Estimating Continuous Distributions by Quantifying Errors in Probability Judgments

Asa B. Palley, Saurabh Bansal

Duke University, The Fuqua School of Business, Pennsylvania State University, Smeal College of Business

Abstract: In many decision problems, the distribution for a continuous random variable must be obtained from expert judgments. We present a novel method for estimating these distributions when an expert provides probability judgments corresponding to a discrete set of fixed outcome values. The decision maker estimates the mean and standard deviation through linear combinations of these fixed values, where the weights are explicit functions of the cumulative probabilities and the expert's judgmental error structure. We show how these errors can be quantified with calibration data using a scale-free model of judgment errors. We test our approach and demonstrate its benefits using data collected in an experimental study.

Accurately Estimating Uncertainties is Critical to Making Good Decisions

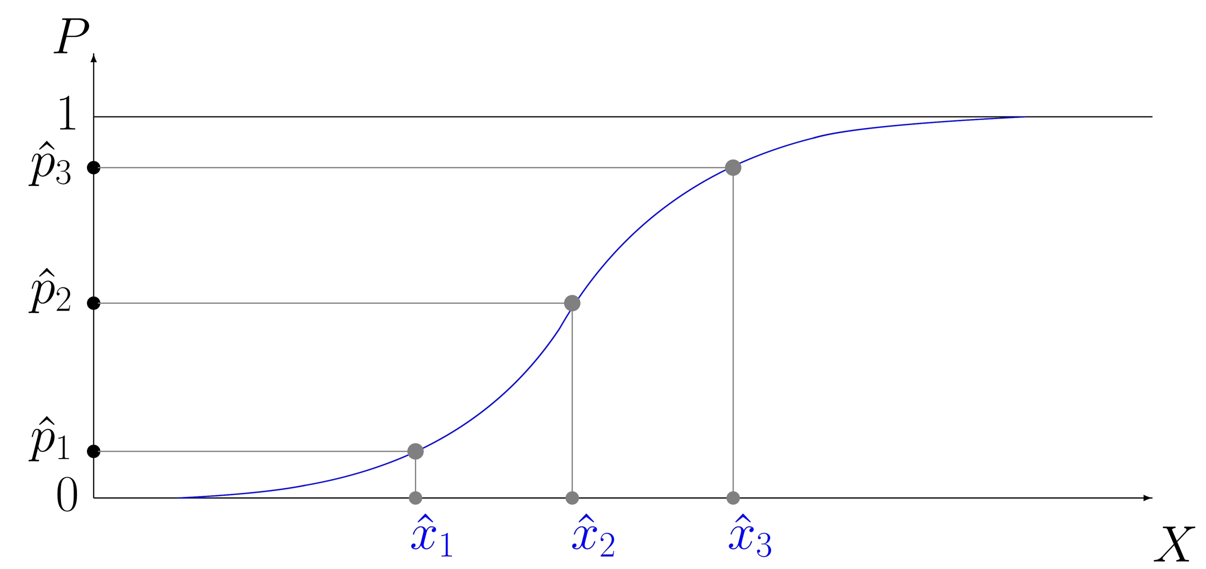
- Inventory decisions with random product demand.

$$q^* = F^{-1}\left(\frac{p-c}{p}\right)$$

- Uncertain returns on investment portfolios.
- Distributions of task durations in project planning.

Eliciting Point Judgments

Expert provides set of judgments on the distribution $\{(\hat{x}_i, \hat{p}_i)\}_{i=1}^N$ such that $\hat{p}_i = \mathbb{P}(X \le \hat{x}_i)$.



Considering Judgmental Errors

- Expert judgments may not be well-calibrated, and can display both over- and under-confidence.
  - 80% CIs provided by financial executives for the stock market contain the realized market return only 36% of the time (Ben-David, Graham, & Harvey, 2013).
  - 90% CIs are over-confident, 70% CIs are well-calibrated, and 50% CIs are under-confident (Budescu & Du, 2007).
- Bansal, Gutierrez, & Keiser (2015) show benefits of accounting for errors in quantile judgments.

Overview of Approach

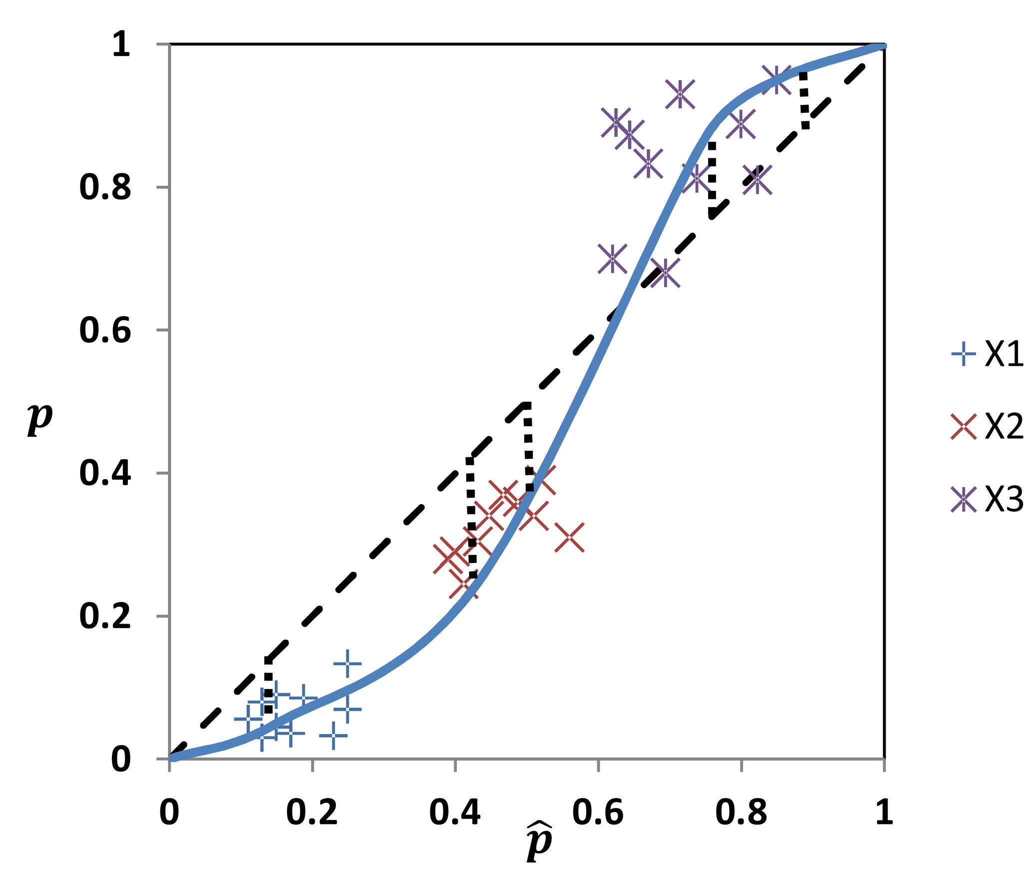
Estimate location and scale parameters $\theta_1 \in \mathbb{R}$ and $\theta_2 \in \mathbb{R}_{++}$ of the distribution of X with standardized CDF $\Phi(z)$ from the fixed values $\hat{\mathbf{x}}$ as a function

$$\hat{\theta}_j = f(\hat{\mathbf{x}}, \hat{\mathbf{p}}, \mathcal{C})$$

of the expert's assessments $(\hat{\mathbf{x}}, \hat{\mathbf{p}})$ and the structure $\mathcal{C}$ of the expert's judgmental errors.

Calibration: Bias in p-domain

For $r = 1, 2, \ldots, R$ known distributions, select a set of fixed values $\hat{x}_{r1}, \hat{x}_{r2}, \ldots, \hat{x}_{rN}$ and elicit corresponding CDF judgments $\hat{p}_{r1}, \hat{p}_{r2}, \ldots, \hat{p}_{rN}$.



Estimate the parameters $\boldsymbol{\gamma}$ of the de-biasing curve $g(\cdot)$ using a scale-free model of judgment error:

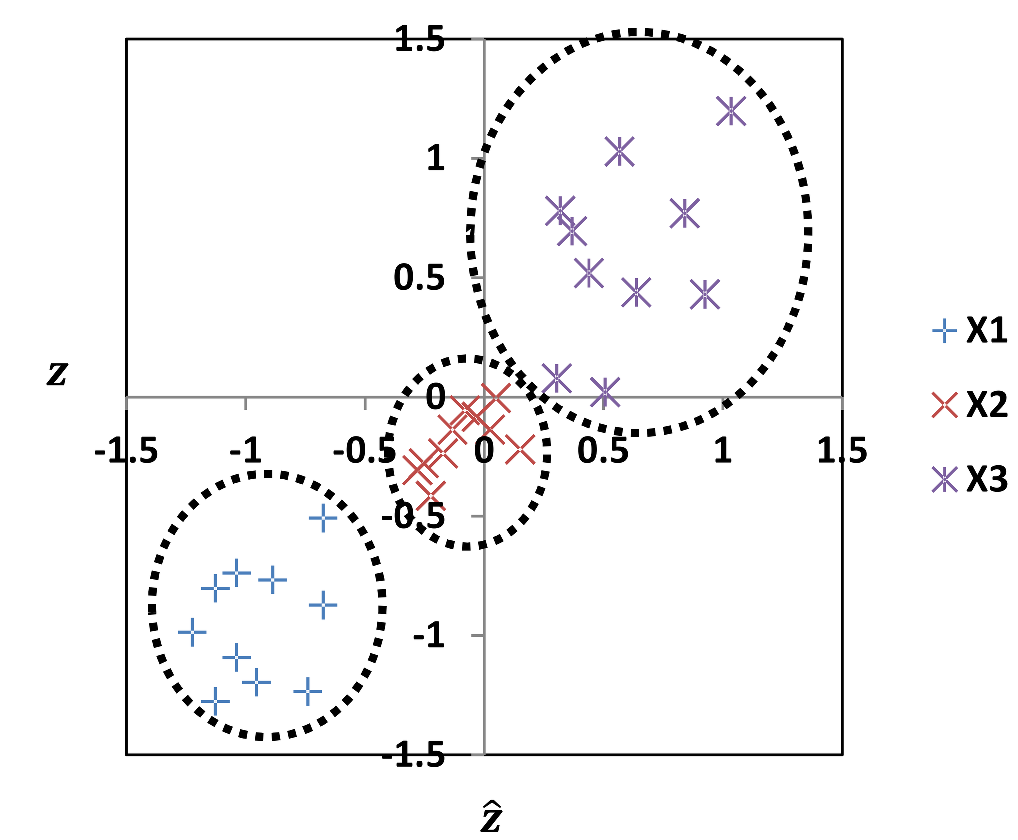
$$g(\hat{p}; \boldsymbol{\gamma}) = p + \psi.$$

Calibration: Error in z-domain

Transform mean-zero residual errors ψ from the p-domain to mean-zero residual errors ε in the z-domain

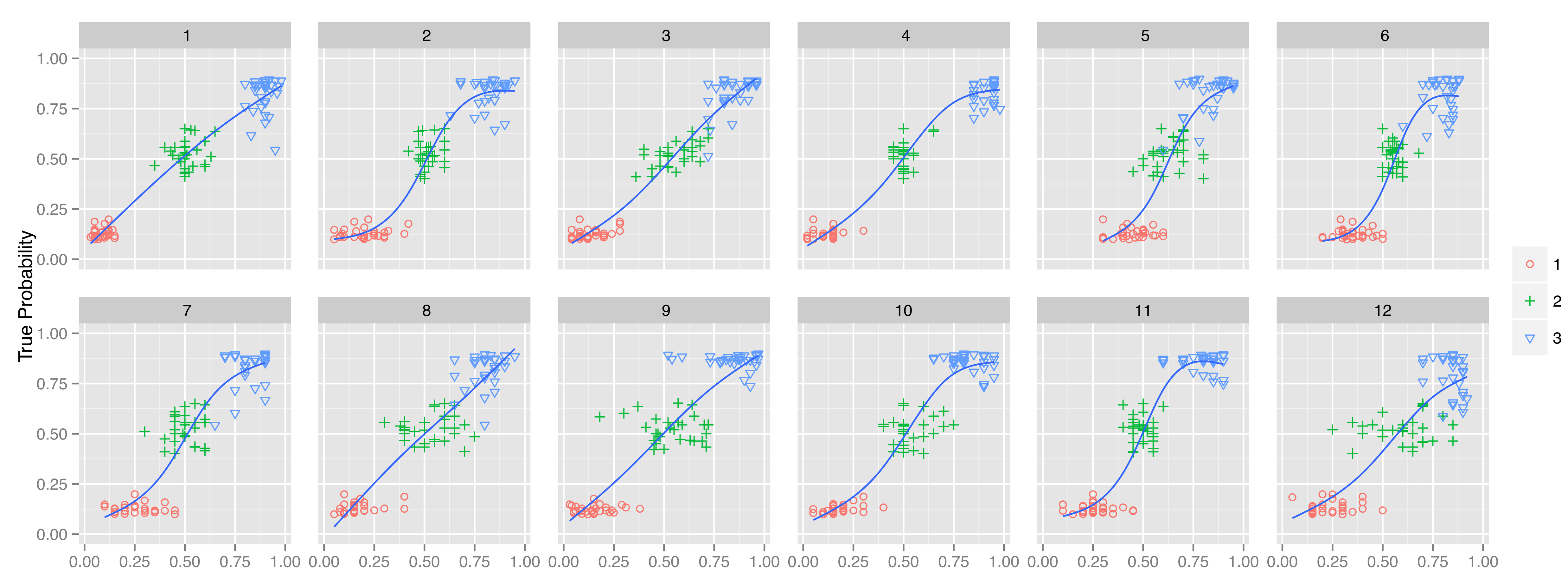
$$\varepsilon_i = \underbrace{\Phi^{-1}(p_i)}_{z_i} - \underbrace{\Phi^{-1}(g(\hat{p}_i)) + \lambda(g(\hat{p}_i))}_{\hat{z}_i}.$$

$\lambda(\hat{p})$ is a bias correction to account for the change of domain. Cluster residual errors by their location in the distribution and estimate variance-covariance matrix Ω.



Experiment

- 30 Normally distributed demand distributions with a variety of known means (ranging from 80 to 126) and standard deviations (ranging from 8 to 42).
- For every distribution, generated video of 25 random realizations, which each flashed on the screen for 2 seconds.
- 12 Ph.D. students provided 3 judgments in the right tail, central region, and left tail of the distribution.
- Participation fee of \$40 plus a potential bonus of \$20 for accuracy (used Binarized QSR; Hossain & Okui, 2013).



Optimal Weights

Proposition: Weights that minimize the variance in $\hat{\theta}_j = \mathbf{w}_j^{*\mathrm{T}}\hat{\mathbf{x}}$ are given by

$$\mathbf{w}_j^{*\mathrm{T}} = \boldsymbol{a}_j(\hat{\mathbf{Z}}\Omega^{-1}\hat{\mathbf{Z}}^{\mathrm{T}})^{-1}\hat{\mathbf{Z}}\Omega^{-1},$$

where $\mathbf{a}_1 = [1, 0]$, $\mathbf{a}_2 = [0, 1]$, and $\hat{\mathbf{Z}} =$

$$\begin{bmatrix} 1 & \ldots & 1 \\ \Phi^{-1}(g(\hat{p}_1)) + \lambda(g(\hat{p}_1)) & \ldots & \Phi^{-1}(g(\hat{p}_N)) + \lambda(g(\hat{p}_N)) \end{bmatrix}$$

Estimates Using Holdout Procedure

- Estimating μ
  - Average RMSE decreases by 16% (Direct curve-fitting=8.37, Calibrated weights=7.07, Direct video data=5.35).
  - Average APE decreases by 19% (Direct curve-fitting=6.25%, Calibrated weights=5.04%, Direct video data=3.87%).
- Estimating σ
  - Average RMSE decreases by 53% (Direct curve-fitting=11.20, Calibrated weights=5.28, Direct video data=2.64).
  - Average APE decreases by 51% (Direct curve-fitting=36.4%, Calibrated weights=17.8%, Direct video data=9.7%).

Discussion and Future Work

- Proposed scale-free model to quantify judgment errors and a method for weighting judgments to estimate the parameters of a variable of interest.
- Tested effectiveness of the method in an experiment, found benefits for estimating the mean, *very large* improvements for estimating the standard deviation.

Ongoing Work

- Error-in-probabilities model works exactly analogously for combining quantile judgments.
- Empirical comparison of quantile and probability elicitation modes both with and without calibration.

References

[1] Ben-David, I., Graham, J. R., Harvey, C. R. 2013. Managerial miscalibration. *Quarterly Journal of Economics*, 1547–1584.

[2] Budescu, D. V., Du, N. 2007. Coherence and consistency of investors' probability judgments. *Management Science* 53(11), 1731–1744.

[3] Bansal, S., Gutierrez, G. J., Keiser, J. R. 2015. Using expert assessments to estimate probability distributions.

[4] Hossain, T., Okui, R. 2013. The binarized scoring rule. *Review of Economic Studies* 80(3), 984–1001.