Simple Bayesian Reference Class Forecasting

for Binary and Continuous Business Variables

Shaun Comfort MD, MBA

Science Enablement Director PDS, Roche-Genentech

Disclaimer

Dr. Comfort is an employee of Roche-Genentech and the opinions he expresses in this presentation are his own and not necessarily those of his employer.

Agenda

1) Introduction
2) Reference Class Forecasting “re-framed” as Bayesian Estimation
3) Example 1 – Probability of Success Estimation
4) Example 2 – Peak Sales Estimation
5) Putting It All Together – Risk Adjusted Peak Sales
6) Discussion and Conclusion

The Problem

Harvard Business Review

Mergers And Acquisitions

Deals Without Delusions

by Dan Lovallo, Patrick Viguerie, Robert Uhlaner, and John Horn

From the Magazine (December 2007)

Summary. Reprint: R0712G Pursuing a merger or acquisition is inherently difficult. Things get even harder when executives are blind to their own faulty assumptions, say Lovallo—a professor at the University of Western Australia Business School and a senior adviser to... more

Review of General Psychology
2003, Vol. 7, No. 4, 331–363

1089-2680/03/$12.00 DOI: 10.1037/1089-2680.7.4.331

One Hundred Years of Social Psychology Quantitatively Described

F. D. Richard
University of North Florida

Charles F. Bond Jr. and
Juli J. Stokes-Zoota
Texas Christian University

This article compiles results from a century of social psychological research, more than 25,000 studies of 8 million people. A large number of social psychological conclusions are listed alongside meta-analytic information about the magnitude and variability of the corresponding effects. References to 322 meta-analyses of social psychological phenomena are presented, as well as statistical effect-size summaries. Analyses reveal that social psychological effects typically yield a value of *r* equal to .21 and that, in the typical research literature, effects vary from study to study in ways that produce a standard deviation in *r* of .15. Uses, limitations, and implications of this large-scale compilation are noted.

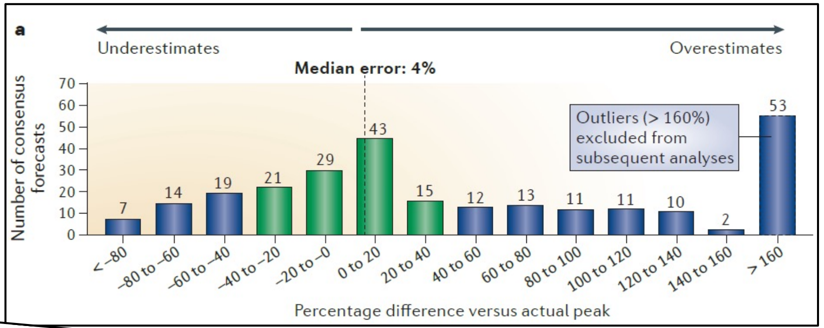
From Peter W. G. Morris, Jeffrey K. Pinto, and Jonas Söderlund, 2011, eds., *The Oxford Handbook of Project Management* (Oxford University Press), pp. 321-344.

CHAPTER 13

OVER BUDGET, OVER TIME, OVER AND OVER AGAIN

MANAGING MAJOR PROJECTS

BENT FLYVBJERG



Why Your IT Project May Be Riskier than You Think

New research shows surprisingly high numbers of out-of-control tech projects — ones that can sink entire companies and careers.

By
Bent Flyvbjerg and Alexander Budzier
BT Centre for Major Programme Management
Saïd Business School
University of Oxford

NEWS & ANALYSIS

FROM THE ANALYST'S COUCH

Pharmaceutical forecasting: throwing darts?

Myoung Cha, Bassel Rifai and Pasha Sarraf

Image from fckncg/Alamy

Harvard Business Review

Decision Making And Problem Solving

Delusions of Success: How Optimism Undermines Executives' Decisions

by Dan Lovallo and Daniel Kahneman

From the Magazine (July 2003)

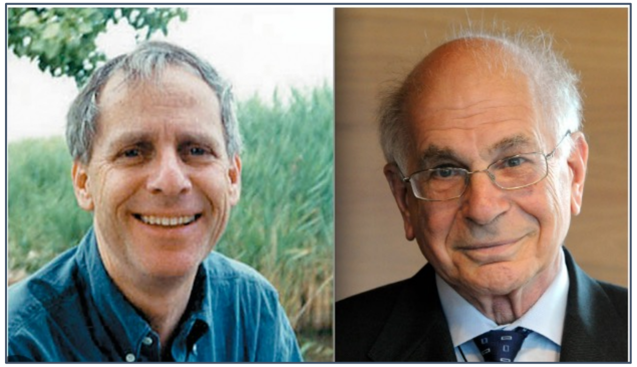
Summary. Reprint: R0307D The evidence is disturbingly clear: Most major business initiatives—mergers and acquisitions, capital investments, market entries—fail to ever pay off. Economists would argue that the low success rate reflects a rational assessment of risk, with the... more

Normative Prediction Principles

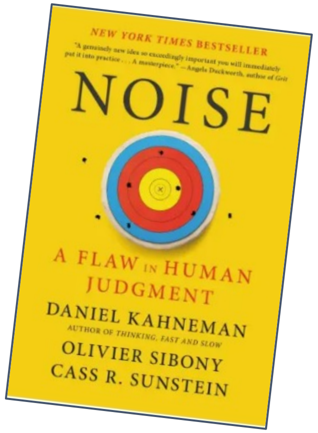
1) The “extremeness” of predictions should be moderated by “predictability”
2) Predictability is commonly measured by “predictive validity” (ie, correlation between outcomes and forecasts)
    a) When predictability is perfect ($r = 1.0$), the forecast is the best prediction
    b) When predictability is zero ($r = 0.0$), the mean (ie, prior or base rate) is the best forecast
    c) For intermediate situations, a weighted average of the forecast and base rate prediction is best
3) This became the basis for Reference Class Forecasting

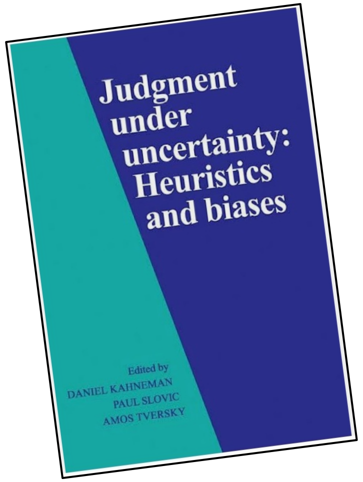
Kahneman & Tversky Reference Class Forecasting (RCF)

- RCF developed by Daniel Kahneman and Amos Tversky to correct consistent human errors in forecasting:



  - Over-reliance on “inside view” (ie, specific, unique details and features) of projects when forecasting completion times, probability of success, investment returns, etc
  - Under-reliance on known distributions of historical outcomes for similar projects or “outside view”
  - Common manifestation of inside-view is the well documented “planning fallacy”
  - Affects expert and layperson forecasts equally





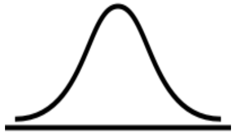
Source: Daniel Kahneman, 2011, Thinking, Fast and Slow (New York: Farrar, Straus and Giroux)

Kahneman & Tversky Corrective (RCF) Method

- Their procedure adopted Truman Kelley's "true score" regression approach:
  - *… the best estimate is obtained by regressing the observed [forecast] in the direction of the mean [of the reference class] Wainer. Chance 2000.*
- The result is a weighted average between the mean of the reference class distribution and the SME forecast

Obtain "inside view" SME forecast for specific task (eg, PTS)

Obtain distribution of outcomes from similar class of projects (eg, Reference Class)



Estimate "predictive validity" (ie, correlation) between SME forecasts and actual outcomes

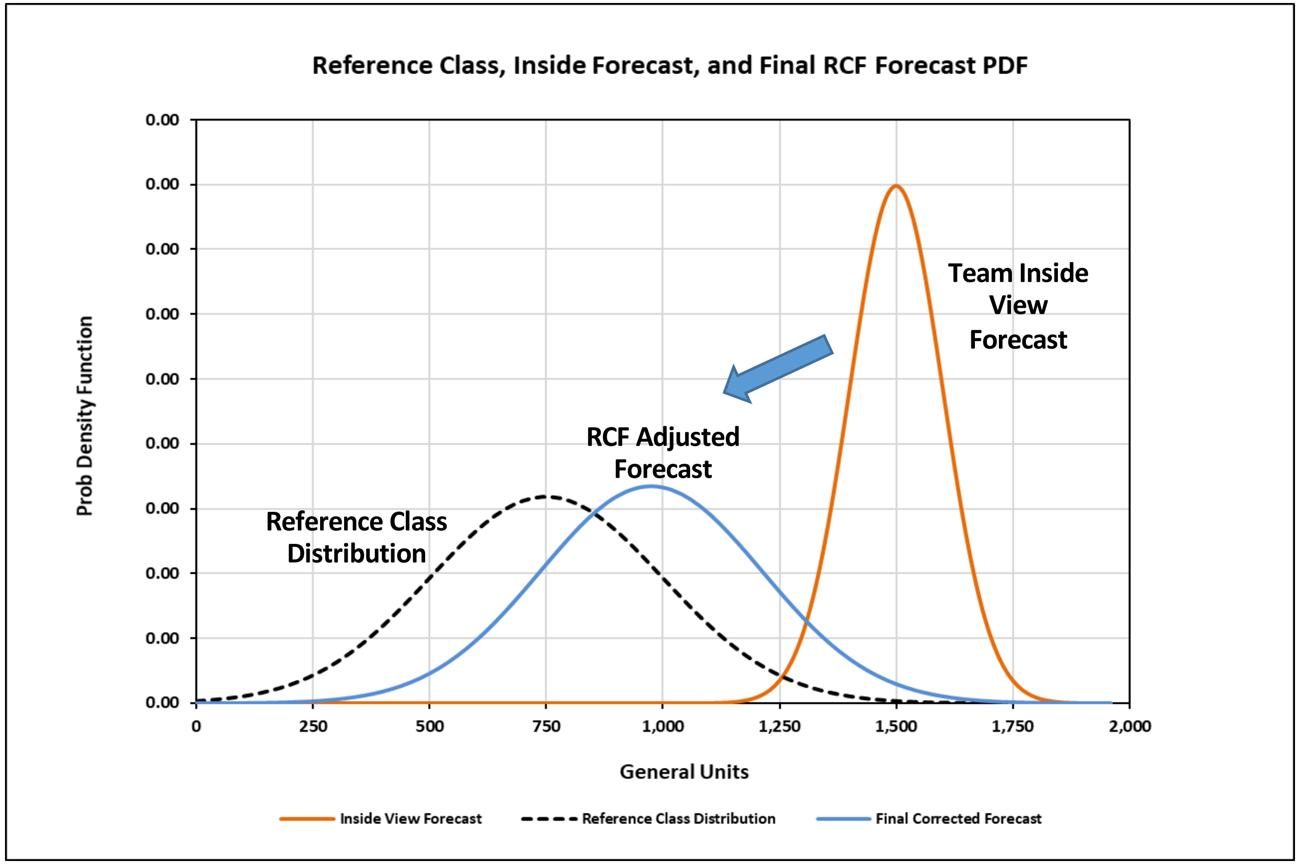
$$\boldsymbol{PV_{est}}$$

Combine RCF (Outside View) with SME (Inside View) Forecast

$$\boldsymbol{Best_{est} = PV \times SME_{Est} + (1 - PV) \times RCF_{Est}}$$

The Effect of Reference Class Forecasting?

- The inside-view team forecast probability density function (PDF) is "regressed" towards the reference class distribution proportional to the predictive validity



Example #1 – XYZ BioPharma PPOS Correction

- You are a consultant for XYZ BioPharma Corporation asked to evaluate the team's predicted probability of success (PPOS) for an important upcoming clinical trial. The company keeps detailed records of all development team forecasts and final study outcomes that can be used as our reference class.

- Analysis of the historical data** shows that over the last 7 years the company conducted 81 studies with 32 successes, for 32/81 ≈ 40% frequency of success (FOS) with the corresponding team forecasts showing an average predicted success rate of 60%.

- The current team's PPOS for the planned study is 70%. Assessment of the historical PPOS forecasts and outcomes show that the predictive validity of the team forecasts = 0.29.

**All data are hypothetical, generated using monte-carlo methods and/or expert input

Example #1 – XYZ BioPharma PPOS Correction

- *Using the KT RCF formula described earlier, the revised PPOS 70% forecast becomes 48%:*

$$\mathrm{PPOS}_{\mathrm{est}} = 0.29\times 70\% + (1-0.29)\times 40\% = 48\%$$

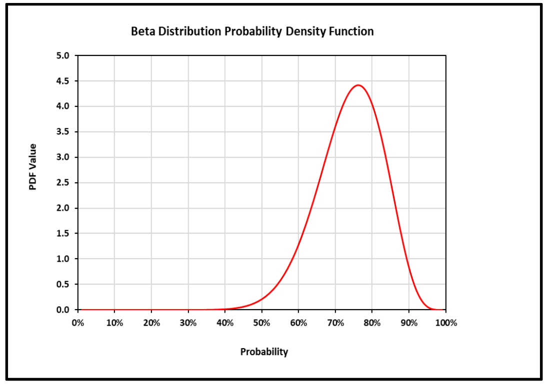
- *As expected, the team forecast has been regressed towards the reference class mean*

Bayesian Inference Conjugate Models

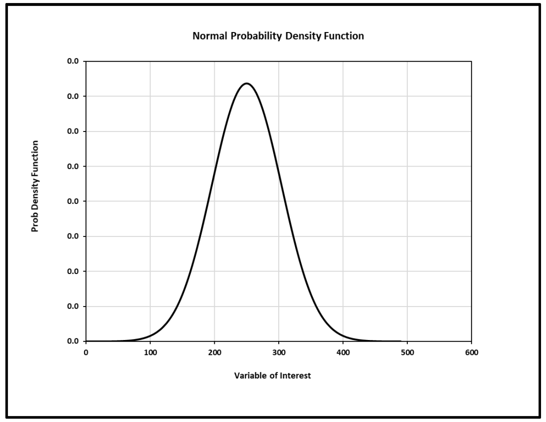
- Bayes Rule: $p(\theta|x) \propto p(\theta)p(x|\theta)$; *posterior* $\propto$ *prior x likelihood*
- If posterior and prior are in same distribution family (eg, normal), they are termed "conjugate distributions"
- Conjugate distributions allow for simple, closed form solutions for posterior

Important Conjugate Models for RCF

- Bernoulli Outcomes:
  - Beta prior simple updating rule
    - Beta Posterior: *Beta(a', b') = Beta(a + ns, b + nf);*
    - *where a = successes, b = failures, ns = new successes, nf = new failures*



- Continuous Normal Outcomes:
  - Normal prior simple updating rule
    - Normal prior: $\mathrm{N}(\mathrm{m}_0, \mathrm{s}_0^2)$, Normal observation: $\mathrm{N}(y, \sigma^2)$
    - Normal posterior: $\mathrm{N}\left(m_p, \mathrm{s}_\mathrm{p}^2\right)$ where:
      - $m_p = \left(\frac{pr_s}{pr_p}\right) m_0 + \left(\frac{pr_\sigma}{pr_p}\right)$y, where pr = precision = 1/variance
      - $\frac{1}{s_p^2} = \frac{1}{s_0^2} + \frac{1}{\sigma^2}$; or posterior precision = sum of precisions



Recasting RCF using Bayesian Conjugate Models

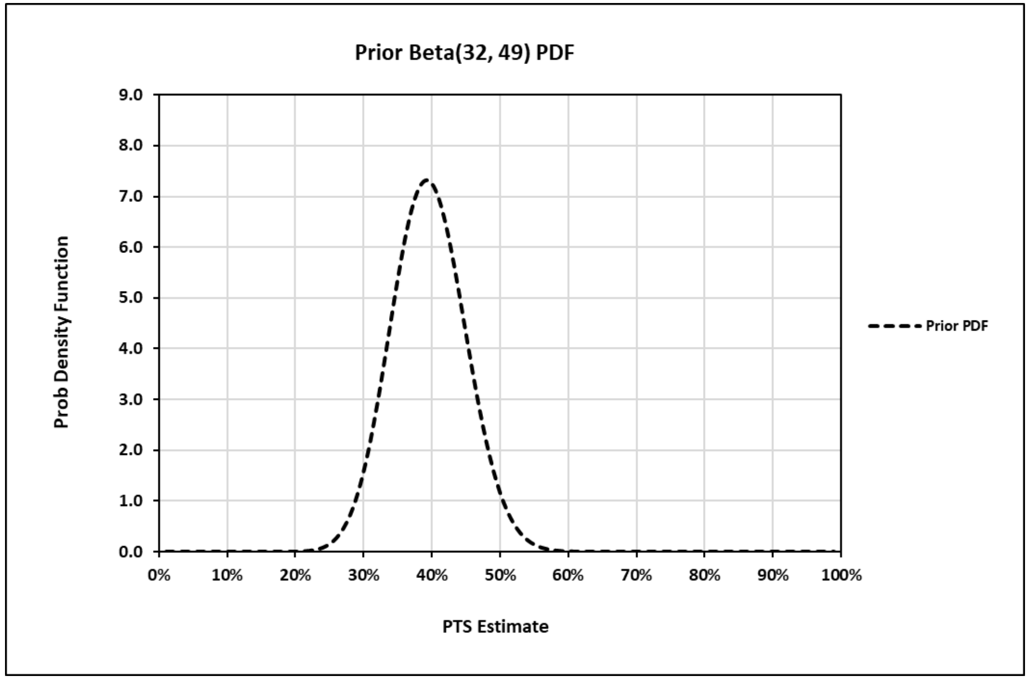
- There are two connections between RCF and Bayes rule that are useful:
  - A) Bayes rule and RCF approaches are both weighted averages between a prior and additional information
  - B) Both approaches must "weigh" the value of the additional information relative to the prior (think base-rate)
- This allows one to link the two concepts:

Number	Kahneman – Tversky Concept	Bayes Rule	Comment(s)
1	Reference Class (Base Rate) Distributional Data	Informed Prior distribution	The prior is an "informed" distribution for the quantity of interest, based on previous or relevant data in contrast to a flat or uniform prior
2	Inside View (Team) Forecast	New information or data	The subjective forecast provides new information from the team that can be used to calculated a posterior (best estimate) distribution
3	Predictive Validity	• Bernoulli Model – Estimates Effective Sample Size (ie, weight) for Beta Distributions • Normal Model – Estimates precision (ie, weight) for Gaussian Distributions	Predictive validity is used with the priors to estimate the effective sample size or precision (ie, the impact weights) for new information (ie, the uncorrected forecast)
4	Corrected/Recalibrated Forecast	Posterior Distribution	The posterior distribution is the best estimate for the forecast of interest and can be used to calculate point estimates, uncertainty ranges, etc

Source: Adapted from Comfort S. Estimating Predictive Probability of Success. Foresight. Issue #72, 2024.

Example #1 – Revisited from Bayesian Perspective

1) Create a beta distribution prior equivalent to the base rate:
   - Equivalent sample size = 32 successes + 49 failures = 81 total trials
   - Informed Prior = Beta(32, 49) distribution shown below



Source: Adapted from Comfort S. Estimating Predictive Probability of Success. Foresight. Issue #72, 2024.

Example #1 – Revisited from Bayesian Perspective

2) Estimate posterior and team forecast effective sample size (ESS):
   - Posterior ESS = $N_{prior}/(1 - Pv) = 81/(1 - 0.29) \approx 114$

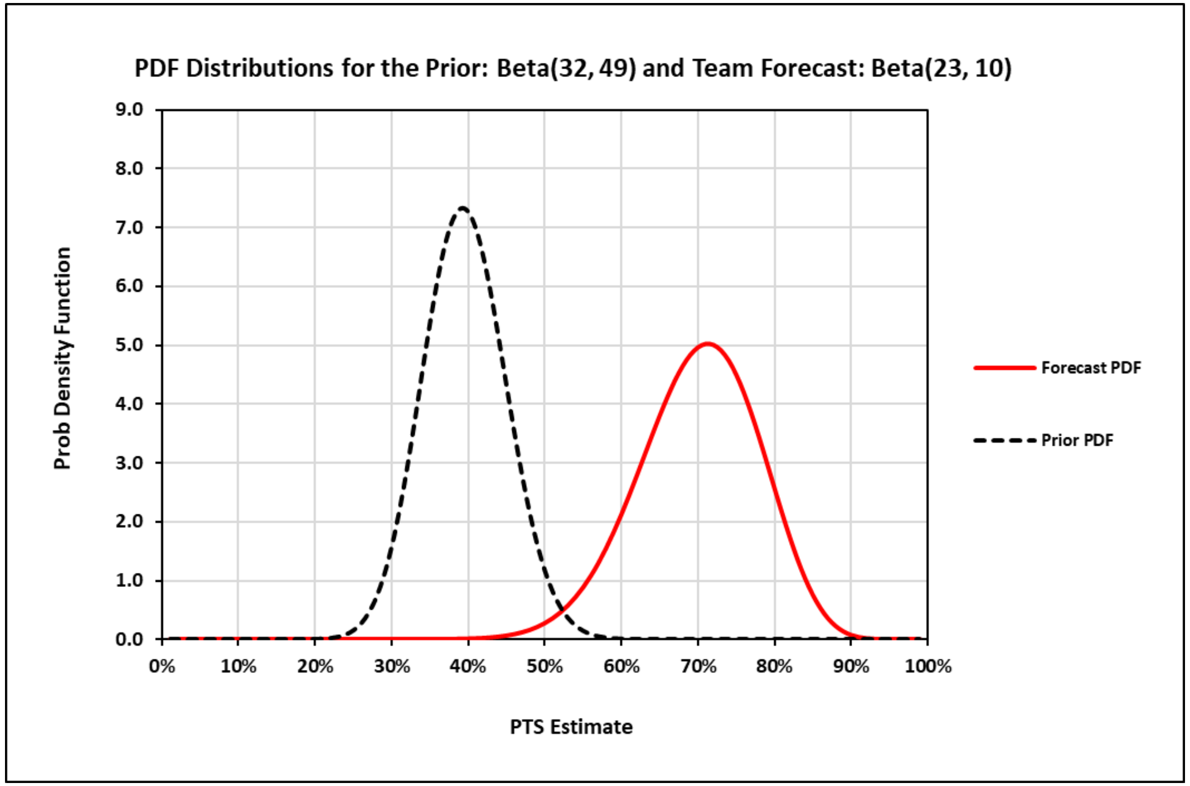
3) Since $N_{posterior} = N_{prior} + N_{forecast}$, solve for $N_{forecast}$ (ESS):
   - Forecast ESS = $N_{forecast} = 114 - 81 = 33$

4) Set team forecast = forecast Beta distribution mean to find the hyperparameters (*a* and *b* values):
   - Forecast Mean = 70% = *a_value*/ESS = *a_value*/33 => *a_value* ≈ 23
   - Therefore *b_value* = 33 – 23 = 10

Source: Adapted from Comfort S. Estimating Predictive Probability of Success. Foresight. Issue #72, 2024.

Example #1 – Revisited from Bayesian Perspective

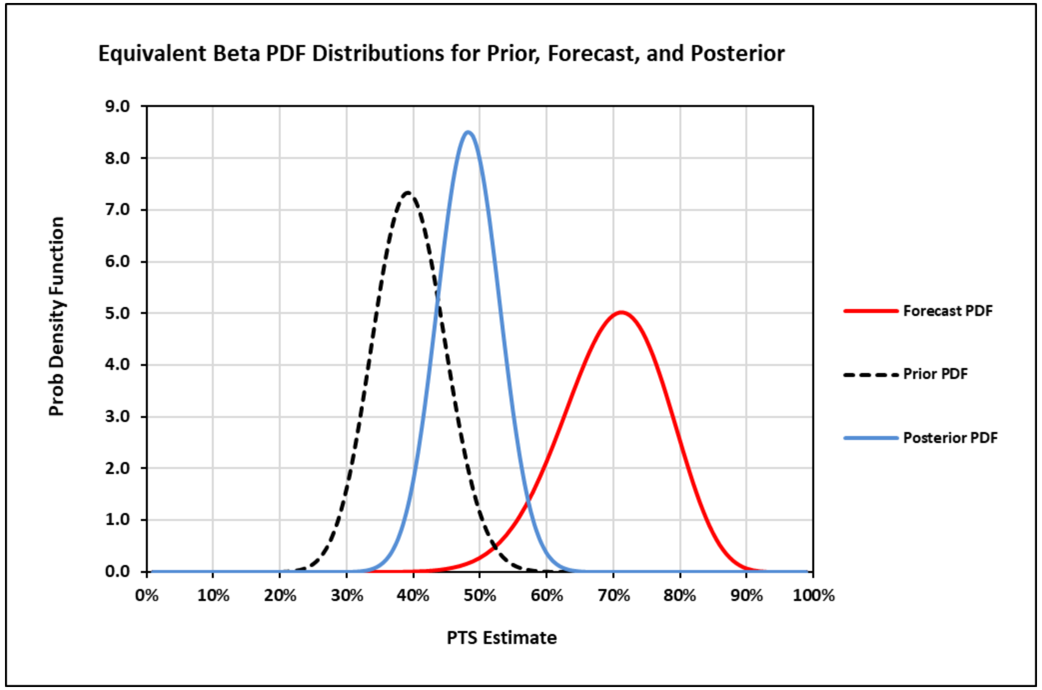
- The resulting beta distributions for the prior and team forecast are shown below:



Source: Adapted from Comfort S. Estimating Predictive Probability of Success. Foresight. Issue #72, 2024.

Example #1 – Revisited from Bayesian Perspective

5) Finally, determine hyperparameters for posterior by updating prior:
   - Posterior = Beta(32 + 23, 49 + 10) = Beta(55, 59)
   - The Prior, Team Forecast, and Posterior PDFs are shown below



Source: Adapted from Comfort S. Estimating Predictive Probability of Success. Foresight. Issue #72, 2024.

Example #1 – Bayesian Perspective Final Points

- Bayesian results are probability distributions, not point estimates
- Relevant point estimates (eg, mean, standard deviation, etc) are easily determined (see table example below)
- The KT RCF Mean and Bayesian Mean will be equal
- The distributions are critical to showing the uncertainty in the estimates for both decision makers and teams

KT RCF PPOS Estimates		Bayesian PPOS Estimates				
Variable	**Mean**	**Variable**	**Mean**	**P10**	**P50**	**P90**
Team PPOS	70%	**Team PPOS**	70%	60%	70%	80%
Base Rate	40%	**Prior PPOS**	40%	33%	39%	47%
Recal PPOS	48%	**Post PPOS**	48%	42%	48%	54%

Source: Adapted from Comfort S. Estimating Predictive Probability of Success. Foresight. Issue #72, 2024.

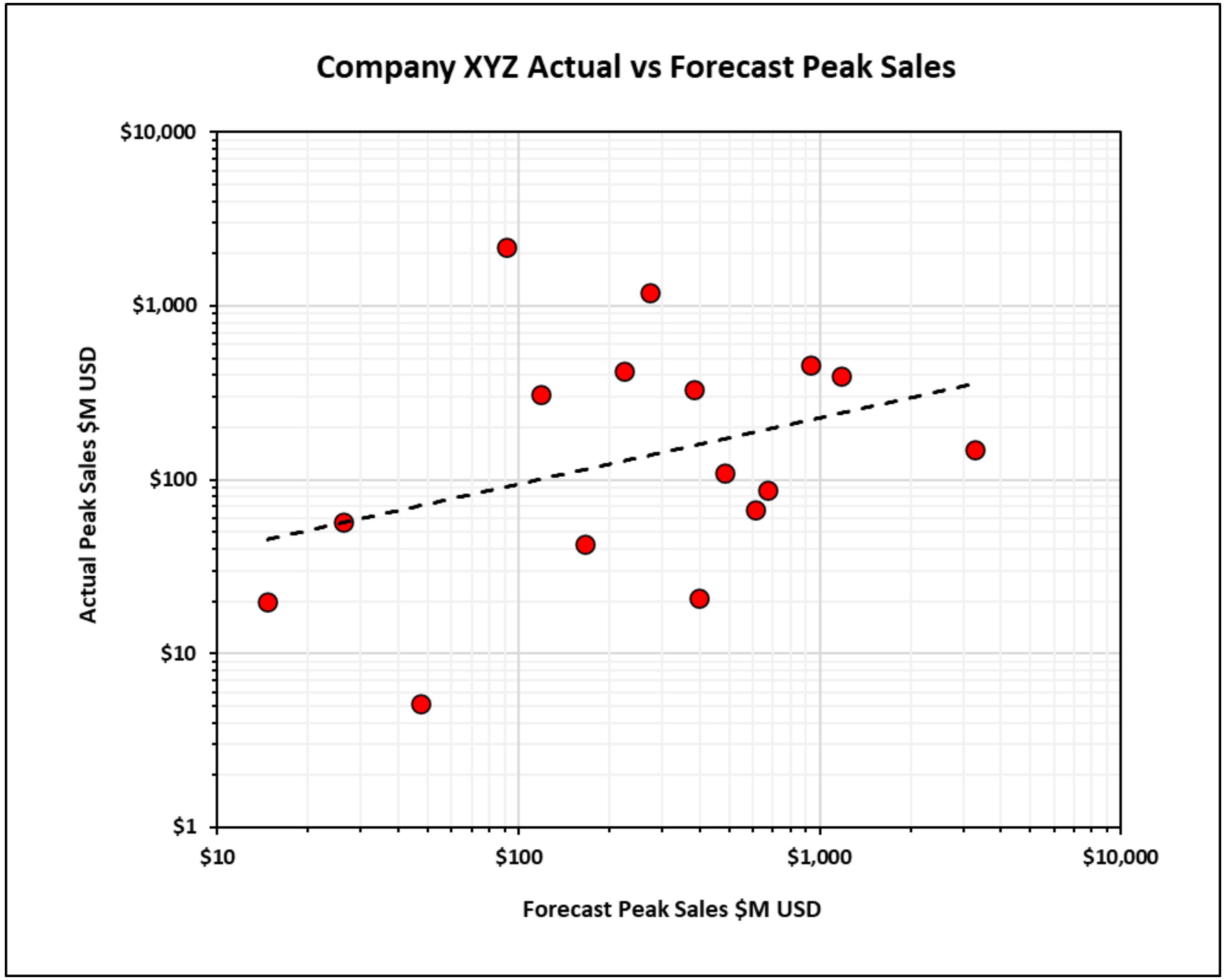
Example #2 – XYZ BioPharma Peak Sales (PkS) Correction

- You are now asked to evaluate the team's predicted PkS for the product, assuming a successful trial from Example #1.
- Similar to POS, the company keeps detailed records of all previous PkS forecasts and outcomes that can be used as our reference class.
- The current team's PkS forecast at 5 years post-launch = $750 M USD**. No high or low uncertainty bounds are provided
- The following slide shows the result of historical analysis of PkS forecasts and outcomes

**All data are hypothetical, generated using monte-carlo methods and/or expert input

Example #2 – XYZ PkS Forecast Accuracy

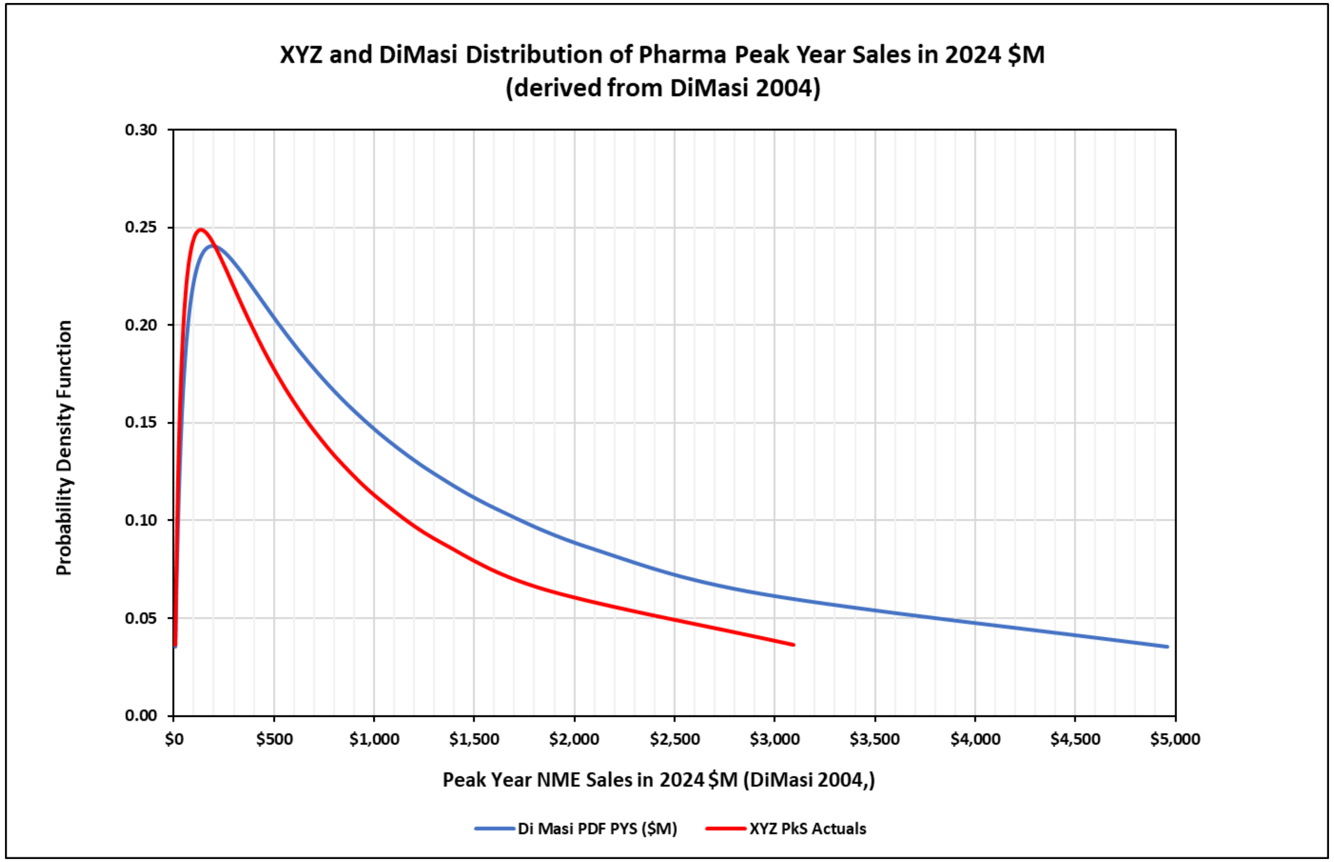
- XYZ past PkS data** collected for 16 products, inflation corrected to current year, and analyzed
- Correlation analysis shows *r* = predictive validity = 0.34
- Descriptive analysis shows actual PkS outcomes follow log normal distribution with Ln Mean = 4.9 and SD = 1.6
- Implies strongly right skewed distribution with Mean ≈ $483 M and SD ≈ $1670 M



**All data are hypothetical, generated using monte-carlo methods and/or expert input

Example #2 – Historical Forecast Accuracy, cont

- As sanity check, public data from DiMasi et al 2004 used to estimate distribution of PkS, inflated to current year
- DiMasi results also show log normal behavior with Ln Mean = 5.3 and Sd = 1.7; equivalent to Mean ≈ $840 M and SD ≈ $3500 M
- Di Masi and implied XYZ probability density functions (PDFs) shown in graph to right
- Both distributions exhibit significant right skewing



**All XYZ data are hypothetical, generated using monte-carlo methods and expert input

Example #2 – Bayesian Updating with Normal Model

- Published Pharma PkS and XYZ actual PkS appear as strongly skewed normal distributions, allowing use of Bayesian Normal Conjugate model to correct team PkS forecasts
- Perform the math in "normal" space and convert to "logs" for visualization and understanding*
- Here the conjugate normal model uses precision (ie, 1/variance) with predictive validity to "weigh" the prior and posterior:
  - Prior precision = 1/variance = $1/1670^2 \approx 3.6e\text{-}7$
  - Posterior precision = $\text{Precision}_{prior}/(1 - Pv) = 3.6e\text{-}7/(1 - 0.34) \approx 5.2e\text{-}7$

*Note – Calculations here are performed in 'normal' space to avoid back-and-forth conversion between natural and log-normal parameters.

Example #2 – Bayesian Updating with Normal Model, cont.

- Posterior precision = Prior + Forecast precisions, so we can solve for team forecast precision:
  - Forecast Precision = $\text{Precision}_{\text{posterior}} - \text{Precision}_{\text{prior}}$ = 1.6e-7

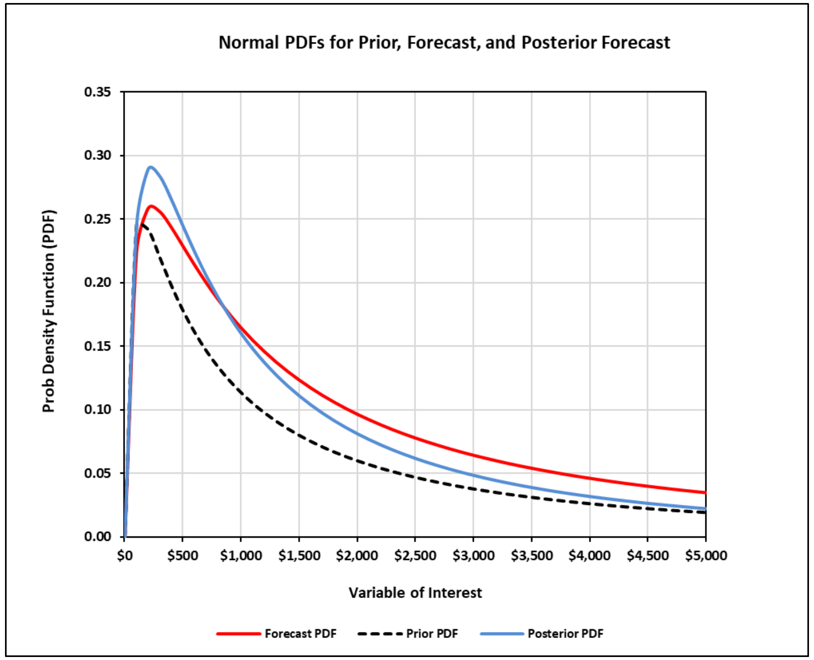
- Finally use normal updating rule from Slide 10 to specify posterior as weighted average of prior and team forecast
  - Posterior Mean = $\left(\frac{precision_{prior}}{precision_{post}}\right) mean_{prior} + \left(\frac{precision_{forecast}}{precision_{post}}\right) mean_{forecast}$

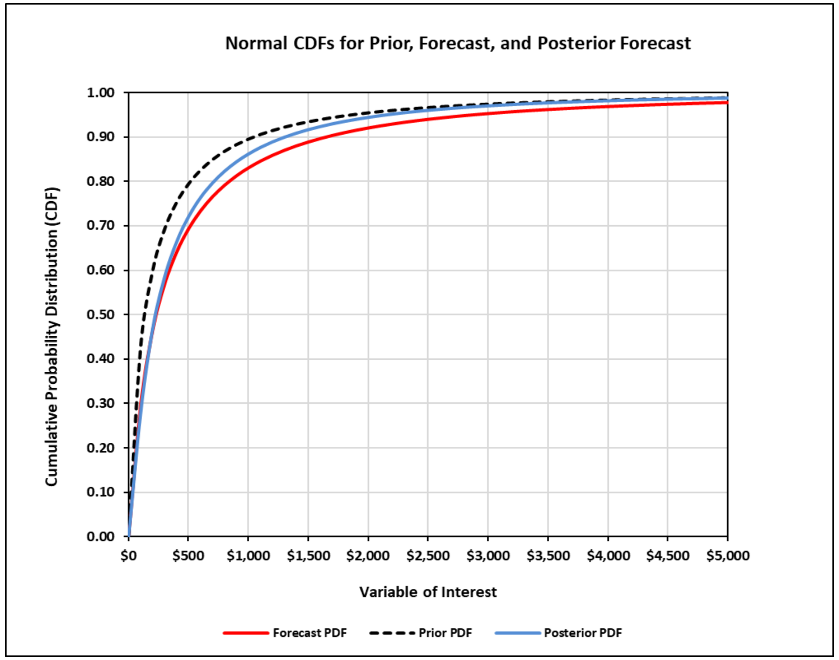
- Numeric results are shown in the table below along with Bayesian Triplot (next slide) and KT RCF result

	Mean	**Precision**	**SD**
Prior	\$483	3.6e-7	\$1,670
Forecast	\$750	1.8e-7	\$2,327
Posterior	\$574	5.4e-7	\$1,299
KT RCF	\$483	N/A	N/A

Example #2 – Bayesian Updating with Normal Model, cont.

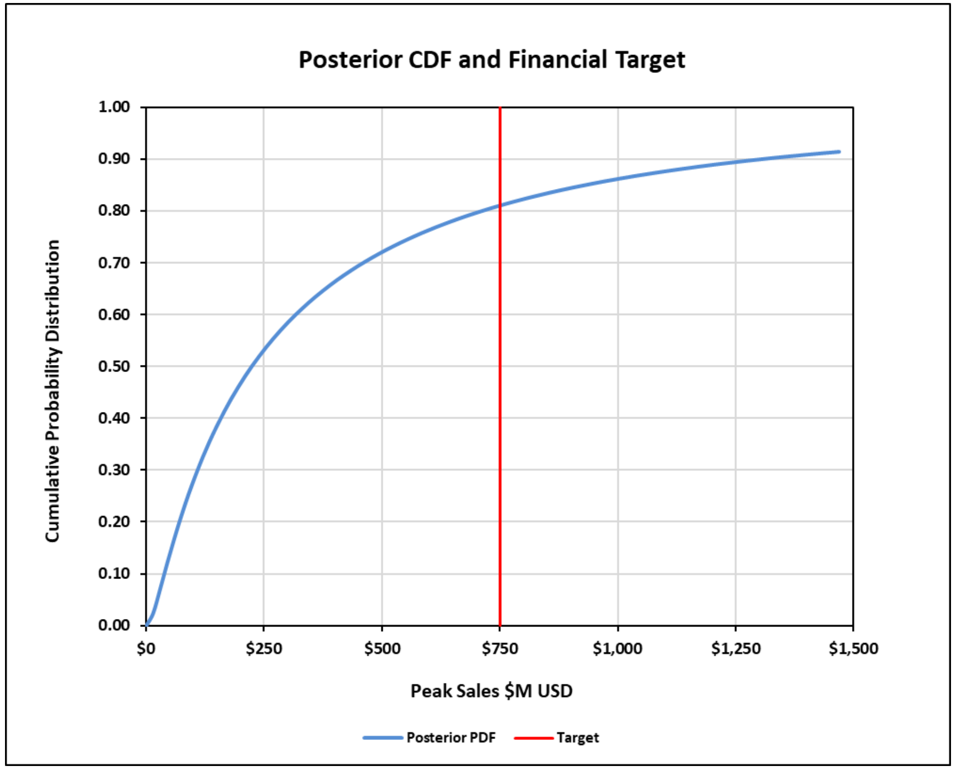
- Note due to the skewed nature of the distributions, PDF peaks correspond to medians and not means
- The graphs below illustrate the Bayesian PDF Triplot (left) and the associated Cumulative Probability Distributions (CDFs, right)
- The final corrected conditional XYZ PkS mean forecast is now approximately \$573 M with an SD ≈ \$1,300 M





Example #2 – How to Use the Results?

- One straightforward use is to ask for the conditional probability of "exceedance"
- Here we use the team original PkS forecast of $750M USD and ask:
  - What is Probability of exceeding or falling short of the team's result?



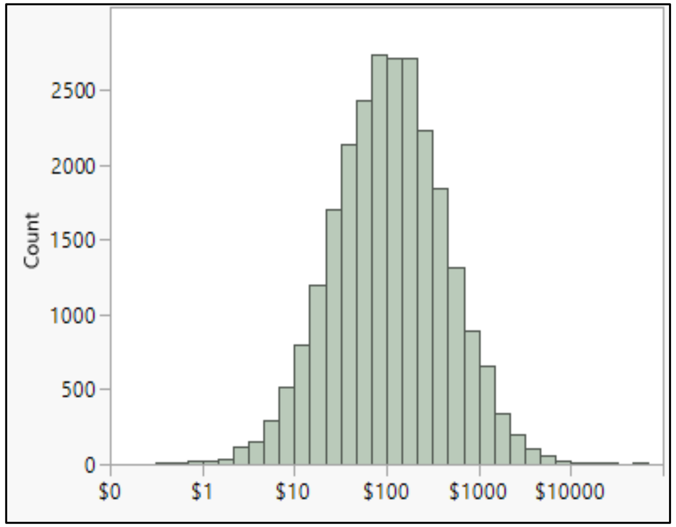
- Visual inspection and calculating the area under the curve shows meeting the forecast is very unlikely (ie, Prob < 25%)

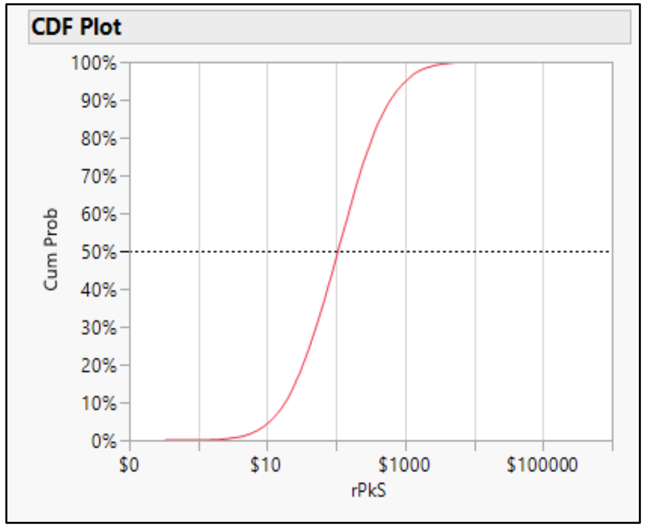
Prob PkS ≥ Forecast	**19%**
Prob PkS < Forecast	**81%**

- This information can be useful for decision makers

Putting It All Together: Risk Adjusted Peak Sales

- To estimate the unconditional PkS (ie, before study outcome known):
  - Draw 25,000 random samples from the POS and PkS distributions and multiply them
  - Generate histogram of the risk-adjusted (aPkS) distribution
  - Graphs below show the resulting aPkS density and cumulative distributions:
- Resulting Risk Adjusted PkS Mean ≈ $280 M and Median ≈ $100 M





Summary Statistics

Mean	277.67253
Std Dev	943.66912
Std Err Mean	5.9682875
Upper 95% Mean	289.37072
Lower 95% Mean	265.97433
N	25000
Variance	890511.4
N Missing	0
Median	106.30061
Range	100257.16
Interquartile Range	225.31036

Observations

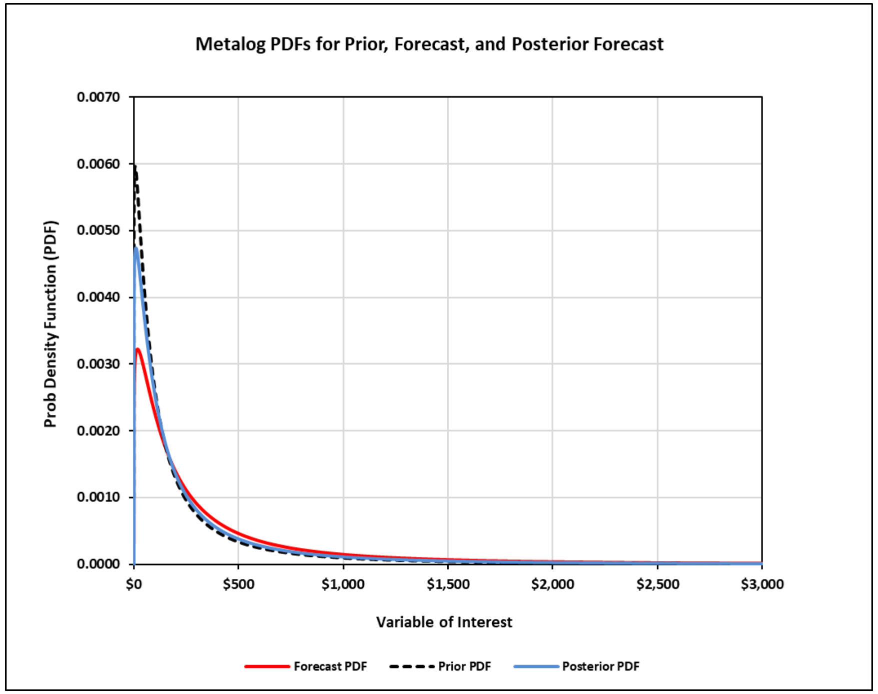
- Kahneman & Tversky’s RCF procedure can be effectively recast as Bayesian inference using Conjugate Distributions
  - Feasible for both Binary and Continuous data
  - Produces full distributions suitable for estimating means, credible intervals, etc
  - Approach is simple, transparent, and easily implemented in standard excel
  - Provides decision makers with quantitative, visual de-biasing method for evaluating “promoter” forecasts
- *Caviats:*
  - *The Normal Conjugate model can be “clunky” to implement (eg, with highly skewed non-negative data, etc)*
  - *Log-Normal conversion may ‘over predict’ outliers (eg, long-broad PkS tail?)*
  - *Metalog distributions may be a suitable alternative?*

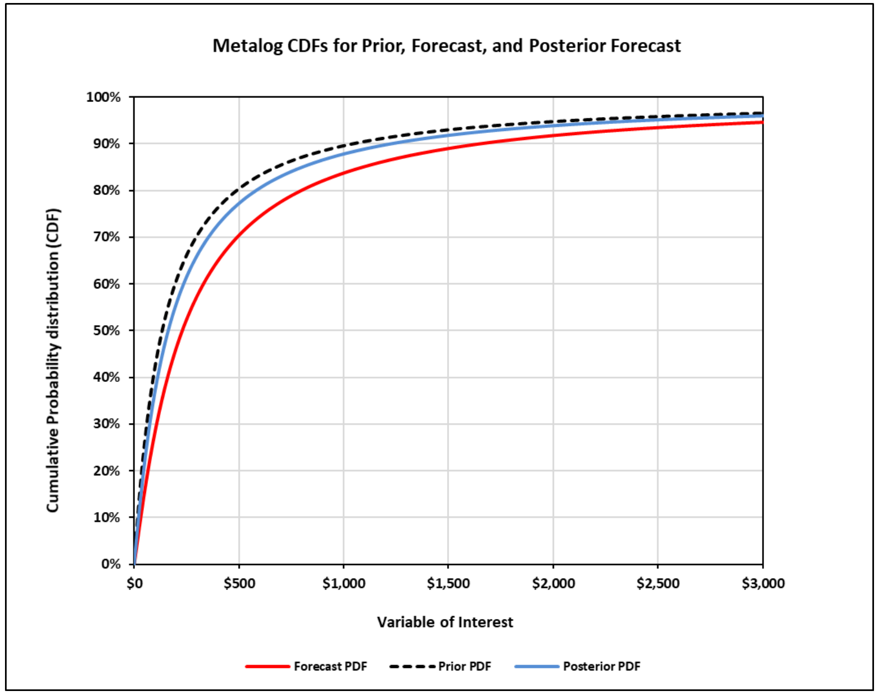
Example #2 – RCF with Metalogs?

- The graphs below illustrate the same PDF Triplots (left) and the associated Cumulative Probability Distributions (CDFs, right) from Example #2 – using SPT Meta-logs
- Performing the same "exceedance" analysis shows:

Prob PkS ≥ Forecast	**16%**
Prob PkS < Forecast	**84%**

- Results are similar although general PDF/CDF shapes differ





Concluding Thoughts

- Kahneman & Tversky’s RCF procedure can be effectively recast as Bayesian inference using Conjugate Distributions:
  - Predictive validity provides “weight” for new observations (ie, forecasts)
  - Bernoulli data effectively modeled with Conjugate Beta Distributions
  - Continuous data can be modeled with Conjugate Normal Distributions – *but*
    - Metalog distributions (SPT3 or higher) should be considered for situations with highly skewed, unbounded or bounded, continuous data
    - At minimum, consider Metalog formulation as a “sanity” check to compare with Conjugate Normal approach

Special Thanks to:

Mr. Darren Dorrell; Roche-GNE
Dr. Tom Keelin; SDP

Questions?

Selected References

1) Bolstad W and Curran J. Introduction to Bayesian Statistics. 3rd Ed. Wiley and Sons, 2017.
2) Kahneman et al. Noise: A flaw in human judgment. Little, Brown, Spark. 2021.
3) Tetlock, PE. Expert Political Judgment. Princeton University Press. 2005.
4) Kahneman D. and Tversky A. Intuitive Prediction: Biases and Corrective Procedures. Judgment under Uncertainty: Heuristics and biases. Cambridge University Press (1982).
5) Kahneman D. and Tversky A. The Evidential Impact of Base Rates. Judgment under Uncertainty: Heuristics and biases. Cambridge University Press (1982).
6) Flyvbjerg B. Over Budget, Over Time, Over and Over Again: Managing Major Projects. The Oxford Handbook of Project Management. Oxford University Press (2011).
7) Comfort, S. Estimating Predictive Probability of Success. Foresight. Issue 72, 2024.
8) Keelin, TW. The Metalog Distributions. Decision Analysis 13(4):243-277