Optimal Transport in Risk Analysis

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

Stanford University (Management Science and Engineering), and Columbia University (Department of Statistics and Department of IEOR).

Goal:

Goal: Present a comprehensive framework for decision making under model uncertainty...

This presentation is an invitation to read these two papers:

https://arxiv.org/abs/1604.01446

https://arxiv.org/abs/1610.05627

Example: Model Uncertainty in Ruin Probabilities

- $R(t)$ = the reserve (perhaps multiple lines) at time t.

Example: Model Uncertainty in Ruin Probabilities

- $R(t)$ = the reserve (perhaps multiple lines) at time t.
- Ruin probability (in finite time horizon T)

$$u_T = P_{true}\left(R(t) \in B \text{ for some } t \in [0, T]\right).$$

Example: Model Uncertainty in Ruin Probabilities

- $R(t)$ = the reserve (perhaps multiple lines) at time t.
- Ruin probability (in finite time horizon T)

$$u_T = P_{true}\left(R(t) \in B \text{ for some } t \in [0, T]\right).$$

- B is a set which models bankruptcy.

Example: Model Uncertainty in Ruin Probabilities

- $R(t)$ = the reserve (perhaps multiple lines) at time t.
- Ruin probability (in finite time horizon T)

$$u_T = P_{true}\left(R(t) \in B \text{ for some } t \in [0, T]\right).$$

- B is a set which models bankruptcy.
- **Problem:** Model (P_{true}) may be complex, intractable or simply unknown...

A Distributionally Robust Risk Analysis Formulation

- **Our solution:** Estimate u_T by solving

$$\sup_{D_c(P_0,P)\leq\delta} P\left(R\left(t\right) \in B \text{ for some } t \in \left[0, T\right]\right),$$

where P_0 is a *suitable* model.

A Distributionally Robust Risk Analysis Formulation

- **Our solution:** Estimate u_T by solving

$$\sup_{D_c(P_0,P)\leq\delta} P\left(R\left(t\right)\in B \text{ for some } t\in\left[0,T\right]\right),$$

  where P_0 is a *suitable* model.

- P_0 = proxy for P_{true}.

A Distributionally Robust Risk Analysis Formulation

- **Our solution:** Estimate u_T by solving

$$\sup_{D_c(P_0,P)\leq\delta} P\left(R\left(t\right) \in B \text{ for some } t \in \left[0, T\right]\right),$$

  where P_0 is a *suitable* model.

- P_0 = proxy for P_{true}.
- P_0 right trade-off between fidelity and tractability.

A Distributionally Robust Risk Analysis Formulation

- **Our solution:** Estimate u_T by solving

$$\sup_{D_c(P_0,P)\leq\delta} P\left(R\left(t\right) \in B \text{ for some } t \in \left[0, T\right]\right),$$

  where P_0 is a *suitable* model.

- P_0 = proxy for P_{true}.
- P_0 right trade-off between fidelity and tractability.
- δ is the distributional uncertainty size.

A Distributionally Robust Risk Analysis Formulation

- **Our solution:** Estimate u_T by solving

$$\sup_{D_c(P_0,P)\leq\delta} P\left(R\left(t\right) \in B \text{ for some } t \in [0, T]\right),$$

where P_0 is a *suitable* model.

- P_0 = proxy for P_{true}.
- P_0 right trade-off between fidelity and tractability.
- δ is the distributional uncertainty size.
- $D_c\left(\cdot\right)$ is the distributional uncertainty region.

Desirable Elements of Distributionally Robust Formulation

- Would like $D_c(\cdot)$ to have wide flexibility (even non-parametric).

Desirable Elements of Distributionally Robust Formulation

- Would like $D_c(\cdot)$ to have wide flexibility (even non-parametric).
- Want optimization to be tractable.

Desirable Elements of Distributionally Robust Formulation

- Would like $D_c(\cdot)$ to have wide flexibility (even non-parametric).
- Want optimization to be tractable.
- *Want to preserve advantages of using P_0.*

Desirable Elements of Distributionally Robust Formulation

- Would like $D_c(\cdot)$ to have wide flexibility (even non-parametric).
- Want optimization to be tractable.
- *Want to preserve advantages of using* P_0*.*
- Want a way to estimate δ.

Connections to Distributionally Robust Optimization

- Standard choices based on divergence (such as Kullback-Leibler) - Hansen & Sargent (2016)

$$D\left(v||\mu\right) = E_v\left(\log\left(\frac{dv}{d\mu}\right)\right).$$

Connections to Distributionally Robust Optimization

- Standard choices based on divergence (such as Kullback-Leibler) - Hansen & Sargent (2016)

$$D\left(v||\mu\right) = E_v\left(\log\left(\frac{dv}{d\mu}\right)\right).$$

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

Connections to Distributionally Robust Optimization

- Standard choices based on divergence (such as Kullback-Leibler) - Hansen & Sargent (2016)

$$D\left(v||\mu\right) = E_v\left(\log\left(\frac{dv}{d\mu}\right)\right).$$

- Robust Optimization: Ben-Tal, El Ghaoui, Nemirovski (2009).
- **Big problem: Absolute continuity may typically be violated...**

- Standard choices based on divergence (such as Kullback-Leibler) - Hansen & Sargent (2016)

$$D\left(v||\mu\right) = E_v\left(\log\left(\frac{dv}{d\mu}\right)\right).$$

- Robust Optimization: Ben-Tal, El Ghaoui, Nemirovski (2009).
- **Big problem: Absolute continuity may typically be violated...**
- Think of using Brownian motion as a proxy model for $R\left(t\right)$...

Connections to Distributionally Robust Optimization

- Standard choices based on divergence (such as Kullback-Leibler) - Hansen & Sargent (2016)

$$D\left(v||\mu\right) = E_v\left(\log\left(\frac{dv}{d\mu}\right)\right).$$

- Robust Optimization: Ben-Tal, El Ghaoui, Nemirovski (2009).
- **Big problem: Absolute continuity may typically be violated...**
- Think of using Brownian motion as a proxy model for $R\left(t\right)$...
- **We advocate using optimal transport costs (e.g. Wasserstein distance).**

Elements of Optimal Transport Costs

- $\mathcal{S}_X$ and $\mathcal{S}_Y$ be Polish spaces.

Elements of Optimal Transport Costs

- $\mathcal{S}_X$ and $\mathcal{S}_Y$ be Polish spaces.
- $\mathcal{B}_{\mathcal{S}_X}$ and $\mathcal{B}_{\mathcal{S}_Y}$ be the associated Borel σ-fields.

Elements of Optimal Transport Costs

- $\mathcal{S}_X$ and $\mathcal{S}_Y$ be Polish spaces.
- $\mathcal{B}_{\mathcal{S}_X}$ and $\mathcal{B}_{\mathcal{S}_Y}$ be the associated Borel σ-fields.
- $c(\cdot) : \mathcal{S}_X \times \mathcal{S}_X \to [0, \infty)$ be lower semicontinuous.

Elements of Optimal Transport Costs

- $\mathcal{S}_X$ and $\mathcal{S}_Y$ be Polish spaces.
- $\mathcal{B}_{\mathcal{S}_X}$ and $\mathcal{B}_{\mathcal{S}_Y}$ be the associated Borel σ-fields.
- $c(\cdot) : \mathcal{S}_X \times \mathcal{S}_X \to [0, \infty)$ be lower semicontinuous.
- $\mu(\cdot)$ and $v(\cdot)$ Borel probability measures defined on $\mathcal{S}_X$ and $\mathcal{S}_Y$.

Elements of Optimal Transport Costs

- $\mathcal{S}_X$ and $\mathcal{S}_Y$ be Polish spaces.
- $\mathcal{B}_{\mathcal{S}_X}$ and $\mathcal{B}_{\mathcal{S}_Y}$ be the associated Borel σ-fields.
- $c(\cdot) : \mathcal{S}_X \times \mathcal{S}_X \to [0, \infty)$ be lower semicontinuous.
- $\mu(\cdot)$ and $v(\cdot)$ Borel probability measures defined on $\mathcal{S}_X$ and $\mathcal{S}_Y$.
- Given π a Borel prob. measure on $\mathcal{S}_X \times \mathcal{S}_Y$,

$$\pi_X(A) = \pi(A \times \mathcal{S}_Y) \quad \text{and} \quad \pi_Y(C) = \pi(\mathcal{S}_X \times C).$$

Definition of Optimal Transport Costs

- Define

$$D_c(\mu, v) = \min_{\pi}\{E_{\pi}(c(X, Y)) : \pi_X = \mu \text{ and } \pi_Y = v\}.$$

Definition of Optimal Transport Costs

- Define

$$D_c(\mu, v) = \min_{\pi}\{E_\pi(c(X, Y)) : \pi_X = \mu \text{ and } \pi_Y = v\}.$$

- This is the so-called Kantorovich problem (see Villani (2008)).

Definition of Optimal Transport Costs

- Define

$$D_c(\mu, v) = \min_{\pi} \{E_\pi (c(X, Y)) : \pi_X = \mu \text{ and } \pi_Y = v\}.$$

- This is the so-called Kantorovich problem (see Villani (2008)).
- If $c(\cdot)$ is a metric then $D_c(\mu, v)$ is a Wasserstein distance of order 1.

Definition of Optimal Transport Costs

- Define

$$D_c(\mu, v) = \min_{\pi}\{E_\pi(c(X,Y)) : \pi_X = \mu \text{ and } \pi_Y = v\}.$$

- This is the so-called Kantorovich problem (see Villani (2008)).
- If $c(\cdot)$ is a metric then $D_c(\mu, v)$ is a Wasserstein distance of order 1.
- If $c(x,y) = 0$ if and only if $x = y$ then $D_c(\mu, v) = 0$ if and only if $\mu = v$.

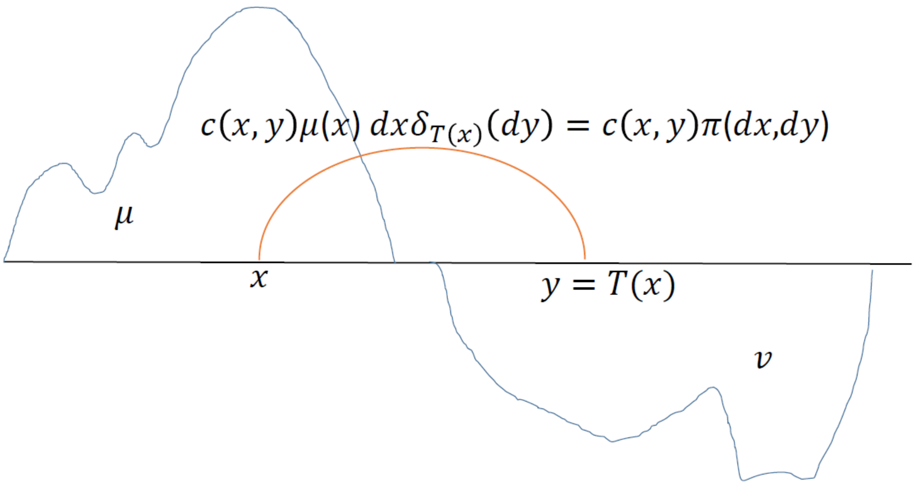
Definition of Optimal Transport Costs

- Define

$$D_c(\mu, v) = \min_{\pi} \{ E_\pi(c(X, Y)) : \pi_X = \mu \text{ and } \pi_Y = v \}.$$

- This is the so-called Kantorovich problem (see Villani (2008)).
- If $c(\cdot)$ is a metric then $D_c(\mu, v)$ is a Wasserstein distance of order 1.
- If $c(x, y) = 0$ if and only if $x = y$ then $D_c(\mu, v) = 0$ if and only if $\mu = v$.
- *Kantorovich's problem is a "nice" infinite dimensional linear programming problem.*

Illustration of Optimal Transport Costs



Fundamental Theorem of Robust Performance Analysis

Theorem (B. and Murthy (2016))

Suppose that $c(\cdot)$ is lower semicontinuous and that $h(\cdot)$ is upper semicontinuous with $E_{P_0}|f(X)| < \infty$. Then,

$$\sup_{D_c(P_0,P)\leq\delta} E_P(f(Y)) = \inf_{\lambda\geq 0} E_{P_0}[\lambda\delta + \sup_z\{f(z) - \lambda c(X,z)\}].$$

Moreover, (π_*) and dual λ_* are primal-dual solutions if and only if

$$\begin{aligned} f(y) - \lambda_* c(x,y) &= \sup_z\{f(z) - \lambda_* c(x,z)\} \quad (x,y)\text{-}\pi_* \text{ a.s.} \\ \lambda_*(E_{\pi_*}[c(X,Y) - \delta]) &= 0. \end{aligned}$$

Implications for Risk Analysis

Theorem (B. and Murthy (2016))

Suppose that $c(\cdot)$ *is lower semicontinuous and that* B *is a closed set. Let* $c_B(x) = \inf\{c(x,y) : y \in B\}$, *then*

$$\sup_{D_c(P_0,P)\leq\delta} P(Y \in B) = P_0\left(c_B(X) \leq 1/\lambda^*\right),$$

where $\lambda^* \geq 0$ *satisfies (under mild assumptions on* $c_B(X)$*)*

$$\delta = E_0\left[c_B(X)\, I\left(c_B(X) \leq 1/\lambda^*\right)\right].$$

- Suppose that

$$
\begin{aligned}
c(x,y) &= d_J(x(\cdot), y(\cdot)) = \text{Skorokhod } J_1 \text{ metric.} \\
&= \inf_{\phi(\cdot) \text{ bijection}} \left\{ \sup_{t\in[0,1]} |x(t) - y(\phi(t))|, \sup_{t\in[0,1]} |\phi(t) - t| \right\}.
\end{aligned}
$$

Application 1: Back to Classical Risk Problem

- Suppose that

$$
\begin{aligned}
c(x,y) &= d_J(x(\cdot), y(\cdot)) = \text{Skorokhod } J_1 \text{ metric.} \\
&= \inf_{\phi(\cdot)\ \text{bijection}} \Big\{ \sup_{t\in[0,1]} |x(t) - y(\phi(t))|, \sup_{t\in[0,1]} |\phi(t) - t| \Big\}.
\end{aligned}
$$

- If $R(t) = b - Z(t)$, then ruin during time interval $[0,1]$ is

$$
B_b = \{ z(\cdot) : b \le \sup_{t\in[0,1]} z(t) \}.
$$

Application 1: Back to Classical Risk Problem

- Suppose that

$$
\begin{aligned}
c(x,y) &= d_J(x(\cdot), y(\cdot)) = \text{Skorokhod } J_1 \text{ metric.} \\
&= \inf_{\phi(\cdot) \text{ bijection}} \left\{ \sup_{t\in[0,1]} |x(t) - y(\phi(t))|, \sup_{t\in[0,1]} |\phi(t) - t| \right\}.
\end{aligned}
$$

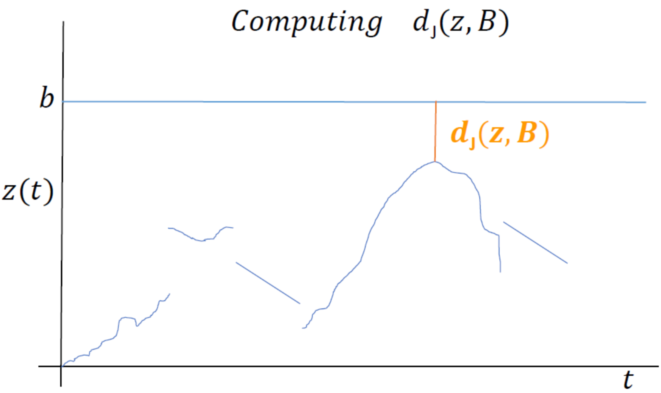
- If $R(t) = b - Z(t)$, then ruin during time interval $[0,1]$ is

$$
B_b = \{z(\cdot) : b \leq \sup_{t\in[0,1]} z(t)\}.
$$

- **Let $P_0(\cdot)$ be the Wiener measure want to compute**

$$
\sup_{D_c(P_0,P)\leq\delta} P(Z \in B_b).
$$

Application 1: Computing Distance to Bankruptcy



- **So:** $\{c_{B_b}(z) \leq 1/\lambda_*\} = \{\sup_{t\in[0,1]} z(t) \geq b - 1/\lambda^*\}$, and

$$\sup_{D_c(P_0,P)\leq\delta} P(Z \in B_b) = P_0\left(\sup_{t\in[0,1]} Z(t) \geq b - 1/\lambda^*\right).$$

- Note **any coupling** π so that $\pi_X = P_0$ and $\pi_Y = P$ satisfies

$$D_c\left(P_0, P\right) \leq E_\pi\left[c\left(X, Y\right)\right] \approx \delta.$$

- Note **any coupling** π so that $\pi_X = P_0$ and $\pi_Y = P$ satisfies

$$D_c\left(P_0, P\right) \leq E_\pi\left[c\left(X, Y\right)\right] \approx \delta.$$

- So use any coupling between *evidence* and P_0 or expert knowledge.

- Note **any coupling** π so that $\pi_X = P_0$ and $\pi_Y = P$ satisfies

$$D_c\left(P_0, P\right) \leq E_\pi\left[c\left(X, Y\right)\right] \approx \delta.$$

- So use any coupling between *evidence* and P_0 or expert knowledge.
- We discuss choosing δ non-parametrically in a moment

- Given arrivals and claim sizes let $Z(t) = m_2^{-1/2} \sum_{k=1}^{N(t)} (X_k - m_1)$

Algorithm 1 To embed the process $(Z(t) : t \geq 0)$ in Brownian motion $(B(t) : t \geq 0)$
Given: Brownian motion $B(t)$, moment m_1 and independent realizations of claim sizes $X_1, X_2, \ldots$

Initialize $\tau_0 := 0$ and $\Psi_0 := 0$. For $j \geq 1$, recursively define,

$$\tau_{j+1} := \inf\left\{ s \geq \tau_j : \sup_{\tau_j \leq r \leq s} B_r - B_s = X_{j+1} \right\}, \text{ and } \Psi_j := \Psi_{j-1} + X_j.$$

Define the auxiliary processes

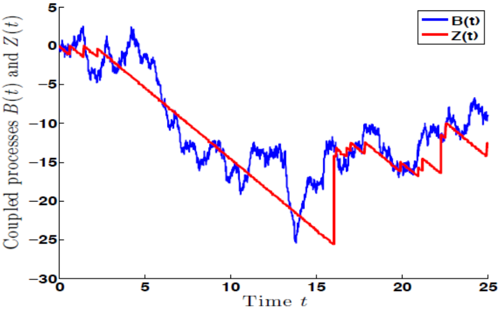
$$\tilde{S}(t) := \sum_{j>0} \sup_{\tau_j \leq s \leq t} B(s) \mathbf{1}(\tau_j \leq t < \tau_{j+1}) \text{ and } \tilde{N}(t) := \sum_{j \geq 0} \Psi_j \mathbf{1}(\tau_j \leq t < \tau_{j+1}).$$

Let $A(t) := \tilde{N}(t) + \tilde{S}(t)$, and identify the time change $\sigma(t) := \inf\{s : A(s) = m_1 t\}$. Next, take the time changed version $Z(t) := \tilde{S}(\sigma(t))$.

Replace $Z(t)$ by $-Z(t)$ and $B(t)$ by $-B(t)$.

Application 1: Coupling in Action

Figure 4. A coupled path output by Algorithm 1

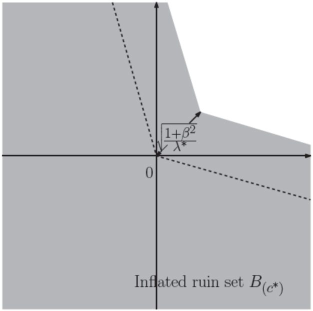


Application 1: Numerical Example

- Assume Poisson arrivals.
- *Pareto claim sizes with index* **2.2** $(P(V > t) = 1/(1+t)^{2.2})$.
- Cost $c(x, y) = d_J(x, y)^2$ <– note power of 2.
- Used Algorithm 1 to calibrate (estimating means and variances from data).

b	$\frac{P_0(\text{Ruin})}{P_{true}(\text{Ruin})}$	$\frac{P^*_{robust}(\text{Ruin})}{P_{true}(\text{Ruin})}$
100	1.07×10^{-1}	12.28
150	2.52×10^{-4}	10.65
200	5.35×10^{-8}	10.80
250	1.15×10^{-12}	10.98

- Paper: Quantifying Distributional Model Risk via Optimal Transport (B. & Murthy '16) **https://arxiv.org/abs/1604.01446** contains more applications
- Multidimensional risk processes (explicit evaluation of $c_B(x)$ for d_J metric).
- Control: $\min_\theta \sup_{P:D(P,P_0)\leq\delta} E[L(\theta, Z)]$ <– robust optimal reinsurance.



(b)Computation of worst-case ruin using the

Connection to machine learning helps further understand why optimal transport costs are sensible choices...

Paper:
Robust Wasserstein Profile Inference and Applications to Machine Learning (B., Murthy & Kang '16)
https://arxiv.org/abs/1610.05627

- Consider estimating $\beta_* \in R^m$ in linear regression

$$Y_i = \beta X_i + e_i,$$

where $\{(Y_i, X_i)\}_{i=1}^n$ are data points.

- Consider estimating $\beta_* \in R^m$ in linear regression

$$Y_i = \beta X_i + e_i,$$

where $\{(Y_i, X_i)\}_{i=1}^n$ are data points.

- Optimal Least Squares approach consists in estimating β_* via

$$MSE(\beta) = \min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \beta^T X_i \right)^2.$$

- Consider estimating $\beta_* \in R^m$ in linear regression

$$Y_i = \beta X_i + e_i,$$

where $\{(Y_i, X_i)\}_{i=1}^n$ are data points.

- Optimal Least Squares approach consists in estimating β_* via

$$MSE\,(\beta) = \min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \beta^T X_i \right)^2.$$

- Apply the distributionally robust estimator based on optimal transport.

Connection to Sqrt-Lasso

Theorem (B., Kang, Murthy (2016)) Suppose that

$$c\left(\left(x,y\right),\left(x',y'\right)\right)=\begin{cases}\left\|x-x'\right\|_q^2 & \text{if} \quad y=y' \\ \infty & \text{if} \quad y\neq y'\end{cases}.$$

Then, if $1/p+1/q=1$

$$\max_{P:D_c(P,P_n)\leq\delta} E_P^{1/2}\left(\left(Y-\beta^T X\right)^2\right)=\sqrt{MSE\left(\beta\right)}+\sqrt{\delta}\left\|\beta\right\|_p^2.$$

Remark 1: This is sqrt-Lasso (Belloni et al. (2011)).
Remark 2: Also representations for support vector machines, LAD lasso, group lasso, adaptive lasso, and more!

For Instance... Regularized Estimators

Theorem (B., Kang, Murthy (2016)) Suppose that

$$c\left(\left(x,y\right),\left(x',y'\right)\right) = \begin{cases} \left\|x - x'\right\|_q & \text{if} \quad y = y' \\ \infty & \text{if} \quad y \neq y' \end{cases}.$$

Then,

$$\sup_{P:\ \mathcal{D}_c(P,P_n)\leq\delta} E_P\left[\log(1+e^{-Y\beta^T X})\right] = \frac{1}{n}\sum_{i=1}^{n}\log(1+e^{-Y_i\beta^T X_i}) + \delta\left\|\beta\right\|_p.$$

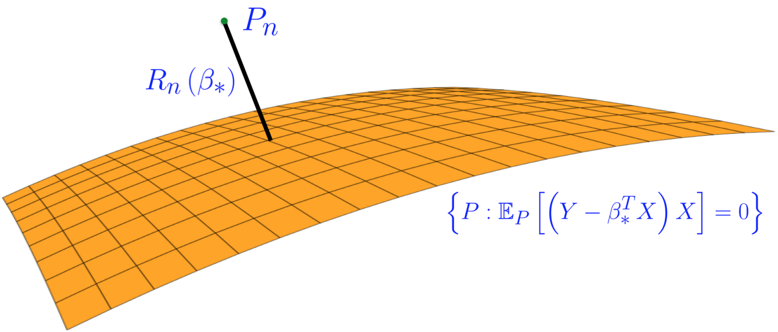
Remark 1: This is regularized logistic regression (see also Esfahani and Kuhn 2015).

Connections to Empirical Likelihood

Paper:
https://arxiv.org/abs/1610.05627

Also chooses δ *optimally* introducing
an extension of Empirical Likelihood
called "Robust Wasserstein Profile Inference".

The Robust Wasserstein Profile Function



Pick $\delta = 95\%$ quantile of $R_n(\beta_*)$ and we show that

$$nR_n(\beta_*) \approx_d \frac{E[e^2]}{E[e^2] - (E|e|)^2} \|N(0, Cov(X))\|_q^2.$$

Conclusions

- Presented systematic approach for quantifying model misspecification.

Conclusions

- Presented systematic approach for quantifying model misspecification.
- Approach based on optimal transport

$$\sup_{D(P,P_0)\leq\delta} P\left(X \in B\right) = P_0\left(c_B\left(X\right) \leq 1/\lambda^*\right).$$

Conclusions

- Presented systematic approach for quantifying model misspecification.
- Approach based on optimal transport

$$\sup_{D(P,P_0)\leq\delta} P\left(X\in B\right) = P_0\left(c_B\left(X\right)\leq 1/\lambda^*\right).$$

- Closed forms solutions, tractable in terms of P_0, feasible calibration of δ.

Conclusions

- Presented systematic approach for quantifying model misspecification.
- Approach based on optimal transport

$$\sup_{D(P,P_0)\leq\delta} P\left(X\in B\right)=P_0\left(c_B\left(X\right)\leq 1/\lambda^*\right).$$

- Closed forms solutions, tractable in terms of P_0, feasible calibration of δ.
- New statistical estimators, connections to machine learning & regularization.

Conclusions

- Presented systematic approach for quantifying model misspecification.
- Approach based on optimal transport

$$\sup_{D(P,P_0)\leq\delta} P\left(X\in B\right) = P_0\left(c_B\left(X\right)\leq 1/\lambda^*\right).$$

- Closed forms solutions, tractable in terms of P_0, feasible calibration of δ.
- New statistical estimators, connections to machine learning & regularization.
- Extensions of Empirical Likelihood & connections to optimal regularization.