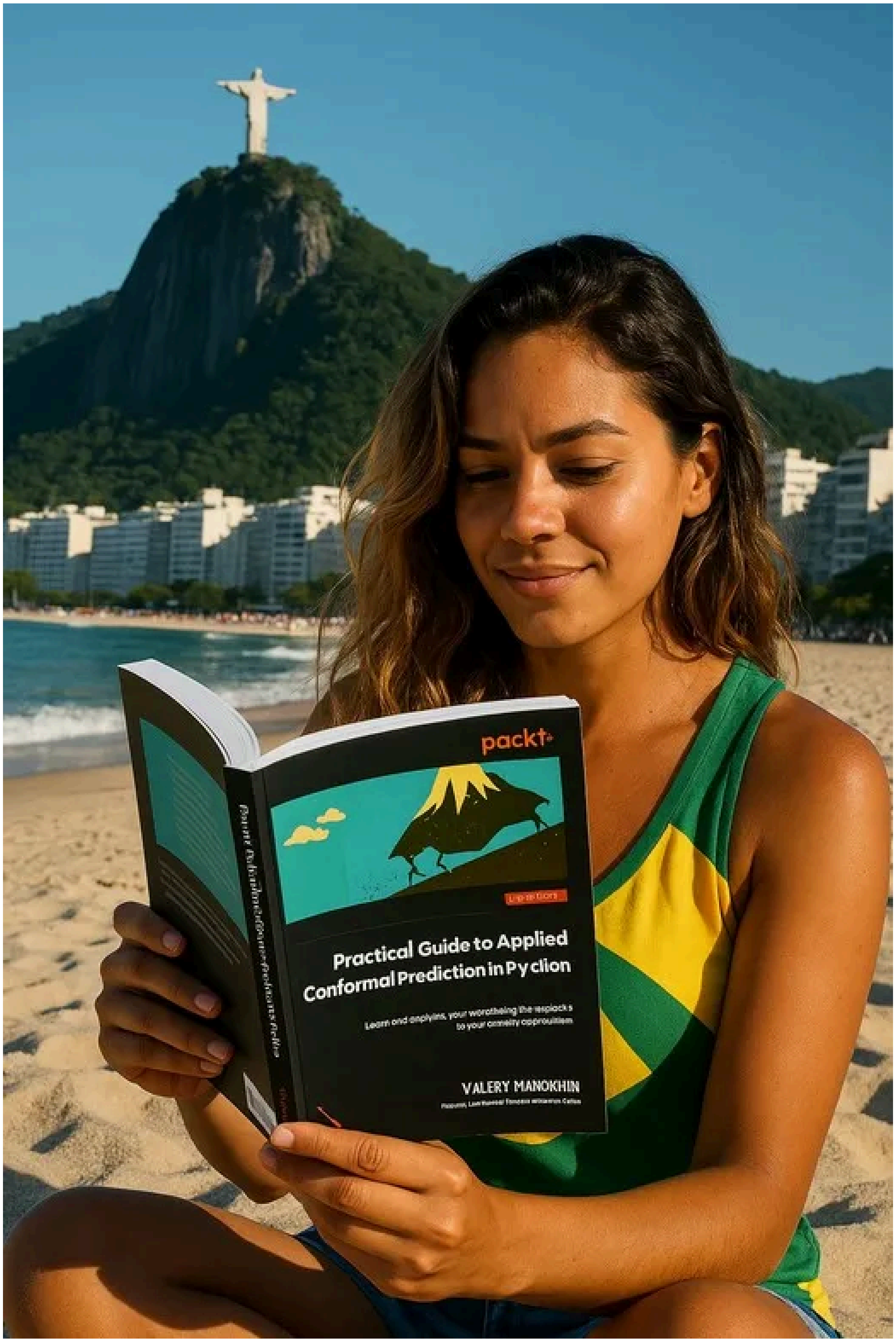
Predicting Full Probability Distributions with Conformal Prediction

Valeriy Manokhin, PhD, MBA, CQF Follow 13 min read · Mar 28, 2025



Introduction

In machine learning, accurate **uncertainty quantification** is as important as the predictions themselves. Traditional models often provide only point estimates or simple prediction intervals, leaving decision-makers in the dark about the full range of possible outcomes. **Conformal Predictive Distributions (CPDs)** offer a breakthrough: they output an entire *cumulative distribution* for each individual prediction, complete with **guaranteed calibration**.

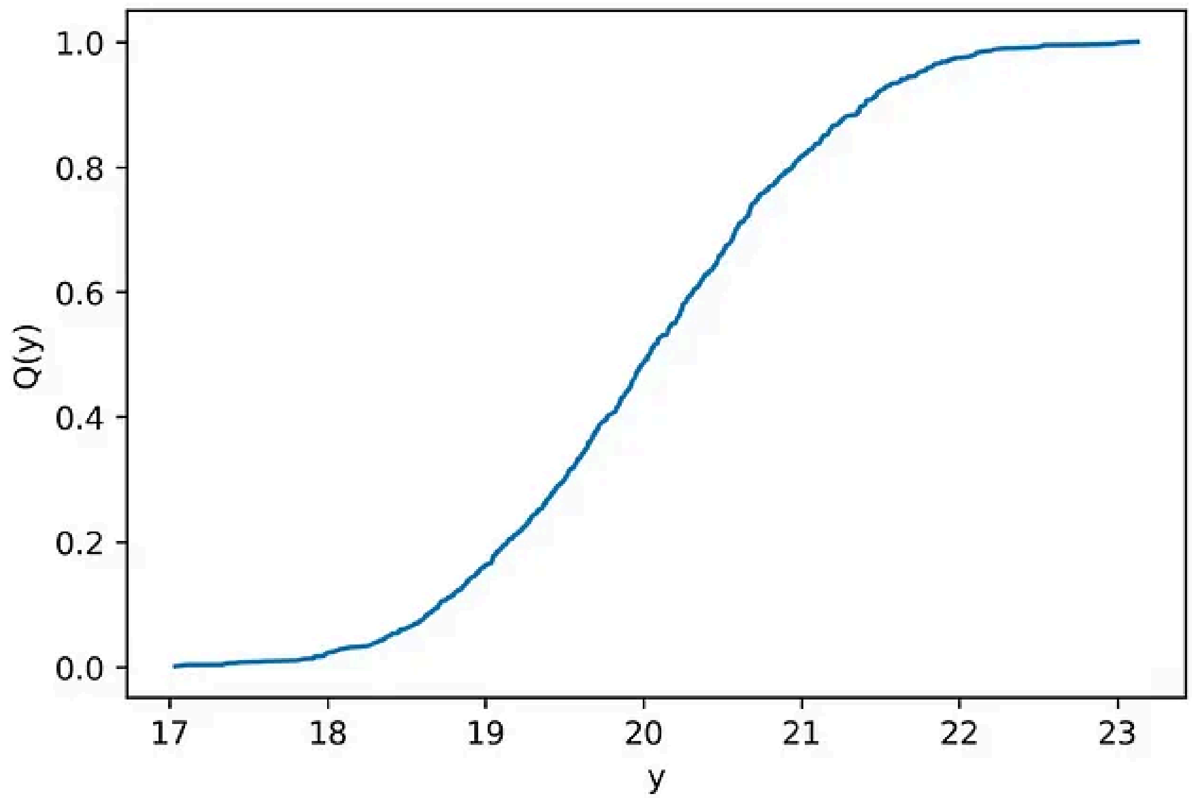
First introduced by Vovk & Manokhin in 2017, CPDs have quickly gained traction — our original paper on nonparametric predictive distributions using conformal prediction has now amassed over 100 citations, reflecting its substantial academic and practical impact.

In this article, we explore what CPDs are, why they are so significant, and how they are being used across domains like healthcare, finance, and climate. We'll also look at notable extensions and how CPDs are implemented in practice (e.g. via the `crepes` library), providing a clear and updated guide beyond our earlier Medium piece on this topic.

What are Conformal Predictive Distributions (CPDs)?

Conformal Predictive Distributions arise from the framework of **Conformal Prediction** — a model-agnostic technique that guarantees *valid* prediction intervals or sets under minimal assumptions (essentially just that data are exchangeable).

Instead of yielding just an interval (e.g. a lower and upper bound for a regression prediction), a conformal predictive **system** produces a full **distribution function** Q(y) for each new example. This function can be thought of as the model's predicted CDF: for any potential outcome value y, Q(y) gives the probability that the true target will be less than y.

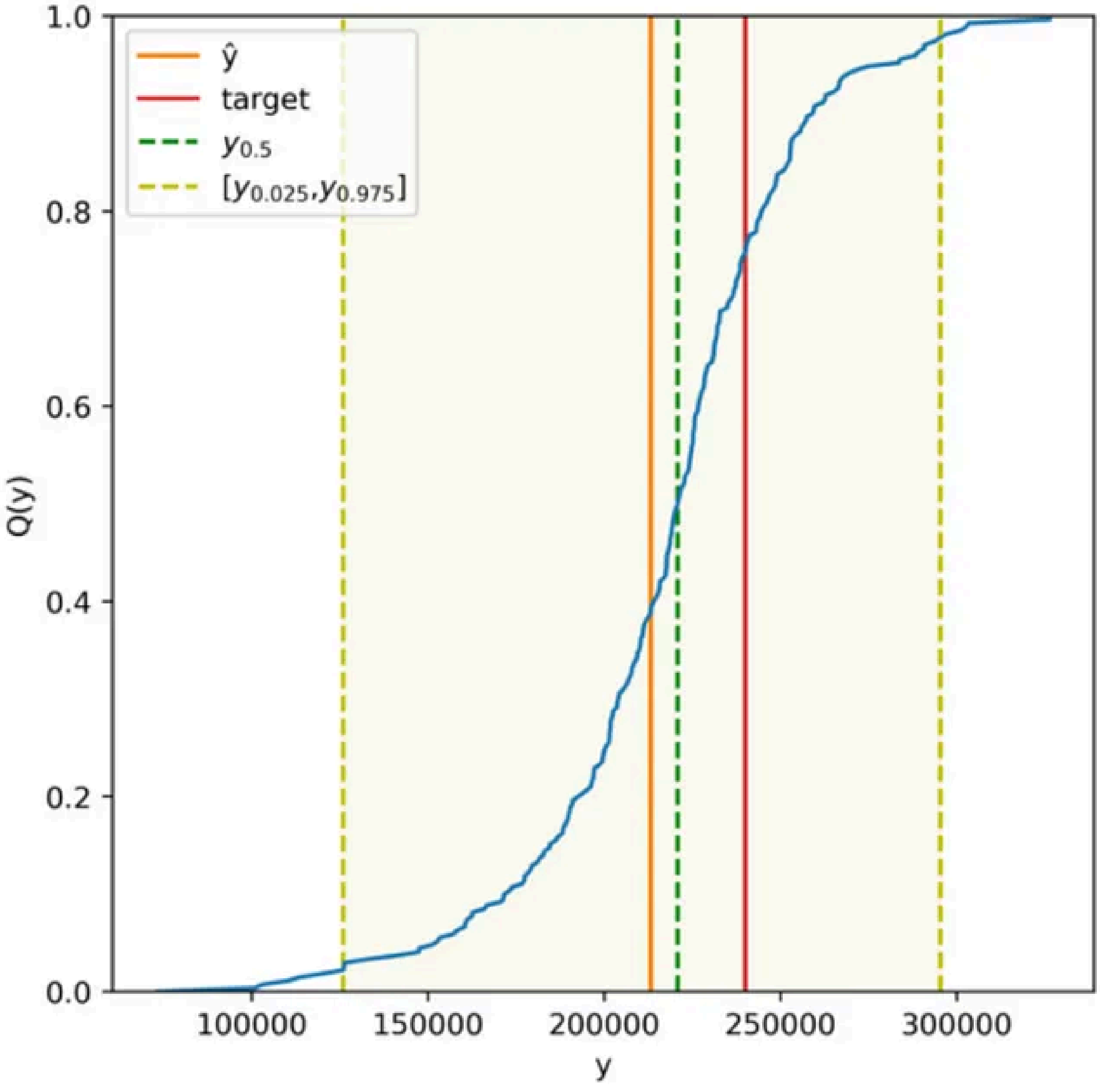


Conformal Predictive Distribution

In other words, CPDs provide **complete uncertainty information** for each prediction, not just a single estimate or fixed interval. Importantly, these distributions come with a **validity guarantee** — meaning that they are calibrated so that, for example, the true value falls below the 90th percentile of the predicted distribution about 90% of the time (over many predictions). This guarantee holds *distribution-free,* i.e. without making strong assumptions on the data or model, which is a key strength of conformal methods.

Example of a conformal predictive distribution (blue curve) produced for a single regression prediction (house price). The CPD is the cumulative distribution function Q(y), representing the predicted probability that the true outcome will be less than y. Vertical lines mark the model's point prediction (orange), the actual target value (red), the median predicted value green dashed, and the 95%

*prediction interval yellow dashed. CPDs thus generalize prediction intervals into a full distribution for each prediction, enabling more informed decisions with **quantified uncertainty** at every possible outcome level.*



Example of a conformal predictive distribution for a test point prediction

As the above figure illustrates, a CPD gives a **complete picture** of uncertainty for an individual prediction. Rather than saying "we predict 200k with ±50k uncertainty," a CPD lets us answer nuanced questions: *What's the probability the value is less than 180k? More than 250k?* By querying the CPD, we can compute **any** percentile or tail probability of interest.

In fact, CPDs allow one to determine the probability of *any event* (e.g. "the patient's blood pressure will exceed 140" or "the stock price will be under $50") by reading off the corresponding cumulative probability. This level of detail is why CPDs have been called *"the Holy Grail of prediction"* — they deliver *full and complete quantification of uncertainty tailored to each test case*. Crucially, they do so **without sacrificing validity**: the CPD is built on conformal prediction's rigorous guarantees, so its confidence levels are empirically grounded. The method simply requires data to be exchangeable (an assumption even milder than the usual i.i.d.), meaning CPDs can be applied alongside most machine learning models with no special retraining

Why Are CPDs Important? (Reliability and Calibration)

The ability to output a full predictive distribution with guaranteed calibration makes CPDs especially valuable in high-stakes and data-critical domains. Here are some key reasons why CPDs are gaining attention:

- **Rigorous Uncertainty Calibration:** CPDs come with a mathematical guarantee of coverage. Unlike Bayesian predictive distributions which in practice are almost always miscalibrated, CPDs *always* achieve the intended coverage frequency by design. This means decision-makers can trust the probabilities from a CPD (e.g. a "70% chance of rain tomorrow" truly means 70% in the long run), which is essential for risk-sensitive applications.
- **Model-Agnostic and Distribution-Free:** CPDs can be attached to *any* underlying regression or forecasting model — whether a simple linear

model or a complex deep neural network — without altering the model's training. The conformal wrapping ensures validity no matter the model or the underlying data distribution. This makes CPDs a convenient add-on for providing uncertainty for existing ML systems, without requiring probabilistic modeling expertise.

- **Actionable Insights for Each Prediction:** By providing the full distribution, CPDs let users extract rich information for each individual case. For instance, one can derive **prediction intervals** at any confidence level (just take the appropriate quantiles from the CDF), or compute the **probability of exceeding a critical threshold** (useful in domains like medicine for dosage limits, or finance for value-at-risk calculations). This granularity transforms how predictions are used in decision-making. Instead of a one-size-fits-all uncertainty estimate, CPDs tailor the uncertainty to each new input, which can be critical if some cases are inherently more uncertain than others.
- **Trust in Critical Applications:** Perhaps most importantly, CPDs bolster **trust** in AI predictions, especially in critical fields. In healthcare, for example, a model that predicts not just an expected outcome but a well-calibrated distribution of possible outcomes can assist clinicians in understanding worst-case vs. best-case scenarios for a patient. As one review notes, conformal methods (including CPDs) have been used in clinical settings to provide insight into the confidence of individual predictions. This is vital because medical decisions are made per patient, and using an ML model without reliable per-patient uncertainty could lead to serious errors. CPDs help ensure *no prediction goes out without a known error rate*.
- **Similarly, in finance**, having the full distribution means institutions can avoid "black swan" surprises — for instance, they won't unknowingly bet the bank on a model that underestimates tail risks. Biased or

overconfident predictions in finance (say for stock prices or credit risk) can result in multi-million dollar losses, so the validity of CPDs is a major advantage. In short, CPDs address the long-standing need for **reliable uncertainty quantification** in AI, which is increasingly recognized as crucial for deploying models responsibly.

Applications and Impact Across Domains

Since their introduction, CPDs have been applied and tested in numerous domains. Below we highlight some key areas and examples where CPDs (and conformal prediction more broadly) are making a difference:

- **Healthcare & Medicine:** In medical AI, CPDs (via conformal prediction) are being explored to provide *per-patient* risk distributions for diagnoses, prognoses, and treatment outcomes. A 2022 review identified multiple studies using conformal prediction in clinical medicine, demonstrating that these methods can supply important insight into the accuracy of individual patient predictions. This means a diagnostic model could report, for example, a patient's probability distribution for a lab result or disease risk, rather than a single number. Such insight is crucial — using an uncalibrated prediction model in healthcare could lead to misdiagnosis or inappropriate treatment if the model's uncertainty is not understood. CPDs' validity guarantees are therefore invaluable in this domain, ensuring that the "confidence" reported alongside a prediction truly reflects reality. In critical cases (like predicting if a tumor is malignant), having a reliable confidence level for each prediction can guide doctors on whether to trust the model or order further tests.
- **Finance & Insurance:** The finance industry has a natural appetite for understanding risk and uncertainty. CPDs have started gaining traction for tasks like *risk management, portfolio forecasting,* and *stress testing*.

Because CPDs output a full distribution, analysts can derive metrics like Value-at-Risk at any level or simulate various scenarios directly from the model's predictive distribution. Recent work on conformalized time-series prediction highlights how predictive simulations can aid in stress-testing financial models. For instance, instead of forecasting a stock price will be $100 (point estimate) or even providing a 95% interval, a CPD-based approach can give the probability of the stock being below $90 or above $110, etc., allowing more nuanced investment decisions. This helps prevent **overconfidence** in models: traders and risk officers can see the worst-case tails. CPDs mitigate this by flagging the true uncertainty. Insurers similarly benefit by getting full claim distribution forecasts (useful for setting premiums and reserves with known confidence).

- **Climate & Energy:** Climate science and energy systems often require **probabilistic forecasts** — e.g., what is the distribution of temperature or wind power production for tomorrow? CPDs have shown promise here as well. Recent research applied *Conformal Predictive Distribution Systems* to short-term **wind speed forecasting**, comparing them with traditional methods like Quantile Regression Forests. The conformal approach produced slightly more conservative but *more efficient* probability distributions than the benchmark, all at a lower computational cost. In practice, this means wind farm operators get a calibrated distribution of possible wind speeds (or power outputs), helping them plan for variability with confidence. The CPD approach in that study achieved better Continuous Ranked Probability Score (CRPS) than the alternatives, indicating a more accurate overall distribution. Beyond wind, CPDs (and conformal methods) are being explored for weather forecasting, rainfall prediction, and electricity load forecasting. The ability to integrate ensemble weather models with conformal calibration yields forecasts that are both sharp and reliable. In the context of climate change projections, having a distribution-free calibrated approach is incredibly valuable, since model uncertainties are high and data distributions can shift — CPDs naturally handle these challenges by guaranteeing validity even under different distributions or non-standard data.

(Other domains are also leveraging CPDs. For example, in ***manufacturing and supply chain****, CPDs can predict full demand distributions to optimize inventory (avoiding stockouts or oversupply). In* ***autonomous driving****, they could quantify the probability of different outcomes (distances, pedestrian movements) to ensure safe decisions. The versatility of CPDs means any field requiring reliable predictions under uncertainty can potentially benefit.)*

Notable Research and Developments in CPDs

The growing interest in conformal predictive distributions is reflected in a flurry of research expanding and refining the technique. Here are a few **notable papers and advances** that have cited or built upon the original CPD methodology:

- **Cross-Conformal Predictive Distributions (Vovk & Manokhin 2018):** To improve the stability of CPDs especially on limited data, researchers introduced *cross-conformal predictive distributions*. This approach leverages cross-validation (multiple folds of calibration) to smooth out variability in the predictive distribution. By aggregating conformal predictions from several splits, cross-conformal methods maintain validity while often yielding tighter distributions. Vovk and colleagues presented this extension shortly after the original CPD paper, demonstrating how ensemble conformal estimates can enhance efficiency without sacrificing coverage.

- **Conformal Predictive Distributions with Kernels (Vovk & Manokhin 2017):** Another line of work has combined CPDs with kernel methods. The idea is to apply CPDs in conjunction with kernel-based nonconformity measures (inspired by kernel ridge regression or Gaussian processes) to capture nonlinear patterns in data. This "CPD with kernels" approach was discussed by Vovk and collaborators as a way to blend classical kernel regression uncertainty with conformal guarantees. It underscores the flexibility of CPDs: one can plug in sophisticated nonparametric learners and still obtain a valid predictive distribution

- **Mondrian Conformal Predictive Distributions (Boström *et al.* 2021):** The standard CPD method assumes a single calibration pool for all examples (or possibly a normalized version that scales by difficulty). *Mondrian CPDs*partition the feature space into categories (analogous to a decision tree split, or any relevant stratification) and build separate CPDs for each category. This increases flexibility, allowing the shape of the predictive distribution to differ across regions of the input space (addressing heteroscedastic situations where uncertainty varies with the input). Boström and colleagues showed that Mondrian CPDs, when applied to regression forests, significantly improved probabilistic accuracy (as measured by CRPS) compared to standard CPDs. Notably, Mondrian CPDs achieved prediction intervals as tight as those from a normalized conformal regressor, while also improving point predictions of the underlying model. This extension indicates how CPDs can be made more *adaptive*, yielding sharper predictions in practice by accounting for context (e.g., "easy" vs "hard" examples have different distribution widths).

- **Conformal Predictive Distribution Trees (Johansson *et al.* 2023):** Marrying interpretability with uncertainty quantification, this recent work introduced CPD Trees — a form of decision tree where each leaf node contains a conformal predictive distribution. Essentially, the tree splits the data into subpopulations (like any regression tree), and instead of a constant prediction in each leaf, it produces a calibrated distribution for that leaf. This gives human-interpretable segments of the feature space (rules defining the leaf) along with an *algorithmic confidence distribution* for each segment. CPD Trees were shown to provide informative predictions with the transparency of a tree model, which is valuable for explainability. They demonstrate that even inherently interpretable models can be enhanced with conformal distribution outputs, combining the best of both worlds (interpretability + reliable uncertainty). This is a promising direction for fields that require explanations for predictions alongside uncertainty (e.g., policy decisions, personalized medicine).

These developments, among others, show an active and growing research community around CPDs. The methodology has been extended to **time-series forecasting**, where CPDs are used for sequential predictions and updated as new data comes in, and to **classification problems** (producing predictive probability distributions over classes, though classification conformal methods typically yield prediction sets or Venn–Abers predictors for probabilities). The fact that CPDs are being integrated into different algorithms (forests, trees, neural nets, etc.) and different settings (online learning, non-iid data, covariate shift adaptation) indicates a maturing field. Each new paper cites the original CPD concept as the foundation, underlining the academic impact of the 2017 work.

Implementation in Practice (The `crepes` Library and More)

One of the strengths of conformal prediction in general is that it's **practical** — there are open-source tools that make it easy to apply CPDs to your own data. A prime example is the Python package **crepes** (Conformal Regressors and Predictive Systems) developed by Henrik Boström and colleagues. This library implements conformal prediction for regression, including the capability to produce **Conformal Predictive Systems (CPS)** which underpin CPDs. With `crepes`, any regression model (scikit-learn regressors, etc.) can be wrapped to output prediction intervals and distributions with minimal code. After fitting a model and calibrating a conformal predictive system, you can obtain **full CPDs** simply by calling the `predict_cps` function with the option to return the distribution. For example:

```
from crepes import ConformalRegressor
# ... (train your base model as regressor)
cps = ConformalRegressor(base_model)
cps.calibrate(X_calib, y_calib)         # use calibration set for conformal
# Get full CPD for each test instance:
cpd_list = cps.predict_cps(X_test, return_cpds=True)
```

By setting `return_cpds=True`, the library returns the entire distribution for each test instance. Internally, this distribution is represented by the set of p-values over the calibration residuals — essentially giving a stepwise CDF (each calibration point contributes a step). The result can be viewed or post-processed to extract interval estimates, median, quantiles, etc. In the earlier figure, we plotted one such CPD: the blue curve was derived from those p-values for a single test example. The `crepes` documentation even provides utilities to plot or evaluate CPDs easily, and it supports advanced variants like Mondrian categories (just by passing a stratification function during calibration).

Beyond `crepes` in Python, there are other resources too. R packages and MATLAB implementations for conformal prediction exist, and the **"Awesome Conformal Prediction"** repository curates tutorials — including a tutorial on CPDs – and examples (e.g., a Kaggle Notebook demonstrating CPDs on a housing price dataset. These tools mean that the barrier to trying CPDs is low.

Practitioners can start with their favorite model and add a conformal prediction step to get distributional outputs that are guaranteed to be calibrated.

Conclusion

Conformal Predictive Distributions represent a significant leap forward in how we think about predictive uncertainty.

In just a few years since their introduction, they have proven their worth both in theory and in application: delivering *reliable, distribution-free predictive distributions* for everything from medical diagnoses to wind farm energy output.

By providing the **full distribution** of a prediction with guaranteed coverage, CPDs equip us with a deeper understanding of model predictions — effectively turning black-box predictions into quantitatively **trustworthy forecasts**. The milestone of 100+ citations for the original CPD paper attests to the strong interest and validation from the research community.

More importantly, the growing roster of use-cases across healthcare, finance, climate science and beyond shows that CPDs are not just an academic curiosity but a practical tool for **robust AI**.

Moving forward, we anticipate even wider adoption of CPDs as awareness grows.

With user-friendly libraries like `crepes` and ongoing research tackling new challenges (like CPDs under distribution shift or in reinforcement learning), the approach is becoming increasingly accessible.

For data scientists and organizations, the message is clear: if you need **trustworthy uncertainty estimates** for critical decisions, Conformal Predictive Distributions offer a proven, easy-to-implement solution.

By embracing CPDs, one can move from mere predictions to **probabilistic predictions** – enabling smarter decisions that explicitly account for uncertainty.

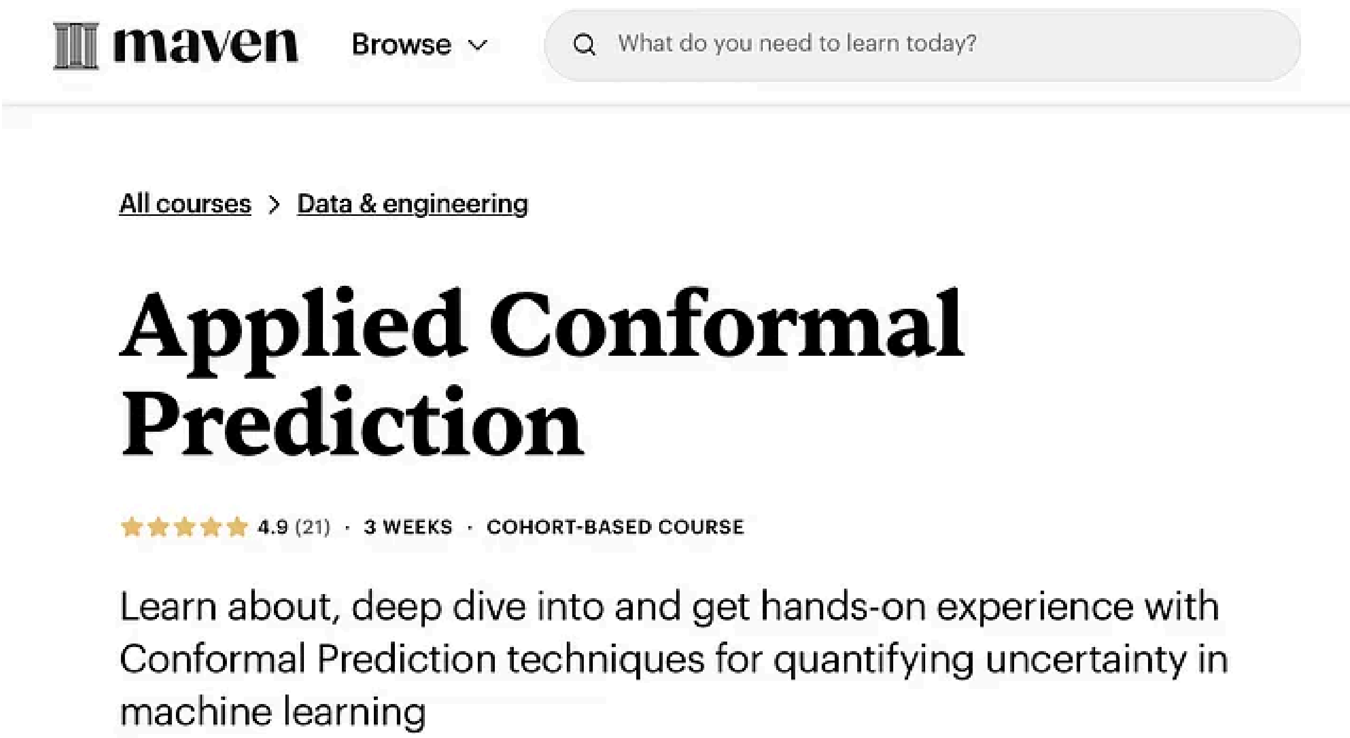
In an era where AI is ever more prevalent, methods like CPDs that guarantee reliability will be key to bridging the gap between complex models and real-world demands for accountability and confidence.

The full probability distribution is no longer just a Bayesian dream; it's available to any predictive model, thanks to conformal prediction. And that is a development worth celebrating

Sources: The concept of conformal predictive distributions and the Least Squares Prediction Machine was introduced by Vovk & Manokhin (2017) in proceedings.mlr.press

If you're eager to master conformal prediction and harness its power for uncertainty quantification in real-world applications, consider enrolling in my highly rated course, "**Applied Conformal Prediction**", or explore my comprehensive book, **“Practical Guide to Applied Conformal Prediction in Python**.

Both resources are designed to help you learn, apply, and integrate cutting-edge uncertainty frameworks directly into your industry use cases.



References

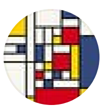
1. **Applied Conformal Prediction, highly rated course.**
2. **“Practical Guide to Applied Conformal Prediction in Python**.
3. **Awesome Conformal Prediction**
4. Nonparametric predictive distributions based on conformal prediction
5. Cross-conformal predictive distributions
6. *Computationally efficient versions of conformal predictive distributions*

7. Conformal predictive distributions with kernels

8. `'crepes` : Conformal Regressors and Predictive Systems’

9. Kaggle Notebook showcasing Conformal Predictive Distributions on Playground Series Season 3, Episode 1 (California Housing data) competittion https://www.kaggle.com/code/predaddict/conformal-predictive-distributions-pss3-e1

Conformal Prediction | Data Science | Machine Learning



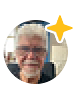
Written by Valeriy Manokhin, PhD, MBA, CQF

2.8K followers · 330 following

Follow

Principal Data Scientist, PhD in Machine Learning, creator of Awesome Conformal Prediction 👍Tip: hold down the Clap icon for up x50

Responses (3)



Alex Mylnikov

What are your thoughts?

Ashmpatel
Apr 28

Very interesting work and the wide range of applications is nice. Looking forward to reading the book. Thankyou

Reply

Marlonlp
Apr 20 (edited)

Thank u for the post, for me trying to get into conformal prediction this is very useful 👍

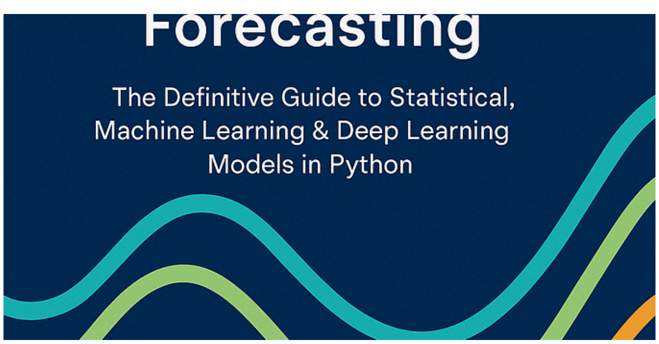
1 reply Reply

Fabio Balloni
Apr 17

cool another ChatGPT article....

Reply

More from Valeriy Manokhin, PhD, MBA, CQF



Valeriy Manokhin, PhD, MBA, CQF

Valeriy Manokhin, PhD, MBA, CQF

Why I'm Writing a New Forecasting Book

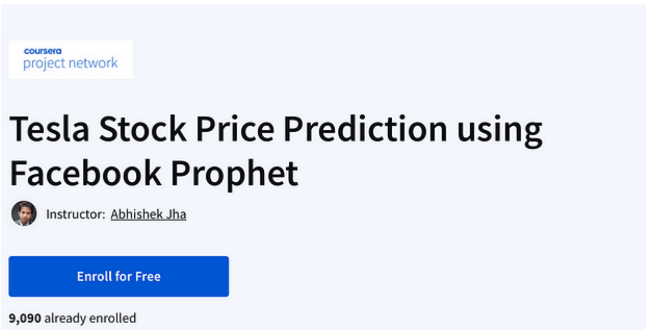
A data science team spent two years replicating an Uber research paper to build a…

Apr 13 140 7

🧪 Why ROC AUC Might Be Misleading — and What to Use…

📍 Introduction: The Rise (and Fall?) of AUC

Mar 30 94 2

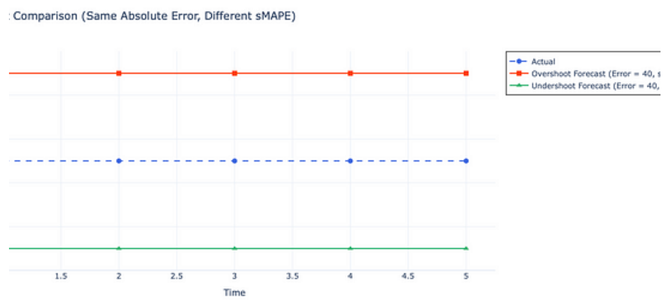


Valeriy Manokhin, PhD, MBA, CQF

The Snake Oil of Applying Facebook Prophet to Financial…

Disclaimer: Hey everyone, a friendly heads-up: the buzz around using Facebook…

Apr 19 89 2



Valeriy Manokhin, PhD, MBA, CQF

MAPE and the Makridakis Competitions: Why Metrics Matter

Introduction: The Metric Behind the Message

Apr 20 88 1

See all from Valeriy Manokhin, PhD, MBA, CQF

Recommended from Medium

Valeriy Manokhin, PhD, MBA, CQF

ARIMA: A 1970s Revolution or 1920s Repackaging? Myth-Bustin...

Introduction

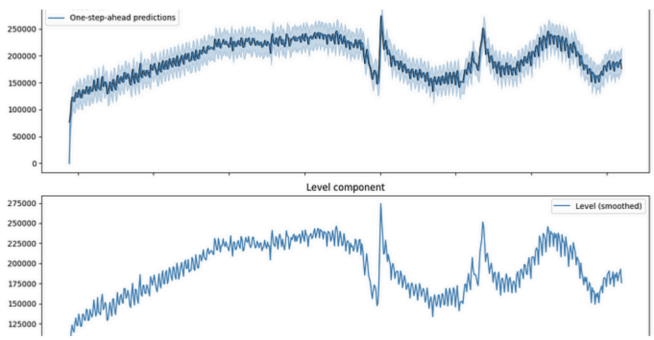
Apr 19 104 2

In Booking.com Engineering by Ivan Shubin

Anomaly Detection in Time Series Using Statistical Analysis

Setting up alerts for metrics isn't always straightforward. In some cases, a simple...

Apr 15 387 7



Kyle Jones

Understanding State Space Models for Time Series Analysis in Python

How to track hidden trends in noisy data

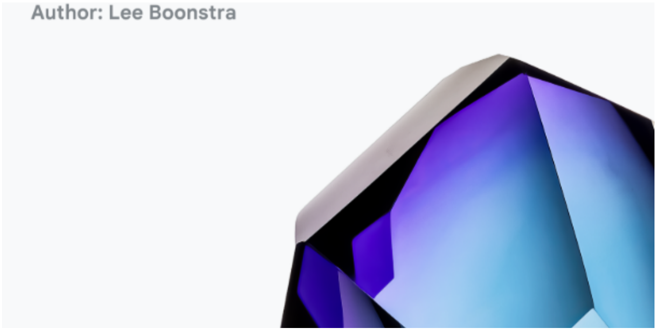
4d ago 29

Unicorn Day

How Markov Chains Can Decode Stock Market Patterns

We've all heard it: "Past performance is not indicative of future results." It's across every...

3d ago 116 5



In Coding Nexus by Algo Insights

Google's 69-Page Prompt Engineering Masterclass: What's…

I've been writing about tech for three years, and let me tell you, nothing's grabbed me…

Apr 13 396 12

Cassie Kozyrkov

Statistics for people in a hurry

All of statistics in under 10 minutes. If you prefer an 🎧audio🎧 version, there's one her…

Apr 8 1.2K 23

See more recommendations