Quantile Sensitivity Estimation

Bernd Heidergott and Warren Volk-Makarewicz

Department of Econometrics and Operations Research
Vrije Universiteit Amsterdam

NET-COOP, 23rd. November 2009

vrije Universiteit amsterdam

Problem in Brief

- Let $X_\theta \in \mathbb{R}^d$, with distribution function $F_\theta(x)$, parameterised by $\theta = [a, b] \subset \mathbb{R}$.
- Map $g : \mathbb{R}^d \to \mathbb{R}$, Discrete event system (DES) $g_\theta(X) := g(X_\theta)$.
- Definition: Quantile (left tail), level $\alpha \in (0, 1)$

$$q_{\alpha,\theta}(g(X)) = \sup\{y : F_{g,\theta}(y) \leq \alpha\} \tag{1}$$

- Estimate quantile via the order statistic: Sample size n. $\nu = \lceil \alpha n \rceil$
- $[g(X)]_{\nu:n}$, ν^{th} smallest realisation. E.g:

$$[g(X)]_{(1)} \leq [g(X)]_{(2)} \leq \dots \leq [g(X)]_{(n)}$$

- From theory: $\mathbb{E}[g(X)]_{\nu:n} \approx q_{\alpha,\theta}(g(X))$, for large n.
- What happens when we change the parameter θ ?
- This leads to the sensitivity of the quantile with respect to θ

$$\frac{\partial}{\partial\theta} q_{\alpha,\theta}(g(X)) = \frac{\partial}{\partial\theta}\mathbb{E}[g(X)]_{\nu:n}$$

This talk will help answer this question. We use the Weak Differentiation approach.

Literature Review: Quantiles (Key References)

History

- Bahadur (1966): Order statistic, iid data. Strongly consistent and asymptotically normal.
- Sen (1972), Heidelburger & Lewis (1985): Dependent sequences.
- David (1980), Serfling (1980): Main texts in area.
- Google: 320,000 results for Quantile Estimation.

Literature Review: Quantiles (Key References)

Applications

- Quality of Service: Ensure minimal service requirement for non-preferred customers.
- Value at Risk: Financial solvency.
- Health Care: Ensure minimal requirement for arrival times of ambulances.

Literature Review: Sensitivity Analysis

1. Finite Differencing
   - (Forward) Estimate of 'slope' of $(g_{\theta+\Delta}(X) - g_{\theta}(X))/\Delta$.
   - Δ too large, concern with bias. Δ too small, variance of estimator is too prohibitive.
   - Variance reduction techniques and stratified sampling does not help (Jäckel, 2002; Zazanis & Suri, 1993).
2. Stochastic Sensitivity Analysis.
   - IPA: Infinitesimal Perturbation Analysis (Glasserman, 1991).
     - Sample path approach. Requires differentiation of g.
   - SF: Score Function Method (Rubinstein & Shapiro, 1993).
   - WD: Weak Differentiation (Pflug, 1996; Leahu, 2008).
     - Distribution approach. Probability distribution is differentiated.

Stochastic Sensitivity Analysis

IPA Estimator

- Probability Space $(\Omega = \mathbb{R}, \mathcal{F} = \sigma(X), \mathbb{P})$.
- $X_\theta \in \mathbb{R}^d$. Let $g'(\cdot)$ be the derivative of g. This leads to the appropriate condition.

$$\begin{aligned}\frac{\partial}{\partial\theta}\mathbb{E}_{\mathbb{P}}[g(X_\theta)] &= \mathbb{E}_{\mathbb{P}}\left[\frac{\partial}{\partial\theta}g(X_\theta)\right] \\ &= \mathbb{E}\left[g'(x)\Big|_{x=X_\theta}\frac{\partial}{\partial\theta}X_\theta\right]\end{aligned}$$

Stochastic Sensitivity Analysis

SF Estimator

- Probability Space $(\Omega = \mathbb{R}, \mathcal{F} = \sigma(X), \mathbb{P}_\theta)$.
- $X \subset \mathbb{R}^d$, probability measure $\mathbb{P}_\theta$. Must require $\mathbb{P}_\theta$ to be dominated by another probability measure $\mathbb{Q}$. $\theta = [a, b] \subset \mathbb{R}$.
- Map $g : \mathbb{R}^d \to \mathbb{R}$, $g \in L^2(\mathbb{P}_\theta \cap \mathbb{P}'_\theta)$.

$$\begin{aligned}\frac{\partial}{\partial\theta}\mathbb{E}_{\mathbb{P}_\theta}[g(X)] &= \mathbb{E}_{\mathbb{P}'_\theta}[g(X)] \\ &= \mathbb{E}_{\mathbb{Q}}\left[g(X)\frac{d\mathbb{P}_\theta}{d\mathbb{Q}}\right] \qquad (1)\end{aligned}$$

$\frac{d\mathbb{P}_\theta}{d\mathbb{Q}}$ is the Radon-Nikodym derivative.

Stochastic Sensitivity Analysis

WD Estimator

- Probability Space $(\Omega = \mathbb{R}, \mathcal{F} = \sigma(X), \mathbb{P}_\theta)$.
- $X \subset \mathbb{R}^d$, probability measure $\mathbb{P}_\theta$. Derivative, $\mathbb{P}'_\theta$ for $\theta = [a, b] \subset \mathbb{R}$ is a signed measure which can be written as difference of two probability measures $c(\theta)(\mathbb{P}^+_\theta - \mathbb{P}^-_\theta)$.
- The measures $\mathbb{P}^+_\theta$ and $\mathbb{P}^-_\theta$ in the weak derivative of $\mathbb{P}_\theta$ have to satisfy the following: $\forall h \in C_b$

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\mathbb{E}_{\mathbb{P}_\theta}[h(X)] &= \mathbb{E}_{\mathbb{P}'_\theta}[h(X)] \\
&= c(\theta)\left(\mathbb{E}_{\mathbb{P}^+_\theta}[h(X)] - \mathbb{E}_{\mathbb{P}^-_\theta}[h(X)]\right) \\
&= c(\theta)\left(\mathbb{E}[h(X^+)] - \mathbb{E}[h(X^-)]\right) \qquad (1)
\end{aligned}
$$

Quantile Sensitivity Analysis: IPA

- Hong, 2009. Hong & Liu, (2 papers).
- $\mathbb{P}_\theta = f(x)Leb(dx)$. DES $g(X)$ with assumption of g being Lipschitz continuous.
- Define $a = q_{\alpha,\theta}(g(X))$.
- IPA Sensitivity of Quantile.

$$q'_\alpha(\theta) = \mathbb{E}_{\mathbb{P}}\left[\frac{\partial}{\partial\theta} g(X_\theta) \middle| g(X_\theta) = a\right]$$

Quantile Sensitivity Analysis: IPA

- IPA Quantile Sensitivity Estimator
  Suppose we have n iid rv's $X = (X_i)_{i=1}^{n}$. Denote $n = mk$, and divide X into k equal samples of m elements. Defining $\nu = \lceil \alpha m \rceil$, with $\alpha \in (0,1)$, the estimator is of the form

$$\bar{D}_{mk} = \frac{1}{k} \sum_{j=1}^{k} D_{m,j} \tag{2}$$

  where $D_{m,j}$ is the limit of the (forward) finite-difference sensitivity estimate of the order statistic.

Quantile Sensitivity Analysis: IPA

Result has been subsequently expanded to:

- X has ϕ-mixing dependency.
- Evaluating sensitivities of CVaR (expected loss given an event with probability α occurred).
- Quantile sensitivity using kernel estimation.

Variance of Sensitivities in the exponential case $(X \sim Exp(\theta))$

- *WD*: Weak Differentiation, standard approach: $X^+ = \exp(\theta), X^+ = X, X^- \sim \gamma(2, \theta)$ independent from X^+.
- WD_{Corr}: WD, standard approach with correlation: $X^+ = X, X^- = X + Y$, Y independent $Exp(\theta)$ RV from X.
- WD_{HJ}: WD, with Hahn - Jordan Decomposition probability measures.
- *IPA*: Infinitesimal Perturbation Analysis.

Lemma

Given the above distribution and sensitivity estimation approaches with regards to the performance function $g(x) = x^p$, $p \in \mathbb{N}$.

$$\begin{array}{ll} p = 1 & \mathrm{Var}_{WD_{HJ}} < \mathrm{Var}_{WD_{Corr}} = \mathrm{Var}_{IPA} < \mathrm{Var}_{WD} \\ p \geq 2 & \mathrm{Var}_{WD_{HJ}} < \mathrm{Var}_{WD_{Corr}} < \mathrm{Var}_{WD} < \mathrm{Var}_{IPA} \end{array}$$

Comparison between IPA and WD

- SF will not be considered. Unbiased, but in general variance is prohibitively high.

IPA	WD
Unbiased	Unbiased
Estimation is simple	Estimation is involved
Quick to compute	Can be quick to compute
Low variance	Usually outperforms IPA

Note: Behaviour of WD estimator depends on choice of probability measures.

Comparison between IPA and WD

IPA	WD
F continuous	F_θ continuous and discrete
g differentiable (no jumps)	g is merely measurable
Need to know behaviour of RV	Need to know $\mathbb{P}_\theta$ and how to simulate $X^{\pm}$.

Quantile Sensitivity via WD

- Simulation budget n of iid random variables $X = (X_i)_{i=1}^n$, $X_i \subset \mathbb{R}$.
- X_i has probability measure $\mathbb{P}_\theta$
- k samples with m elements in each sample $n = km$.
- Also depict α-level quantile as order statistic.
- $q_{\alpha,\theta}(X) = \mathbb{E}_{\mathbb{P}_\theta^m}[X_{\nu:m}]$. $\nu = \lceil \alpha m \rceil$.

Choose to write order statistic as a probability weighted mean (different than in paper)

$$\mathbb{E}_{\mathbb{P}_\theta^m}[X_{\nu:m}] = \frac{1}{\beta(\nu, m-\nu+1)} \int_{\mathbb{R}} F_\theta^{\nu-1}(x)(1-F_\theta(x))^{m-\nu}\mathbb{P}_\theta(dx)$$

$\beta(\cdot,\cdot)$ is the beta function.

WD Quantile Sensitivity Estimator

- Weak Derivative estimator is a Monte Carlo analog of the derivative of the integrand.
- By the product rule, the derivative of the integrand of the probability weighted mean expression is a sum of two terms:

$$\frac{\partial}{\partial\theta}\mathbb{E}_{\mathbb{P}_\theta^m}[X_{\nu:m}] = \frac{1}{\beta(\nu, m-\nu+1)}\frac{\partial}{\partial\theta}\int_{\mathbb{R}} F_\theta^{\nu-1}(x)(1-F_\theta(x))^{m-\nu}\mathbb{P}_\theta(dx)$$

WD Quantile Sensitivity Estimator

- Weak Derivative estimator is a Monte Carlo analog of the derivative of the integrand.
- By the product rule, the derivative of the integrand of the probability weighted mean expression is a sum of two terms:

$$\frac{\partial}{\partial\theta}\mathbb{E}_{\mathbb{P}_\theta^m}[X_{\nu:m}] = \frac{1}{\beta(\nu, m-\nu+1)}\int_{\mathbb{R}}\frac{\partial}{\partial\theta}\left\{F_\theta^{\nu-1}(x)(1-F_\theta(x))^{m-\nu}\right\}\mathbb{P}_\theta(dx)$$

- Highlighted aspect: Modified weighted mean w.r.t original random variable X. Differentiation of F_θ is an MVD derivative yielding $c(\theta)(F_\theta^+(x) - F_\theta^-(x))$.

WD Quantile Sensitivity Estimator

- Weak Derivative estimator is a Monte Carlo analog of the derivative of the integrand.
- By the product rule, the derivative of the integrand of the probability weighted mean expression is a sum of two terms:

$$\frac{\partial}{\partial\theta}\mathbb{E}_{\mathbb{P}_\theta^m}[X_{\nu:m}] = \frac{1}{\beta(\nu, m-\nu+1)}\int_{\mathbb{R}} F_\theta^{\nu-1}(x)(1-F_\theta(x))^{m-\nu}\frac{\partial}{\partial\theta}\mathbb{P}_\theta(dx)$$

- Highlighted aspect: Weak derivative. This is a difference of probability weighted means of X^+, X^- respectively.

WD Quantile Sensitivity Estimator

- $X = (X_i)_{i=1}^n$ is an iid collection of random variables. Denote $n = mk$, and divide X into k equal samples of m elements. Define $\nu = \lceil \alpha m \rceil$, $\alpha \in (0, 1)$.
- Define $D_{\nu:m,j}$, $j = 1, \ldots, k$ as the WD quantile sensitivity estimator of the j^{th} sample.
- Overall estimator is the average of the k samples.

$$D_{\nu:m} = \frac{1}{k} \sum_{j=1}^{k} D_{\nu:m,j}$$

Example: Parameter Quantile Sensitivity of Exponential distribution

- $\mathbb{P}_\theta = Exp(\theta)$.
- Fixed $n = 16384$. Various choices of m, k.
- $\alpha = 0.1, 0.5, 0.9$. $\theta = 1$. 1000 replications.
- WD Simulation approach: For a given sample $\Xi_l : 1 \leq l \leq k$ the ECDF is constructed using all the elements of Ξ_l (or $\Xi_l^{\pm}$) and the empirical probabilities are calculated with the same and corresponding samples. i.e $\hat{F}_\theta(x_i)$ is a multiple of $1/m$.
- Used both 'Correlated' and Hahn-Jordan random number generation.
- Compare with IPA.

Example Result

- $m = 128$, $k = 128$. $\alpha = 0.90$.

	Mean	Bias	Std. Err.
Corr	-2.4800	0.1774	0.0615
HJ	-2.3907	0.0881	0.0558
IPA	-2.3310	0.0284	0.0241

- True Value: $q'_{\alpha,\theta}(X) = \ln(0.1) = -2.3026$.

Analysis of Results

- Perspective: IPA Estimator: $D^{IPA}_{\nu:m} = -X_{\nu:m}/\theta$.

- In previous simulation, five variations were studied.
- For all five variants, for fixed n, m is large, estimator $D_{\nu:m}$ converges to $q'_{\alpha,\theta}(X)$. For k large, standard deviation is reduced.
- Computational time for the ECDF method (quickest variation) is ten times slower than IPA. Thus work-normalised variance is less for IPA.
- Outside the region of one standard error, *HJ* approach yields more unbiased estimates, controlling for everything else.
- For fixed m as k increases, rate of convergence for all approaches is $\mathcal{O}(k^{-1/2})$. This also holds for IPA. If m increases for fixed k, rate of convergence for WD variants is less then IPA $= \mathcal{O}(m^{-1/2})$.

Topics for Further Research

1. Conceptual Issues
   - Good decision rules between trade off of bias and standard error?
   - Alternative approaches in determining $\hat{F}_\theta$.
2. More General Models
   - M/M/1 Queue and Waiting times.
   - Other distributions and other performance functions.
   - Discrete distributions.

References

- Bahadur, 1966, A note of quantiles in large samples. *Ann. Math.Statist.*, 37, 577-580, 1966.
- Sen, 1972, On the Bahadur's representation of sample quantiles for sequences of ϕ-mixing random variables. *J. Multivariate Anal.*, 2, 77-95, 1972.
- David, 1980, *Order Statistics*. Wiley, New York.
- Serfling, 1980, *Approximation Theorems in Mathematical Statistics*. Wiley, New York.
- Heidelburger & Lewis, 1984, Quantile Estimation in dependent sequences. *Operations Research*, 32, 185-209.
- Jäckel, 2002, *Monte Carlo Methods in Finance*. Wiley. United Kingdom.
- Zazanis & Suri, 1993, Convergence rates of finite-difference sensitivities estimates for stochastic systems. *Operations Research*, Wiley, New York, 1980.

References

- Glasserman, 1991, Gradient Estimation via Peturbation Analysis, Kluwer Academic Publishers, Morwell, MA, 1991.
- Rubinstein & Shapiro, 1993, *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization*, Wiley, New York.
- Pflug, 1996, *Optimization of Stochastic Models*, Kluwer, Academic Publishers, Norwell, MA.
- Leahu, 2008, Measure-Valued Differentiation for Finite Products of Measures: Theory and Applications. Number 428 in Tinbergen Institute Research Series. Thela Thesis.
- Hong, 2009, Estimating Quantile Sensitivities, *Operations Research*, 57, 118-130.
- Hong & Liu, 2009, Simulating Sensitivities of Conditional Value-at-Risk, *Management Sci.*, 55, 281-293.
- Hong & Liu, 2009, Kernel Estimation of Quantile Sensitivities, *Naval Res. Logist.*, 56, 511-525.