighted linear combination of the fixed values, selecting weights that minimize the variance in the estimates subject to an unbiasedness constraint. This problem has a unique closed-form solution for any location-scale distribution.

STRUCTURAL RESULTS: We document some important structural properties of the optimal weights. Notably, the weights for estimating the mean add up to 1 and the weights for estimating the standard deviation add up to 0. In addition, by quantifying the uncertainty in the estimates, we establish a direct correspondence between the estimates from the method and an equivalent sample size of observed data. This provides an objective way to rank order experts and quantify their relative expertise in estimating demand or yield distributions.

EXPERIMENTAL VALIDATION: Finally, we test our approach and demonstrate its application and benefits using data collected in an experimental study. Results show that our approach can improve the estimation of the mean by more than 20% and the standard deviation by more than 50%. The benchmarks discussed in Akcay et al. (2011) for the benefit of improved estimation of demand distributions imply that our approach can reduce inventory costs in several systems.

REFERENCES:

Akcay, A., Biller, B., Tayur, S. 2011. Improved Inventory Targets in the Presence of Limited Historical Demand Data. *Manufacturing and Service Operations Management* 13(3); 297-309.

Bansal, S., Gutierrez, G.J., Keiser, J.R. 2015. Using Expert Assessments to Estimate Probability Distributions. *Working Paper, University of Texas at Austin, McCombs School of Business.*

Gaur, V., Kesavan, S., Raman, A., Fisher, M.L. 2007. Estimating Demand Uncertainty Using Judgmental Forecasts. *Manufacturing and Service Operations Management* 9(4); 480-491.