On Robust Risk Analysis

Jose Blanchet (based on work with Y. Kang and K. Murthy)

Columbia University.
Department of Statistics,
Department of IEOR.

A Little Story: Motivation and Goal

We want to develop a systematic, data driven, approach for stress testing.

Formalizing

- Risk factors are encoded as $X \in R^d$ and the exposure is $h(X)$ (a function of the risk factors).

Formalizing

- Risk factors are encoded as $X \in R^d$ and the exposure is $h(X)$ (a function of the risk factors).
- Want to compute (for simplicity)

$$Risk = E_{true}(h(X))$$

Formalizing

- Risk factors are encoded as $X \in R^d$ and the exposure is $h(X)$ (a function of the risk factors).
- Want to compute (for simplicity)

$$Risk = E_{true}(h(X))$$

- $P_{true}(\cdot)$ represents the **unknown** probabilistic law of X <— this is a problem.

Formalizing

- Risk factors are encoded as $X \in R^d$ and the exposure is $h(X)$ (a function of the risk factors).
- Want to compute (for simplicity)

$$Risk = E_{true}(h(X))$$

- $P_{true}(\cdot)$ represents the **unknown** probabilistic law of X <— this is a problem.
- **How do we estimate $E_{true}(h(X))$ combining empirical sample and what-if scenarios in a meaningful way?**

- Let's say the **BANK** has an empirical sample $X_1, ..., X_n$, i.i.d. copies of X.

Empirical Distributions and Stress Testing

- Let's say the **BANK** has an empirical sample $X_1, ..., X_n$, i.i.d. copies of X.
- Natural non-parametric estimator

$$\frac{1}{n}\sum_{j=1}^{n} X_j.$$

Empirical Distributions and Stress Testing

- Let's say the **BANK** has an empirical sample $X_1, ..., X_n$, i.i.d. copies of X.
- Natural non-parametric estimator

$$\frac{1}{n}\sum_{j=1}^{n} X_j.$$

- Assume that the **REGULATOR** produces $Y_1, ..., Y_n$ i.i.d. copies from some r.v. Y. <— Think "WHAT-IF" distribution.

Empirical Distributions and Stress Testing

- Let's say the **BANK** has an empirical sample $X_1, ..., X_n$, i.i.d. copies of X.
- Natural non-parametric estimator

$$\frac{1}{n}\sum_{j=1}^{n} X_j.$$

- Assume that the **REGULATOR** produces $Y_1, ..., Y_n$ i.i.d. copies from some r.v. Y. <— Think "WHAT-IF" distribution.
- **How do we incorporate the scenarios $Y_1, ..., Y_n$ as a form of stress testing?**

Incorporating Scenarios as Stress Testing

- **Step 1:** Define $Z_i = X_i$, for $i = 1, ..., n$ and $Z_{n+j} = Y_j$ for $j = 1, ..., n$ (put *ALL* scenarios X's and Y's together).

Incorporating Scenarios as Stress Testing

- **Step 1:** Define $Z_i = X_i$, for $i = 1, ..., n$ and $Z_{n+j} = Y_j$ for $j = 1, ..., n$ (put ALL scenarios X's and Y's together).
- **Step 2:** Let

$$\mu_n(dx) = \frac{1}{n}\sum_{j=1}^{n}\delta_{\{X_j\}}(dx) \quad <- \text{ empirical. measure.}$$

$$v_n(dz) = \sum_{j=1}^{2n}\delta_{\{Z_j\}}(dx)\,w(j) \quad <- \text{ prob. measure.}$$

Incorporating Scenarios as Stress Testing

- **Step 1:** Define $Z_i = X_i$, for $i = 1, ..., n$ and $Z_{n+j} = Y_j$ for $j = 1, ..., n$ (put *ALL* scenarios X's and Y's together).
- **Step 2:** Let

$$\begin{aligned} \mu_n(dx) &= \frac{1}{n}\sum_{j=1}^{n}\delta_{\{X_j\}}(dx) \text{ <- empirical. measure.} \\ v_n(dz) &= \sum_{j=1}^{2n}\delta_{\{Z_j\}}(dx)\,w(j) \text{ <- prob. measure.} \end{aligned}$$

- **Step 3:** Consider

$$\sup_{v_n(\cdot)}\{E_{v_n}(h(Z)) : d(v_n, \mu_n) \leq \delta\}.$$

Natural Questions...

$$\sup_{v_n(\cdot)} \{ E_{v_n}(h(Z)) : d(v_n, \mu_n) \leq \delta \}.$$

- How to choose uncertainty region?

Natural Questions...

$$\sup_{v_n(\cdot)} \{ E_{v_n}(h(Z)) : d(v_n, \mu_n) \leq \delta \}.$$

- How to choose uncertainty region?
- How is this different from *distributionally* robust optimization?

Natural Questions...

$$\sup_{v_n(\cdot)} \{ E_{v_n}(h(Z)) : d(v_n, \mu_n) \leq \delta \}.$$

- How to choose uncertainty region?
- How is this different from *distributionally* robust optimization?
- OK, so what's the new stuff here?

Choosing The Region and Connections to Distributionally Robust Optimization

- We advocate choosing Wasserstein's distance

$$d\left(v_n, \mu_n\right) = \min\{\sum_{i,j} \pi\left(i,j\right) \left|Z_j - X_i\right|_2 : \pi_Z = v_n, \pi_X = \mu_n\}.$$

Choosing The Region and Connections to Distributionally Robust Optimization

- We advocate choosing Wasserstein's distance

$$d\left(v_n, \mu_n\right) = \min\{\sum_{i,j} \pi\left(i,j\right)\left|Z_j - X_i\right|_2 : \pi_Z = v_n, \pi_X = \mu_n\}.$$

- Other regions based on divergence (such as Kullback-Leibler) - Hansen & Sargent (2016)

$$D\left(v_n || \mu_n\right) = E_{v_n}\left(\log\left(\frac{dv_n}{d\mu_n}\right)\right).$$

Choosing The Region and Connections to Distributionally Robust Optimization

- We advocate choosing Wasserstein's distance

$$d\left(v_n, \mu_n\right) = \min\{\sum_{i,j} \pi\left(i,j\right)\left|Z_j - X_i\right|_2 : \pi_Z = v_n, \pi_X = \mu_n\}.$$

- Other regions based on divergence (such as Kullback-Leibler) - Hansen & Sargent (2016)

$$D\left(v_n || \mu_n\right) = E_{v_n}\left(\log\left(\frac{dv_n}{d\mu_n}\right)\right).$$

- Robust Optimization: Ben-Tal, El Ghaoui, Nemirovski (2009).

Choosing The Region and Connections to Distributionally Robust Optimization

- We advocate choosing Wasserstein's distance

$$d\left(v_n, \mu_n\right) = \min\{\sum_{i,j} \pi\left(i,j\right)\left|Z_j - X_i\right|_2 : \pi_Z = v_n, \pi_X = \mu_n\}.$$

- Other regions based on divergence (such as Kullback-Leibler) - Hansen & Sargent (2016)

$$D\left(v_n || \mu_n\right) = E_{v_n}\left(\log\left(\frac{dv_n}{d\mu_n}\right)\right).$$

- Robust Optimization: Ben-Tal, El Ghaoui, Nemirovski (2009).
- **It is crucial that $v_n\left(Y_j\right) > 0$: WE MUST VIOLATE ABSOLUTE CONTINUITY TO DO STRESS TESTING.**

d

What's New Here?

- Distributionally robust stochastic programming using Wasserstein distance: Kuhn et al. (2015)...

What's New Here?

- Distributionally robust stochastic programming using Wasserstein distance: Kuhn et al. (2015)...
- Distributionally robust stress testing formulation is **NEW** here.

What's New Here?

- Distributionally robust stochastic programming using Wasserstein distance: Kuhn et al. (2015)...
- Distributionally robust stress testing formulation is **NEW** here.
- **Main Contribution (to explain in the sequel):**

We explain how to optimally select δ and obtain confidence intervals.

Profile Function

- **Introducing Wasserstein's Profile Function**

$$
\begin{aligned}
& R_n(\gamma) \\
= \;& \min\{\sum_{i,j} \pi(i,j) \, |Z_j - X_i|_2 : \pi_Z = v_n, \pi_X = \mu_n, E_{v_n}(h(Z)) = \gamma\} \\
= \;& \min\{\sum_{i,j} \pi(i,j) \, |Z_j - X_i|_2 : \pi_X = \mu_n, E_{\pi}(h(Z)) = \gamma\}.
\end{aligned}
$$

Profile Function

- **Introducing Wasserstein's Profile Function**

$$
\begin{aligned}
& R_n(\gamma) \\
= \;& \min\{\sum_{i,j} \pi(i,j)\,|Z_j - X_i|_2 : \pi_Z = v_n, \pi_X = \mu_n, E_{v_n}(h(Z)) = \gamma\} \\
= \;& \min\{\sum_{i,j} \pi(i,j)\,|Z_j - X_i|_2 : \pi_X = \mu_n, E_\pi(h(Z)) = \gamma\}.
\end{aligned}
$$

- Idea borrowed from Empirical Likelihood, Owen (1988).

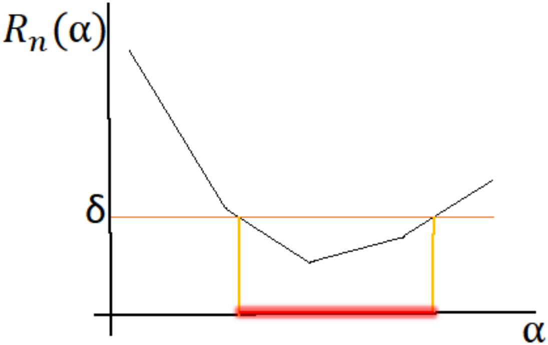
Asymptotic Distribution for Wasserstein Profile Function

Theorem (B. and Kang 2016)

Suppose $h(X)$ *has a density* $f(\cdot)$, $h(Y)$ *has density* $g(\cdot)$, *and* $E\left(h(X)^2 + h(Y)^2\right) < \infty$. *Then under the null hypothesis, i.e.* $H_0 : \gamma = E_{true}(h(X))$,

$$nR_n(\gamma) \to \kappa\chi_1^2.$$

How to Use The Wasserstein Profile Function?



$$P(R_n(\gamma) < \delta) \approx P(\chi^2 < \delta n/\kappa) = .95$$

Under H_0 we have P(γ ϵ Red interval)≈.95

Incorporating Scenarios as Stress Testing

- Compute δ so that $P\left(\chi_1^2 \leq \delta n/\kappa\right) = .95$ and solve the LP

$$\max \sum_{j=1}^{2n} h\left(Z_j\right) w\left(j\right)$$

$$\sum_i \pi\left(i,j\right) \left|X_i - Z_j\right|_2 \leq \delta \ \ \forall j$$

$$\sum_i \pi\left(i,j\right) = w\left(j\right) \ \ \forall j, \ \ \sum_j \pi\left(i,j\right) = \frac{1}{n} \ \ \forall i$$

$$\pi\left(i,j\right) \geq 0 \ \forall i,j$$

Asymptotic Distribution for Wasserstein Profile Function

Theorem (B. and Kang 2016)

Suppose $H(X) = (h_1(X), ..., h_d(X))$ *has a density* $f(\cdot)$*,* $H(Z)$ *has density* $g(\cdot)$*, and* $E\left(h_i(X)^2 + h_i(Y)^2\right) < \infty$*. Then under the null hypothesis, i.e.* $H_0 : \gamma_i = E_{true}(h_i(X))$ *for all* i*, then*

- *When* $d = 1$*,*
$$nR_n(\gamma) \Rightarrow \kappa_1 \chi_1^2.$$
- *When* $d = 2$*,*
$$nR_n(\gamma) \Rightarrow \kappa_2 U^T Var(H(X)) U \text{ with } U \sim N(0, I).$$
- *When* $d \geq 3$*,*
$$n^{1/2+\frac{3}{2d+2}} R_n(\gamma) \Rightarrow \kappa_d \left(\sqrt{U^T Var(H(X)) U}\right)^{1+1/(d+1)}.$$

An Interesting Connection to Machine Learning

Theorem (B. Kang and Murthy 2016)

Suppose you want to find β *such that*

$$\inf_{\beta} \sup_{d_C(P,\mu_n)\leq\delta} E_P[\left\| Y - \beta^T X \right\|_2^2],$$

then choosing a suitably chosen function $C(\cdot)$ *the formulation turns out to be equivalent to genearlized LASSO. And regularization parameter corresponds to* δ*. So one can choose it without using cross-validation.*

Now to Infinite Dimensions

Let's move to robust performance analysis of stochastic processes...

Robust Performance Analysis for Stochastic Processes

Theorem (B. & Murthy (2016))

Suppose X *takes values on a Polish space* $\mathcal{S}$. *Let*

$$C(x,y) : \mathcal{S} \times \mathcal{S} \rightarrow [0,\infty)$$

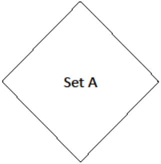
satisfy $C(x,x) = 0$, $C(x,y) \leq C(x,z) + C(z,y)$, *and lower semicontinuous. Consider for* A *closed*

$$\begin{aligned} OPT &= \sup P(Y \in A) \\ \text{s.t. } (X,Y) \text{ satisfies } &: E(C(X,Y)) \leq \delta \text{ and } X \text{ follows } P_0. \end{aligned}$$

Then,

$$OPT = P_0(X \in B(\delta)) = P_0\left(\inf_{y \in A} C(X,y) \leq 1/\lambda^*(\delta)\right).$$

Intuition for Multidimensional rv's



Question:

$\max P(Y \in A)$

$EC(X,Y) \leq \delta$

$X \text{ follows } P_0$

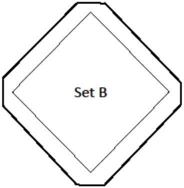
Answer:

$P_0(X \in B)$

Shape of Set B
Depends on C (.)

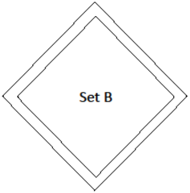
Cost:

$C(X,Y) = |X - Y|_\infty$



Cost:

$C(X,Y) = |X - Y|_2$



Conclusions: We Can Robustify in Great Generality!

- New inference methodology designed to incorporate stress-testing scenarios.

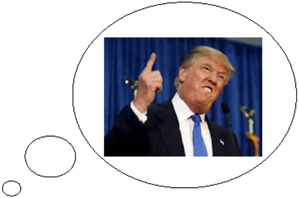
Conclusions: We Can Robustify in Great Generality!

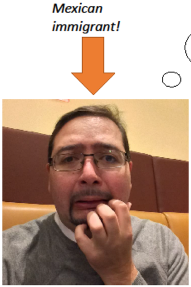
- New inference methodology designed to incorporate stress-testing scenarios.
- New robust performance analysis analysis results for stochastic processes.

Conclusions: We Can Robustify in Great Generality!

- New inference methodology designed to incorporate stress-testing scenarios.
- New robust performance analysis analysis results for stochastic processes.
- Last word of caution: there is stuff that is just too bad to be robustified...

Not Everything Can be Robustified...





Now that we can robustify in great generality...

Nothing can possibly go wrong... Right?