Notes on Quantile Risk Control: A Flexible Framework for Bounding the Probability of High-Loss Predictions

Catherine Chen cyc2152

December 2023

J. C. Snell, T. P. Zollo, Z. Deng, T. Pitassi, and R. Zemel, "Quantile Risk Control: A Flexible Framework for Bounding the Probability of High-Loss Predictions." arXiv, Dec. 27, 2022. Accessed: Dec. 17, 2023. [Online]. Available: http://arxiv.org/abs/2212.13629

Abstract: Presents bounds on quantiles of the loss distribution of a predictor.

1 Introduction

Goal: Produce predictive algorithms that are *rigorous*, i.e. produce bounds that can be trusted with high confidence, and *flexible*, i.e. applicable to an array of loss-related risk measures and adaptive to the difficulty of the instance.

Contribution: Produce lower confidence bounds on the cumulative distribution function (CDF) of a predictor's loss distribution, which can be converted into an upper bound on the quantile function. Provide bounds for risk measures that can be expressed as weighted integrals of the quantile function, a.k.a. quantile-based risk measures (QBRM): expected loss, VaR, CVaR, and VaR-interval.

Definition 1.1 (Expected Loss)**.** Mean loss over the test distribution.

Definition 1.2 (Value at Risk (VaR))**.** The β-VaR measures the maximum loss incurred on a specific quantile, after excluding a $1-\beta$ proportion of high-loss outliers, i.e. maximum loss incurred with probability β.

Definition 1.3 (Conditional Value at Risk (CVaR))**.** The β-CVaR measures the mean loss for the worst $1-\beta$ proportion of the population.

Definition 1.4 (VaR-interval)**.** Optimizing an uncertain loss quantile that belongs to a known range, i.e. for a range of β values.

1.1 Setting and Notation

Assume a black-box predictor $h : \mathcal{Z} \to \hat{\mathcal{Y}}$ that maps from an input space $\mathcal{Z}$ to a space of predictions $\hat{\mathcal{Y}}$.

Assume a loss function $\ell : \hat{\mathcal{Y}} \times \mathcal{Y} \to \mathbb{R}$ that quantifies the quality of a prediction $\hat{Y}$ with respect to the target output Y.

Let (Z, Y) be drawn from an unknown data distribution $\mathcal{D}$ over $\mathcal{Z} \times \mathcal{Y}$ and define the random variable $X \triangleq \ell(h(Z), Y)$ to be the loss induced by h on $\mathcal{D}$.

Recall that the CDF of a random variable X is defined as $F(x) \triangleq P(X \leq x)$; i.e. F is the CDF of the loss RV X

Goal: Produce rigorous upper bounds on the risk $R(F)$ for a class of risk measures $R \in \mathcal{R}$, given a set of validation loss samples $X_{1:n}$.

1.2 Quantile-based Risk Measures

Recall that the quantile function is defined as $F^{-1}(p) \triangleq \inf\{x : F(x) \geq p\}$.

Definition 1.5 (Quantile-based Risk Measure)**.** Let $\psi(p)$ be a weighting function such that $\psi(p) \geq 0$ and $\int_0^1 \psi(p)dp = 1$. The quantile-based risk measure defined by ψ is

$$R_\psi(F) \triangleq \int_0^1 \psi(p)F^{-1}(p)dp. \tag{1}$$

2 Quantile Risk Control

Quantile Risk Control (QRC) achieves rigorous control of quantile-based risk measures (QBRM), by inverting a one-sided goodness-of-fit test statistic to produce a lower confidence bound on the loss CDF. This can subsequently be used to form a family of upper confidence bounds that hold for any QBRM.

More formally, specify a confidence level $\delta \in (0,1)$ and let $X_{(1)} \leq \ldots \leq X_{(n)}$ denote the order statistics of the validation loss samples. QRC consists of the following high-level steps:

1. Choose a one-sided test statistic of the form $S \triangleq \min_{1\leq i\leq n} s_i\left(F\left(X_{(i)}\right)\right)$, where F is the (unknown) CDF of $X_1, \ldots, X_n$.
2. Compute the critical value s_δ such that $P\left(S \geq s_\delta\right) \geq 1-\delta$.
3. Construct a CDF lower confidence bound $\hat{F}_n$ defined by coordinates $\left(X_{(1)}, b_1\right), \ldots, \left(X_{(n)}, b_n\right)$ where $b_i \triangleq s_i^{-1}\left(s_\delta\right)$.
4. For any desired QBRM defined by weighting function ψ, report $R_\psi\left(\hat{F}_n\right)$ as the upper confi dence bound on $R_\psi(F)$.

2.1 CDF Lower Bounds are Risk Upper Bounds

WTS: A lower bound G on the true loss CDF F incurred by a predictor can be used to bound $R_\psi(F)$ for any quantile-based risk measure weighting function.

For CDFs F and G, let $F \succeq G$ denote $F(x) \geq G(x)$ for all $x \in \mathbb{R}$.

Theorem 1. *Let F and G be CDFs such that $F \succeq G$. Then $R_\psi(F) \leq R_\psi(G)$ for any weighting function $\psi(p)$ as defined in Definition 1.5*

Proof. Consider $G^{-1}(p)$. By the definition of the quantile function, $G\left(G^{-1}(p)\right) \geq p$.

$$F \succeq G \implies F(p) \geq G(p) \implies F\left(G^{-1}(p)\right) \geq G\left(G^{-1}(p)\right) \geq p \implies F^{-1}\left(F\left(G^{-1}(p)\right)\right) \geq F^{-1}(p).$$

Since $x \geq F^{-1} \circ F(x)$, this implies $G^{-1}(p) \geq F^{-1}\left(F\left(G^{-1}(p)\right)\right) \implies G^{-1}(p) \geq F^{-1}(p)$.

$R_\psi(F) \leq R_\psi(G)$ holds by Definition 1.5. □

Consider a finite sample setting, $X_{1:n}$:

Definition 2.1 (Lower Confidence Bounds (LCB) on the CDF)**.** Denote $\hat{F}_n$ a $(1-\delta)$-CDF-LCB if for any F, $P\left(F \succeq \hat{F}_n\right) \geq 1-\delta$, where $\hat{F}_n$ is a function of $X_{1:n} \sim^{iid} F$.

Corollary 1. *Suppose that $\hat{F}_n$ is a $(1-\delta)$-CDF-LCB. Then $P\left(R_\psi(F) \leq R_\psi\left(\hat{F}_n\right)\right) \geq 1-\delta$.*

Proof. $F \succeq \hat{G}\left(X_{1:n}\right) \implies R_\psi(F) \leq R_\psi(G)$ by Theorem 1. Result follows immediately. □

2.2 Inverting Goodness-of-fit Statistics to Construct CDF Lower Bounds

WTS: Produce and use a set of lower confidence bounds on the uniform order statistics to bound F.

Let $U_1, \dots, U_n \sim^{iid} \mathrm{U}(0,1)$, with order statistics: $U_{(1)} \leq \dots \leq U_{(n)}$.

Consider a one-sided minimum goodness-of-fit (GoF) statistic of the following form:

$$S \triangleq \min_{1 \leq i \leq n} s_i \left(U_{(i)}\right), \tag{2}$$

where $s_1, \dots, s_n : [0,1] \to \mathbb{R}$ are right continuous monotone nondecreasing functions.

Theorem 2. *Let $s_\delta = \inf_r\{r : P(S \geq r) \geq 1-\delta\}$, where $\delta \in (0,1)$ and S is defined above. Let $X_1, \dots, X_n \sim^{iid} F$, where F is an arbitrary CDF and let $X_{(1)} \leq \dots \leq X_{(n)}$ be the corresponding order statistics. Then $P\left(\forall i : F\left(X_{(i)}\right) \geq s_i^{-1}\left(s_\delta\right)\right) \geq 1-\delta$, where $\forall s \in \mathbb{R}, s_i^{-1}(s) \triangleq \inf_u\{u : s_i(u) \geq s\}$ is the generalized inverse of s_i.*

Proof. For general F, the generalized inverse function if defined as: $F^{-1}(p) \triangleq \inf\{x : F(x) \geq p\}$.

Since

$$P\left(F^{-1}(U) \leq c\right) = P(U \leq F(c)) = F(c).$$

$F^{-1}(U)$ has the same distribution as $X \sim F$, where U is a uniform random variable defined on $[0,1]$. In addition, F^{-1} preserves ordering (by definition of the quantile function), i.e. $F^{-1}\left(U_{(1)}\right) \leq \dots \leq F^{-1}\left(U_{(n)}\right)$. Thus, $X(i)$ has the same distribution as $F^{-1}\left(U_{(i)}\right)$.

Since

$$F \circ F^{-1}(t) \geq t$$

for any $t \in [0,1]$ and any CDF F. Thus,

$$F\left(F^{-1}\left(U_{(i)}\right)\right) \geq U_{(i)} \geq s_i^{-1}\left(s_\delta\right).$$

where the second inequality holds because given the definition of s_δ,

$$P\left(S \geq s_\delta\right) \geq 1-\delta.$$

and definition s_i^{-1} where

$$S \triangleq \min_{1 \leq i \leq n} s_i \left(U_{(i)}\right),$$

$S \geq s_\delta \implies U_{(i)} \geq s_i^{-1}\left(s_\delta\right)$ for all $i \in \{1, \dots, n\}$. Thus,

$$P\left(\forall i, F\left(X_{(i)}\right) \geq s_i^{-1}\left(s_\delta\right)\right) = P\left(\forall i, F\left(F^{-1}\left(U_{(i)}\right) \geq s_i^{-1}\left(s_\delta\right)\right) = P\left(\forall i, U_{(i)} \geq s_i^{-1}\left(s_\delta\right)\right) \geq P\left(S \geq s_\delta\right) \geq 1-\delta$$

since $X(i)$ is of the same distribution as $F^{-1}\left(U_{(i)}\right)$ for all $i \in [n]$. □

2.3 Conservative CDF Completion

WTS: Constraints given by Theorem 2 form a $(1-\delta)$-CDF-LCB via conservative completion of the CDF F given the order statistics of a finite sample from F that is defined by a set of functions $s_1, \dots, s_n$.

Theorem 3. *Let F be an arbitrary CDF satisfying $F\left(x_i\right) \geq b_i$ for all $i \in \{1, \dots, n\}$, where $x_1 \leq \dots \leq x_n$ and $0 \leq b_1 \leq \dots \leq b_n < 1$. Let us denote $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ and $\mathbf{b} = (b_1, b_2, \cdots, b_n)$ and let $x^+ \in \mathbb{R} \cup \infty$ be an upper bound on $X \sim F$, i.e. $F\left(x^+\right) = 1$. Let*

$$G_{(\mathbf{x},\mathbf{b})}(x) \triangleq \begin{cases} 0, & \text{for } x < x_1 \\ b_1, & \text{for } x_1 \leq x < x_2 \\ \dots & \\ b_n, & \text{for } x_n \leq x < x+ \\ 1, & \text{for } x^+ \leq x. \end{cases}$$

Then, $F \succeq G_{(\mathbf{x},\mathbf{b})}$.

Proof.

$$F(x_i) \geq \begin{cases} 0, & \text{for } x < x_1 \implies F(x) \geq 0 \\ \dots & \\ b_i, & \text{for } x_i \leq x < x_{i+1} \implies F(x) \geq F(x_i) \geq b_i \\ \dots & \\ 1, & \text{for } x^+ \leq x. \end{cases}$$

If $x \geq x^+,\ F(x) = 1$. Therefore, $F(x) \geq G_{(\mathbf{x},\mathbf{b})}(x)$. ☐

Corollary 2. *Let S be the test statistic $S \triangleq \min_{1 \leq i \leq n} s_i\left(U_{(i)}\right)$ and $s_\delta = \inf_r\{r : P(S \geq r) \geq 1-\delta\}$. Let $X_1, \dots, X_n \sim$ iid F, where F is an arbitrary CDF and let $X_{(1)} \leq \dots \leq X_{(n)}$ be the corresponding order statistics. Then with probability at least $1-\delta, F \succeq G_{\left(X_{(1:n)}, s_{1:n}^{-1}(s_\delta)\right)}$, where $X_{(1:n)} \triangleq \left(X_{(1)}, \dots, X_{(n)}\right)$ and $s_{1:n}^{-1}(s_\delta) \triangleq \left(s_1^{-1}(s_\delta), \dots, s_n^{-1}(s_\delta)\right)$.*

Proof. By Theorem 3, $\forall i, F\left(X_{(i)}\right) \geq s_i^{-1}(s_\delta) \implies F(x) \geq G_{\left(X_{(1:n)} s_{1:n}^{-1}(s_\delta)\right)}$, which implies

$$P\left(F(x) \geq G_{\left(X_{(1:n)}, s_{1:n}^{-1}(s_\delta)\right)}(x)\right) \geq P\left(\forall i, F\left(X_{(i)}\right) \geq s_i^{-1}(s_\delta)\right).$$

Then, by Theorem 2, $P\left(\forall i, F\left(X_{(i)}\right) \geq s_i^{-1}(s_\delta)\right) \geq 1-\delta$. ☐

Therefore, $\hat{F}_n \triangleq G_{\left(X_{(1:n)}, s_{1:n}^{-1}(s_\delta)\right)}$ is a $(1-\delta)$-CDF-LCB that can be used to select a predictor to minimize a risk measure.

2.4 Interpreting Standard Tests of Uniformity as Minimum-type Statistics S

Several standard tests of uniformity may be viewed as instances of the minimum-type statistic $S \triangleq \min_{1 \leq i \leq n} s_i\left(U_{(i)}\right)$.

1. A one-sided **Kolmogorov-Smirnov (KS) statistic** with the uniform as the null hypothesis can be expressed as

$$D_n^+ \triangleq \max_{1 \leq i \leq n}\left(\frac{i}{n} - U_{(i)}\right) = -\min_{1 \leq i \leq n}\left(U_{(i)} - \frac{i}{n}\right),$$

   Thus $D_n^+ = -S$, where $s_i(u) = u - \frac{i}{n}$.

2. A one-sided **Berk-Jones (BJ) statistic** is defined as

$$M_n^+ \triangleq \min_{1 \leq i \leq n} I_{U_{(i)}}(i, n-i+1)$$

   where $I_x(a,b)$ is the CDF of Beta(a,b) evaluated at x. Thus $M_n^+ = S$ where $s_i(u) = I_u(i, n-i+1)$.

2.5 Bounding Multiple Predictors Simultaenously

Let ϕ be an arbitrary black box function and $t_1, \dots, t_m$ be a set of functions, e.g. different thresholding operations. Then define a finite set of predictors $\mathcal{H} = \{t_1 \circ \phi, \dots, t_m \circ \phi\}$. The loss R.V. for a random predictor $h \in \mathcal{H}$ is denoted by $X^h \sim F^h$ with corresponding validation loss samples $X_{1:n}^h$.

Theorem 4. *Suppose that $\hat{F}_n$ is a $(1-\delta/|\mathcal{H}|)$-CDF-LCB. Then $P\left(\forall h \in \mathcal{H} : F^h \succeq \hat{F}_n^h\right) \geq 1-\delta$.*

Proof. By applying a union bound argument over $h \in \mathcal{H}$:

$$P\left(\exists h \in \mathcal{H} : F^h \not\succeq \hat{G}\left(X_{1:n}^h\right)\right) \leq \sum_{h \in \mathcal{H}} P\left(F^h \not\succeq \hat{G}\left(X_{1:n}^h\right)\right) \leq \delta.$$

☐

2.6 Novel Truncated Berk-Jonest Statistics

Contribution: Novel goodness-of-fit statistics: two forms of a truncated Berk-Jones statistic, targeting a range of quantiles.

1. **Truncated one-sided Berk-Jones** targets a quantile interval $[\beta_{\min}, 1)$, defined as:

$$M_{n,k}^{+} \triangleq \min_{k \leq i \leq n} I_{U_{(i)}}(i, n-i+1)$$

   by dropping lower order statistics that do not affect the bound on $F^{-1}(\beta_{\min})$. The statistic can be realized by using

$$s_i(u) = \begin{cases} I_u(i, n-i+1), & \text{for } k \leq i \leq n \\ 1, & \text{otherwise} \end{cases}$$

   - Given $\beta_{\min}$, define $k^*(\beta_{\min}) \triangleq \min\left\{k : s_k^{-1}\left(s_\delta^k\right) \geq \beta_{\min}\right\}$, where s_δ^k is the critical value of $M_{n,k}^{+}$.
   - Bisection search can be used to compute k^*, then the inversion of $M_{n,k^*(\beta_{\min})}^{+}$ provides a CDF lower bound targeted at quantiles $\beta_{\min}$ and above.

2. **Truncated two-sided Berk-Jones** targets a quantile interval $[\beta_{\min}, \beta_{\max}]$, defined as:

$$M_{n,k,\ell}^{+} \triangleq \min_{k \leq i \leq \ell} I_{U_{(i)}}(i, n-i+1)$$

   - Compute $k^*(\beta_{\min})$ as in the one-sided case
   - Remove higher order statistics using the upper endpoint of $\ell^* \triangleq \min\left\{\ell : s_\ell^{-1}\left(s_\delta^{k^*,\ell}\right) \geq \beta_{\max}\right\}$, where $s_\delta^{k^*,\ell}$ is the corresponding critical value.