ANSRE: ANalysis and Synthesis of Rare Events

Jose Blanchet, Karthyek Murthy, Viet-Anh Nguyen, Fan Zhang

We acknowledge support for the MURI sponsored by the Air Force under award number FA9550-20-1-0397.

Goal:

This talk has three objectives to discuss:
1) The role of robustness and the impact of model error?
2) The proposed approch that we'll follow.
3) Challenges and opportunities.

The Role of Robustness and Impact of Model Error

"All models are wrong, but some are useful" G. Box (1976) JASA.

At the hearth of G. Box's discussion is the trade-off between fidelity and tractability.

This observation is applicable to *every* model,
in extreme events we are particularly exposed by data scarcity.

The Role of Robustness in Rare Event Analysis...

- **Example:** Extreme Value Theory (EVT)

$$M_n = \max\{X_1, \ldots, X_n\} \approx a_n Z + b_n, \quad (1)$$

where a_n, b_n are deterministic sequences and Z is (known!) generalized extreme value distribution.

The Role of Robustness in Rare Event Analysis...

- **Example:** Extreme Value Theory (EVT)

$$M_n = \max\{X_1, ..., X_n\} \approx a_n Z + b_n, \tag{1}$$

  where a_n, b_n are deterministic sequences and Z is (known!) generalized extreme value distribution.

- EVT is used to extrapolate quantiles since a_n and b_n can be estimated by maximum likelihood and Z has a parametric form.

The Role of Robustness in Rare Event Analysis...

- **Example:** Extreme Value Theory (EVT)

$$M_n = \max\{X_1, ..., X_n\} \approx a_n Z + b_n, \tag{1}$$

  where a_n, b_n are deterministic sequences and Z is (known!) generalized extreme value distribution.
- EVT is used to extrapolate quantiles since a_n and b_n can be estimated by maximum likelihood and Z has a parametric form.
- But (1) involves assumptions that are impossible to verify and/or depending on fine (hard to learn) structure in the distribution of X_i.

The Role of Robustness in Rare Event Analysis...

- **Example:** Extreme Value Theory (EVT)

$$M_n = \max\{X_1, ..., X_n\} \approx a_n Z + b_n, \quad (1)$$

  where a_n, b_n are deterministic sequences and Z is (known!) generalized extreme value distribution.
- EVT is used to extrapolate quantiles since a_n and b_n can be estimated by maximum likelihood and Z has a parametric form.
- But (1) involves assumptions that are impossible to verify and/or depending on fine (hard to learn) structure in the distribution of X_i.
- Assumptions sometimes fail to hold (e.g. geometric or Poisson data).

The Role of Robustness in Rare Event Analysis...

- **Example:** Extreme Value Theory (EVT)

$$M_n = \max\{X_1, ..., X_n\} \approx a_n Z + b_n, \tag{1}$$

  where a_n, b_n are deterministic sequences and Z is (known!) generalized extreme value distribution.
- EVT is used to extrapolate quantiles since a_n and b_n can be estimated by maximum likelihood and Z has a parametric form.
- But (1) involves assumptions that are impossible to verify and/or depending on fine (hard to learn) structure in the distribution of X_i.
- Assumptions sometimes fail to hold (e.g. geometric or Poisson data).
- **Consequence: unexpected failures.**

Simulated Data

Data $X = Y + 50I(Y > 5)$ where $P(Y > y) = (1 + y)^{-1.1}$ (i.e. mixtures) estimate $1/1000$ tail quantile of X. Simulated data 2,000 i.i.d. samples of X.

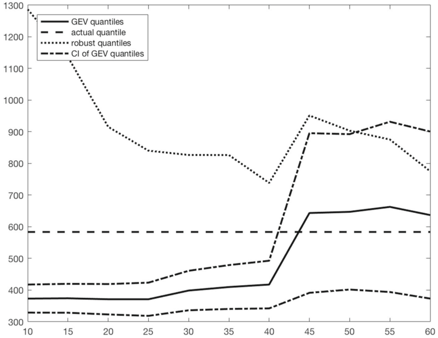


Figure: Block size vs $1/1000$ tail quantile.

B., He and Murthy (2020): https://doi-org/10.1007/s10687-019-00371-1

Need for Robustness: Academic Response Basel 3.5 (2014)

"In its broadest sense, robustness has to do with (in)sensitivity to underlying model deviations and/or data changes. *Furthermore, here, a whole new field of research is opening up; at the moment, it is difficult to point to the right approach*."

Robustness:
insensitivity of *decisions* / *inference* to model error or data changes
trying to minimize the cost of such insensitivity.

Elements of Distributionally Robust Analysis

- Distributionally Robust Performance Analysis

$$\min_{\theta} \max_{D(P,P_0)\leq\delta} E_P\left(f\left(X,\theta\right)\right)$$

Elements of Distributionally Robust Analysis

- Distributionally Robust Performance Analysis

$$\min_{\theta} \max_{D(P,P_0)\leq\delta} E_P\left(f\left(X,\theta\right)\right)$$

- $f(\cdot) <-$ quantity of interest (measures risk).

Elements of Distributionally Robust Analysis

- Distributionally Robust Performance Analysis

$$\min_{\theta} \max_{D(P,P_0)\leq\delta} E_P\left(f\left(X,\theta\right)\right)$$

- $f\left(\cdot\right) <-$ quantity of interest (measures risk).
- $X <-$ stochastic object of interest.

Elements of Distributionally Robust Analysis

- Distributionally Robust Performance Analysis

$$\min_{\theta} \max_{D(P,P_0)\leq\delta} E_P\left(f\left(X,\theta\right)\right)$$

- $f(\cdot)$ <– quantity of interest (measures risk).
- X <– stochastic object of interest.
- θ <– decision variable to minimize risk.

Elements of Distributionally Robust Analysis

- Distributionally Robust Performance Analysis

$$\min_{\theta} \max_{D(P,P_0)\leq\delta} E_P\left(f\left(X,\theta\right)\right)$$

- $f(\cdot) <-$ quantity of interest (measures risk).
- $X <-$ stochastic object of interest.
- $\theta <-$ decision variable to minimize risk.
- $P_0 <-$ benchmark model (wrong but simple, balances tractability & reality).

Elements of Distributionally Robust Analysis

- Distributionally Robust Performance Analysis

$$\min_{\theta} \max_{D(P,P_0)\leq\delta} E_P\left(f\left(X,\theta\right)\right)$$

- $f(\cdot)$ <– quantity of interest (measures risk).
- X <– stochastic object of interest.
- θ <– decision variable to minimize risk.
- P_0 <– benchmark model (wrong but simple, balances tractability & reality).
- $\mathcal{U}_\delta = \{P : D(P, P_0) \leq \delta\}$ <– *distributional uncertainty region*.

Literature Review

- **RO & Divergence-DRO:** Dupuis, James & Peterson '00; Hansen & Sargent '01, '08; Nilim & El Ghaoui '02, '03; Iyengar '05; A. Ben-Tal, L. El Ghaoui, & A. Nemirovski '09; Bertsimas & Sim '04; Bertsimas, Brown, Caramanis '13; Lim & Shanthikumar '04; Lam '13, '17; Csiszár & Breuer '13; Jiang & Guan '12; Hu & Hong '13; Wang, Glynn & Ye '14; Bayrakskan & Love '15; Duchi, Glynn & Namkoong '16; Bandi and Bertsimas '15; Bertsimas, Gupta & Kallus '13.

Literature Review

- **RO & Divergence-DRO:** Dupuis, James & Peterson '00; Hansen & Sargent '01, '08; Nilim & El Ghaoui '02, '03; Iyengar '05; A. Ben-Tal, L. El Ghaoui, & A. Nemirovski '09; Bertsimas & Sim '04; Bertsimas, Brown, Caramanis '13; Lim & Shanthikumar '04; Lam '13, '17; Csiszár & Breuer '13; Jiang & Guan '12; Hu & Hong '13; Wang, Glynn & Ye '14; Bayrakskan & Love '15; Duchi, Glynn & Namkoong '16; Bandi and Bertsimas '15; Bertsimas, Gupta & Kallus '13.
- **Wasserstein-DRO & Moments:** Scarf '58; Shapiro '15; Delage & Ye '10; Hampel '73; Huber '81; Pflug & Wozabal '07; Delage & Ye '10; Mehrotra & Zhang '14; Esfahani & Kuhn '15; Blanchet & Murthy '16; Gao & Kleywegt '16; Duchi & Namkoong '17.

Questions of Interest, Challenges and Opportunities

- How do you select the shape (i.e. $D(P, P_0)$) and the size δ of uncertainty region

$$\mathcal{U}_\delta = \{P : D(P, P_0) \leq \delta\}?$$

Questions of Interest, Challenges and Opportunities

- How do you select the shape (i.e. $D(P, P_0)$) and the size δ of uncertainty region

$$\mathcal{U}_\delta = \{P : D(P, P_0) \leq \delta\}?$$

- How do you solve

$$\max_{D_c(P,P_0)\leq\delta} E_P(f(Y))?$$

Questions of Interest, Challenges and Opportunities

- How do you select the shape (i.e. $D(P, P_0)$) and the size δ of uncertainty region

$$\mathcal{U}_\delta = \{P : D(P, P_0) \leq \delta\}?$$

- How do you solve

$$\max_{D_c(P,P_0)\leq\delta} E_P(f(Y))?$$

- How do you make inference if P_0 is data driven or non-parametric?

Questions of Interest, Challenges and Opportunities

- How do you select the shape (i.e. $D(P, P_0)$) and the size δ of uncertainty region

$$\mathcal{U}_\delta = \{P : D(P, P_0) \leq \delta\}?$$

- How do you solve

$$\max_{D_c(P,P_0)\leq\delta} E_P(f(Y))?$$

- How do you make inference if P_0 is data driven or non-parametric?
- Address these questions in the context of rare events.

Questions of Interest, Challenges and Opportunities

- How do you select the shape (i.e. $D(P, P_0)$) and the size δ of uncertainty region

$$\mathcal{U}_{\delta} = \{P : D(P, P_0) \leq \delta\}?$$

- How do you solve

$$\max_{D_c(P,P_0)\leq\delta} E_P(f(Y))?$$

- How do you make inference if P_0 is data driven or non-parametric?
- Address these questions in the context of rare events.
- Connect these findings to our thrusts.

Selecting the Uncertainty Set: A Simple Example

- How to choose $\{P : D(P, P_0) \leq \delta\}$?

Selecting the Uncertainty Set: A Simple Example

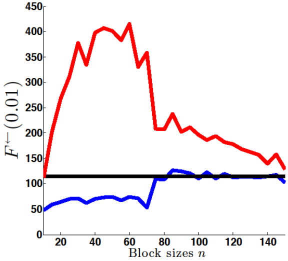
- How to choose $\{P : D(P, P_0) \leq \delta\}$?
- Of course, tractatibility is a concern...

Selecting the Uncertainty Set: A Simple Example

- How to choose $\{P : D(P, P_0) \leq \delta\}$?
- Of course, tractatibility is a concern...
- But also we'd like to understand what are the implications of the chosen "geometry"?

Selecting the Uncertainty Set: A Simple Example

- How to choose $\{P : D(P, P_0) \leq \delta\}$?
- Of course, tractatibility is a concern...
- But also we'd like to understand what are the implications of the chosen "geometry"?
- For example, let's take the simple EVT discussed earlier.



How to Select the Uncertainty Set?

- Naturally, P_0 is dictated by EVT

$$M_n = \max\{X_1, ..., X_n\} \approx b_n Z(\gamma) + a_n, \tag{2}$$

with $Z(\gamma)$ being Weibull: $\gamma < \mathbf{0}$ or Gumbel: $\gamma = \mathbf{0}$ or Frechet: $\gamma > \mathbf{0}$.

How to Select the Uncertainty Set?

- Naturally, P_0 is dictated by EVT

$$M_n = \max\{X_1, ..., X_n\} \approx b_n Z(\gamma) + a_n, \tag{2}$$

with $Z(\gamma)$ being Weibull: $\gamma < \mathbf{0}$ or Gumbel: $\gamma = \mathbf{0}$ or Frechet: $\gamma > \mathbf{0}$.

- $Z(\gamma) = GEV(\gamma)$ includes Weibull, Gumbel, Frechet:

$$P(Z \leq x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right) \quad 1 + \gamma x > 0.$$

How to Select the Uncertainty Set?

- Naturally, P_0 is dictated by EVT

$$M_n = \max\{X_1, ..., X_n\} \approx b_n Z(\gamma) + a_n, \qquad (2)$$

with $Z(\gamma)$ being Weibull: $\gamma < \mathbf{0}$ or Gumbel: $\gamma = \mathbf{0}$ or Frechet: $\gamma > \mathbf{0}$.

- $Z(\gamma) = GEV(\gamma)$ includes Weibull, Gumbel, Frechet:

$$P(Z \leq x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right) \quad 1 + \gamma x > 0.$$

- **The larger the** γ **the heavier the tails <– cases** $\gamma < 0, \gamma = 0, \gamma > 0$: **key for intuition in decision making!**

How to Select the Uncertainty Set?

- Naturally, P_0 is dictated by EVT

$$M_n = \max\{X_1, ..., X_n\} \approx b_n Z(\gamma) + a_n, \tag{2}$$

with $Z(\gamma)$ being Weibull: $\gamma < \mathbf{0}$ or Gumbel: $\gamma = \mathbf{0}$ or Frechet: $\gamma > \mathbf{0}$.

- $Z(\gamma) = GEV(\gamma)$ includes Weibull, Gumbel, Frechet:

$$P(Z \leq x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right) \quad 1 + \gamma x > 0.$$

- **The larger the γ the heavier the tails <– cases** $\gamma < 0, \gamma = 0, \gamma > 0$: **key for intuition in decision making!**
- If (2) holds, then X belongs to the domain of attraction of $GEV(\gamma)$.

Consider Divergence Criteria

- We consider (with P_0 = reference model from standard EVT)

$$\bar{F}_{\alpha}^{*}(u) := \max_{P: D_{\alpha}(P||P_0) \leq \delta} P(M_n > u).$$

Consider Divergence Criteria

- We consider (with P_0 = reference model from standard EVT)

$$\bar{F}_{\alpha}^{*}(u) := \max_{P:D_{\alpha}(P||P_0)\leq\delta} P(M_n > u).$$

- Renyi divergence of degree $\alpha > 1$

$$D_{\alpha}(P||P_0) = \frac{1}{\alpha - 1} \log E_0\left(\left(\frac{dP}{dP_0}\right)^{\alpha}\right).$$

Consider Divergence Criteria

- We consider (with P_0 = reference model from standard EVT)

$$\bar{F}^*_\alpha(u) := \max_{P:D_\alpha(P||P_0)\leq\delta} P(M_n > u).$$

- Renyi divergence of degree $\alpha > 1$

$$D_\alpha(P||P_0) = \frac{1}{\alpha-1}\log E_0\left(\left(\frac{dP}{dP_0}\right)^\alpha\right).$$

- As $\alpha \to 1$ get Kullback-Leibler (KL) divergence

$$D_\alpha(P||P_0) \to D_1(P||P_0) = E_0\left(\frac{dP}{dP_0}\log\left(\frac{dP}{dP_0}\right)\right).$$

Using Renyi Divergence & KL in EVT

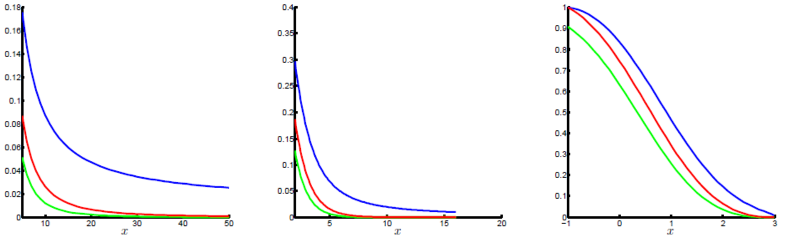
Theorem (B., He, and Murthy (2020))

$\bar{F}^*_\alpha(u)$ preserves domain of attraction if $\alpha > 1$ and $\alpha = 1$ substantially increases the risk estimate:

Reference model using EVT methodology	Robustified tail distribution using Renyi (α): Domain of attraction	Robustified tail distribution using KL: Domain of attraction
$\gamma < 0$ Bounded support Weibull	$\gamma^* = \gamma\,\alpha / (\alpha-1)$ STILL WEIBULL	Slow decay to upper bound of support
$\gamma = 0$ Exponential-types Gumbel	$\gamma^* = 0$ STILL GUMBEL	Frechet → Heavy
$\gamma > 0$ Power law type tails Frechet	$\gamma^* = \gamma\,\alpha / (\alpha-1)$ STILL FRECHET	Logarithmic tail decay → Super Heavy

Illustration of the Kullblack-Leibler Selection vs Renyi

- **Tail CDFs $\bar{F}_{\alpha}(u)$:** *Blue = KL*, **Red = Renyi(5)**, Green = True.



(a) $G_{\frac{1}{3}}(x)$, a Frechet example (b) $G_0(x)$, a Gumbel example (c) $G_{-\frac{1}{3}}(x)$, a Weibull example

Optimal Transport Uncertainty Sets

- Also consider $\{P : D(P_0, P) \leq \delta\}$ using optimal transport:

$$
\begin{aligned}
D_c(P_0, P) &= \min\{\int c(x, y)\, \pi(dx, dy) \\
\text{s.t.} \int_y \pi(dx, dy) &= P_0(dx) \\
\int_x \pi(dx, dy) &= P(dy) \\
\pi(dx, dy) &\geq 0\}.
\end{aligned}
$$

Optimal Transport Uncertainty Sets

- Also consider $\{P : D(P_0, P) \leq \delta\}$ using optimal transport:

$$\begin{aligned} D_c(P_0, P) &= \min\{\int c(x,y)\,\pi(dx,dy) \\ \text{s.t.} \int_y \pi(dx,dy) &= P_0(dx) \\ \int_x \pi(dx,dy) &= P(dy) \\ \pi(dx,dy) &\geq 0\}. \end{aligned}$$

- The Wasserstein distance $W(\mu, v)$ is obtained by choosing $c(x,y)$ to be a metric.

Optimal Transport Uncertainty Sets

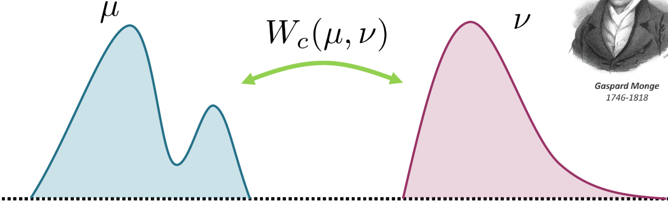
- Also consider $\{P : D(P_0, P) \leq \delta\}$ using optimal transport:

$$
\begin{aligned}
D_c(P_0, P) &= \min\{\int c(x,y)\,\pi(dx,dy) \\
\text{s.t.} \int_y \pi(dx,dy) &= P_0(dx) \\
\int_x \pi(dx,dy) &= P(dy) \\
\pi(dx,dy) &\geq 0\}.
\end{aligned}
$$

- The Wasserstein distance $W(\mu, v)$ is obtained by choosing $c(x,y)$ to be a metric.
- Based on (infinite dimensional) linear programming (tractable in principle).

Optimal Transport: Wasserstein Distance

Leonid Kantorovich
1912–1986



Gaspard Monge
1746-1818

$$W_c(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} {\color{red} d^p(x,y)}\,\pi(\mathrm{d}x,\mathrm{d}y)$$

$$\Pi(\mu,\nu) = \{\pi | \pi(A\times\mathcal{X}) = \mu(A), \pi(\mathcal{X}\times\mathcal{B}) = \nu(B)\ \forall A\subset\mathcal{X}, B\subset\mathcal{X}\}$$

Robust Analysis with Optimal Transport Costs

Theorem (B. and Murthy (2019))

If $c \geq 0$ is lower semicontinuous & $E_{P_0} |f(X)| < \infty$,

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

Moreover, π_* and λ_* can be characterized using complementary slackness.

Math. of Operations Research (2019):
https://doi-org/10.1287/moor.2018.0936
The key outcome is that strong duality holds and it reduces to a one dimensional convex problem.

Far reaching implications!

Unification and Extensions of Regularized Estimators

- *Support Vector Machines*: B., Kang, Murthy '19 - *https://doi-org/10.1287/moor.2018.0936*

Unification and Extensions of Regularized Estimators

- *Support Vector Machines*: B., Kang, Murthy '19 - *https://doi-org/10.1287/moor.2018.0936*
- *Group Lasso*: B., & Kang '16: **https://arxiv.org/abs/1705.04241 (ACML '17)**

Unification and Extensions of Regularized Estimators

- *Support Vector Machines*: B., Kang, Murthy '19 - *https://doi-org/10.1287/moor.2018.0936*
- *Group Lasso*: B., & Kang '16: **https://arxiv.org/abs/1705.04241 (ACML '17)**
- *Generalized adaptive ridge*: B., Kang, Murthy, Zhang '17: **https://arxiv.org/abs/1705.07152**

Unification and Extensions of Regularized Estimators

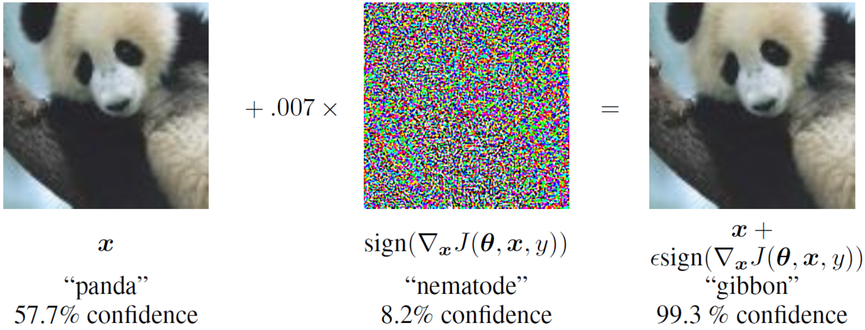
- *Support Vector Machines*: B., Kang, Murthy '19 - *https://doi-org/10.1287/moor.2018.0936*
- *Group Lasso*: B., & Kang '16: **https://arxiv.org/abs/1705.04241 (ACML '17)**
- *Generalized adaptive ridge*: B., Kang, Murthy, Zhang '17: **https://arxiv.org/abs/1705.07152**
- Semisupervised learning: B., and Kang '20: **https://arxiv.org/abs/1702.08848 (OR '20)**

Unification and Extensions of Regularized Estimators

- *Support Vector Machines*: B., Kang, Murthy '19 - *https://doi-org/10.1287/moor.2018.0936*
- *Group Lasso*: B., & Kang '16: **https://arxiv.org/abs/1705.04241 (ACML '17)**
- *Generalized adaptive ridge*: B., Kang, Murthy, Zhang '17: **https://arxiv.org/abs/1705.07152**
- Semisupervised learning: B., and Kang '20: **https://arxiv.org/abs/1702.08848 (OR '20)**
- **Comprehensive review:** Rahimian and Mehrotra '19: **https://arxiv.org/pdf/1908.05659.pdf.**

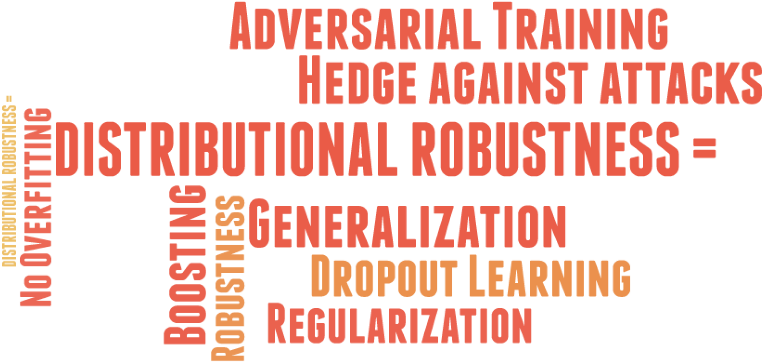
Deep Neural Networks: Adversarial Attacks

- Szegedy, Zaremba, Sutskever, Bruna, Erhan, Goodfellow, and Fergus (2014).



Distributional Robustness and Generalization in Machine Learning

Previous result can be used to recover: norm regularization, adversarial training of neural networks, support vector machines, LASSO, etc. (see *https://doi-org/10.1287/moor.2018.0936).*



Implications for Risk Analysis

Corollary

Let B ANY closed set

$$\begin{aligned} c_B(x) &= \inf\{c(x,y) : y \in B\} \\ &= \textit{Optimal cost of transporting } x \textit{ to } B. \end{aligned}$$

then

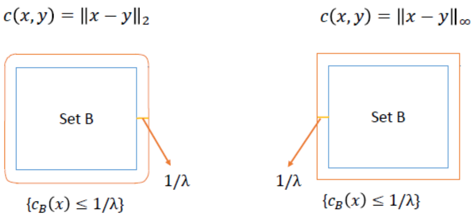
$$\sup_{D_c(P_0,P)\leq\delta} P(Y \in B) = P_0\left(c_B(X) \leq 1/\lambda^*\right),$$

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

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

- **Can choose $c_B(\cdot)$ so that calculation remains tractable...**

Geometry of the set $\{c_B(x) \leq 1/\lambda\}$



Additional Applications: Multidimensional Ruin Problems

- **Important:** $\{c_B(x) \leq 1/\lambda^*\}$ can preserve the geometry of B & original process is preserved!
- Brownian motion, can be robustified & estimate rare events involving heavy-tails.
- Estimate $P_{true}(\max_{0\leq t\leq 1} \text{InsuranceReserve}(t) > b)$ (assume InsuranceReserve(t) is Brownian motion under $P_0(\cdot)$)
- Simulated data is heavy tailed ($P(V > t) = 1/(1+t)^{2.2}$).
- Optimal transport cost: $c(x,y) = \max_{0\leq t\leq t} |x(t) - y(t)|^2$

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

Discussion

Distributionally robust engineering design in material design
$\max_{D(P,P_0)\leq\delta} P\left(\sup_{x\in\Omega}\|Du(x,\omega)\|>b\right);$
$u(\cdot)$ solves a PDE satisfying with random input.

(Zhigang+Vahid).

Discussion

Distributionally Robust Bayesian Inference

Example:

$$\min_{\phi(\cdot)} E \left\| \Theta - \phi\left(X\right) \right\|_2^2 \text{ -> solution } \phi^*\left(X\right) = E\left(\Theta | X\right).$$

If (X, Θ) is Gaussian under P_0, $\phi^*(\cdot)$ is affine.

Now, we have that Nguyen et al. (2020)

$$\min_{\phi(\cdot)} \max_{W_2(P,P_0)\leq\delta} E\left\|\Theta - \phi\left(X\right)\right\|_2^2$$

Nash eq. exists, ϕ^* is also affine.

Question: Inform uncertainty quantification with the Wasserstein distance for a large class of PDEs with random input and also for max-stable processes.

Conclusions

- Distributional robustness can be achieved by optimizing over a non-parametric set of models (called uncertainty region).

Conclusions

- Distributional robustness can be achieved by optimizing over a non-parametric set of models (called uncertainty region).
- The shape of the uncertainty region induces key properties in the inference (Renyi-type divergence preserves domain of attraction).

Conclusions

- Distributional robustness can be achieved by optimizing over a non-parametric set of models (called uncertainty region).
- The shape of the uncertainty region induces key properties in the inference (Renyi-type divergence preserves domain of attraction).
- Optimal transport recovers regularization, adversarial training and has good generalization properties.

Conclusions

- Distributional robustness can be achieved by optimizing over a non-parametric set of models (called uncertainty region).
- The shape of the uncertainty region induces key properties in the inference (Renyi-type divergence preserves domain of attraction).
- Optimal transport recovers regularization, adversarial training and has good generalization properties.
- Optimal transport is flexible, can be used in robustifying rare events for random fields, stochastic processes, random graphs, etc.

Conclusions

- Distributional robustness can be achieved by optimizing over a non-parametric set of models (called uncertainty region).
- The shape of the uncertainty region induces key properties in the inference (Renyi-type divergence preserves domain of attraction).
- Optimal transport recovers regularization, adversarial training and has good generalization properties.
- Optimal transport is flexible, can be used in robustifying rare events for random fields, stochastic processes, random graphs, etc.
- Key computation involves and linear programming in the context of optimal transport.