Asymptotic statistical treatment rules when a quantile is the object of interest

Nikita Pavlov

Penn State University

April 2024

Motivation (1)

- Assume that depending on treatment $\tau \in \{0; 1\}$ outcome is distributed according to $F_\tau(\cdot|\theta_0)$, which belongs to distribution family uniquely determined by parameter $\theta_0 \in \Theta$;
- Let social welfare be given by functional $w_\tau(\theta_0) = W(F_\tau(\cdot|\theta_0))$. Define welfare contrast as $g(\theta_0) = w_1(\theta_0) - w_0(\theta_0)$
- Suppose that we have some data informative about θ_0 such that $g(\theta_0)$ is point-identified;
- In such circumstances, any sequence of treatment rules $\{\delta_n\}_{n=1}^{\infty}$ that is asymptotically equivalent to $I_{g(\theta_0)>0}$ is consistent (loss function $L(\delta_n, \hat{\theta}) \rightarrow_p 0$);
- How to choose among consistent alternatives?

Motivation (2)

- Hirano and Porter (2009) approach this from the perspective of hypothesis testing by using local parametrization and applying associated local asymptotic normality results (Le Cam (1986));
- In particular, they assume that $g(\theta_0) = 0$ and focus on a sequence of welfare contrasts $g(\theta_0 + \frac{h}{\sqrt{n}})$;
- For this sequence, they find sequence of rules that are asymptotically optimal in a sense of minimizing Bayes expected (wrt sample distribution) loss or minmax expected loss;

Motivation (3)

- However, as Manski and Tetenov (2014, 2023) point out, expectation is just one feature of loss distribution;
- Alternatively, one may be more interested in rules that control for risk in one of the tails;
- The purpose of my work is to find asymptotically optimal treatment rules in Hirano and Porter (2009) setting by focusing on quantiles instead of expectation.

Basic Setup (1)

Suppose that a social planner observes certain realization of random covariate vector $X \sim_d F_X$ defined on some space $\mathcal{X}$ and has to assign treatment $\tau \in \{0; 1\}$ according to chosen treatment rule

$$\delta(x) = P(\tau = 1 | X = x)$$

Treatment decision τ corresponds to outcome distribution $F_\tau(\cdot | x; \theta)$ which belongs to a distribution family indexed by parameter $\theta \in \Theta$.

Basic Setup (2)

Suppose that social welfare is given by functional $W(F)$, where F is some given distribution function. Let $w_\tau(\theta, x) = W(F_\tau(\cdot|x;\theta))$. Hence, welfare contrast is given by

$$g(\theta, x) = w_1(\theta, x) - w_0(\theta, x)$$

Assume that we have some i.i.d. data $Z = (Z_1, ..., Z_n) \sim_d P_\theta^n$ on product space $\mathcal{Z}^n$, which is informative about θ.

We focus on a sequence of experiments $\mathcal{E}_n = \{P_\theta^n : \theta \in \Theta\}$.

Basic Setup (3)

Randomised statistical treatment rule is a mapping $\mathcal{Z}^n \times \mathcal{X} \to [0, 1]$:

$$\delta(z^n, x) = Pr(\tau = 1 | Z^n = z^n, X = x)$$

Let $L(\delta, \theta, x)$ be loss function which determines penalties for using rule δ.

We want to choose δ to make distribution $P_\theta^n(L(\delta(z^n, x), \theta, x))$ most favorable according to some criterion.

Basic Setip (4)

One possible criterion is expected value of loss (Hirano and Porter, 2009), $E_{P_\theta^n}(L(\delta(z^n, x), \theta, x))$.

Instead, we may focus on λ-quantile of loss $Q^\lambda_{P_\theta^n}(L)$, λ-CVaR of loss $C^\lambda_{P_\theta^n}(L)$ or some other functional respecting stochastic dominance $G_{P_\theta^n}(L)$. This is the focus of my paper.

To create ordering of rules over chosen criterion, we may either use Bayesian approach (integrate chosen functional over parameter space Θ with respect to some prior measure Π) or focus on its minimax value.

Hirano and Porter (2009) Recap

Hirano and Porter (2009) loot at "problematic" θ's for which welfare contrast is close to 0.

In particular, assume that $g(\theta_0, x) = 0$ and consider sequence of local alternatives of the form $\theta_0 + \frac{h}{\sqrt{n}}$, $h \in R^k$.

By assuming $g(\theta_0, x) = 0$ we're focusing on cases when it is hard to choose treatment even for large sample sizes. Hirano and Porter (2009) propose treatment rule which is asymptotically optimal for these cases.

If $g(\theta_0, x) \neq 0$ there are many treatment rules (including those suggested by Hirano and Porter (2009)) which select correct treatment with probability approaching one.

Quantiles

Fix covariates and for simplicity assume symmetric hypothesis structure of loss function (similar logic applies to a more general form of loss):

$$L(\delta, \theta) = \begin{cases} 1 - \delta(z^n), g(\theta) > 0 \\ \delta(z^n), g(\theta) \leq 0 \end{cases} \tag{1}$$

Suppose that we order rules by integrating λ-quantile of loss function over state space with respect to some prior density $\pi(\theta)$.

That is, our object of interest could be presented as:

$$B_n(\delta, \theta_0, \pi) = \int Q^{\lambda}_{P^n_{\theta_0 + \frac{h}{\sqrt{n}}}} \left(L(\delta(z^n), \theta_0 + \frac{h}{\sqrt{n}}) \right) \pi(\theta_0 + \frac{h}{\sqrt{n}}) dh \tag{2}$$

Main Idea

- Find $B_{\infty}(\delta, \theta_0, \pi)$ such that $B_n(\delta, \theta_0, \pi) \rightarrow B_{\infty}(\delta, \theta_0, \pi)$;
- Find δ^* such that $B_{\infty}(\delta^*, \theta_0, \pi) = inf_{\delta \in \mathcal{D}} B_{\infty}(\delta, \theta_0, \pi)$;
- Find finite sample counterpart of δ^*, δ_n^*, such that $\delta_n^* \rightarrow_p \delta^*$;
- Show that plugging δ_n^* into finite sample Bayesian λ-quantile of loss B_n yields asymptotically optimal results: $B_n(\delta_n^*, \theta_0, \pi) \rightarrow inf_{\delta} B_{\infty}(\delta, \theta_0, \pi)$.

DQM Assumption

Assume $\theta_0 \in \Theta \subseteq R^k$ is such that $g(\theta_0, x) = 0$.

Assume that the sequence of experiments $\mathcal{E}_n = \{P_\theta^n : \theta \in \Theta\}$ satisfies Differentiability in Quadratic Mean (DQM) at θ_0 for some function $s : Z \to R^k$:

$$\int [\sqrt{dP^n_{\theta_0+h}(z)} - \sqrt{dP^n_{\theta_0}(z)} - \frac{1}{2} h' s(z) \sqrt{dP^n_{\theta_0}(z)}]^2 dz = ||o(h^2)||] \quad (3)$$

If (1) holds, usually s is a score function corresponding to $\mathcal{E}_1$ and it could be shown that fisher information $I_{\theta_0} = E_{P_{\theta_0}}[s's]$ exists.

Continuously differentiable density is usually sufficient for DQM to hold.

Main Proposition for Quantiles

Proposition below allows to obtain asymptotic optimality results for quantiles, similarly to Proposition 3.1 of Hirano and Porter (2009) for expected value. Both are based on Theorem 7.10 of van der Vaart (1998).

Proposition 1

Assume that Θ is an open subset or R^k. Let the sequence of experiments $(P^n_\theta : \theta \in \Theta)$ satisfy DQM at θ_0 with non-singular I_{θ_0}. Finally, suppose that treatment rule $\delta(z^n)$ in experiments $(P^n_{\theta_0+\frac{h}{\sqrt{n}}} : h \in R^k)$ converges in distribution under every h.
Than there exists a function $\delta(\Delta)$ in the limit experiment $\mathcal{N}(h, I_{\theta_0}^{-1})$ such that $Q^\lambda_{P^n_{\theta_0+\frac{h}{\sqrt{n}}}}(\delta(z^n)) \to Q^\lambda_{\mathcal{N}(h, I_{\theta_0}^{-1})}(\delta(\Delta))$ under every h.

Quantiles (Continued)

Recall that we're using symmetric loss structure:

$$L(\delta, \theta_0 + \frac{h}{\sqrt{n}}) = \begin{cases} 1 - \delta(z^n), g(\theta_0 + \frac{h}{\sqrt{n}}) > 0 \\ \delta(z^n), g(\theta_0 + \frac{h}{\sqrt{n}}) \leq 0 \end{cases} \tag{4}$$

Hence:

$$Q^{\lambda}_{P^n_{\theta_0 + \frac{h}{\sqrt{n}}}} \left(L(\delta, \theta_0 + \frac{h}{\sqrt{n}}) \right) = \begin{cases} Q^{\lambda}_{P^n_{\theta_0 + \frac{h}{\sqrt{n}}}} (1 - \delta), g(\theta_0 + \frac{h}{\sqrt{n}}) > 0 \\ Q^{\lambda}_{P^n_{\theta_0 + \frac{h}{\sqrt{n}}}} (\delta), g(\theta_0 + \frac{h}{\sqrt{n}}) \leq 0 \end{cases} \tag{5}$$

Quantiles (Continued)

Recall that we focus on Bayes λ-quantile of loss:

$$B_n(\delta, \theta_0, \pi) = \int Q^{\lambda}_{P^n_{\theta_0+\frac{h}{\sqrt{n}}}} \left(L(\delta(z^n), \theta_0 + \frac{h}{\sqrt{n}}) \right) \pi(\theta_0 + \frac{h}{\sqrt{n}}) dh =$$

$$\int_{g(\theta_0+\frac{h}{\sqrt{n}})\leq 0} Q^{\lambda}_{P^n_{\theta_0+\frac{h}{\sqrt{n}}}} (\delta) \pi(\theta_0+\frac{h}{\sqrt{n}}) dh + \int_{g(\theta_0+\frac{h}{\sqrt{n}})>0} Q^{\lambda}_{P^n_{\theta_0+\frac{h}{\sqrt{n}}}} (1-\delta) \pi(\theta_0+\frac{h}{\sqrt{n}}) dh$$

Quantiles (Continued)

Assume welfare contrast $g(\cdot)$ is differentiable at θ_0 and prior density $\pi(*)$ is continuous at θ_0.

Note that since $g(\theta_0) = 0$, $\sqrt{n}g(\theta_0 + \frac{h}{\sqrt{n}}) \to \dot{g}'h$. Hence, by assumptions above and Proposition 1:

$$B_n(\delta, \theta_0, \pi) \to B_\infty(\delta, \theta_0, \pi) =$$

$$\int\limits_{\dot{g}'h \leq 0} Q^{\lambda}_{\mathcal{N}(h, I^{-1}_{\theta_0})}(\delta(\Delta))\, \pi(\theta_0) dh + \int\limits_{\dot{g}'h > 0} Q^{\lambda}_{\mathcal{N}(h, I^{-1}_{\theta_0})}(1 - \delta(\Delta))\, \pi(\theta_0) dh \quad (6)$$

We are interested in finding $\delta(\Delta)$ minimizing (6).

Optimal Asymptotic Treatment Rule for Quantile Case

Proposition 2

Assume asymptotic Bayesian λ-quantile of loss $B_\infty(\delta, \theta_0, \pi)$ is given by (6). Than $\delta^*(\Delta) = I\{\frac{\dot{g}'\Delta}{\sigma_g} > \Phi^{-1}(\lambda)\}$, where $\sigma_g = \dot{g}' I_{\theta_0}^{-1} \dot{g}$, is an optimal asymptotic treatment rule:
$inf_{\delta \in \mathcal{D}} B_\infty(\delta, \theta_0, \pi) = B_\infty(\delta^*, \theta_0, \pi)$.

Optimal Asymptotic Treatment Rule for Quantile Case: Proof Sketch (1)

Note that $\delta^*(\Delta) = I\{\frac{\dot{g}'\Delta}{\sigma_g} > \Phi^{-1}(\lambda)\}$ is a Bernoulli random variable:

$$\delta^*(\Delta) = \begin{cases} 1 \text{ w.p. } P_{\mathcal{N}(h, I_{\theta_0}^{-1})}(\frac{\dot{g}'\Delta}{\sigma_g} > \Phi^{-1}(\lambda)) \\ 0 \text{ w.p. } P_{\mathcal{N}(h, I_{\theta_0}^{-1})}(\frac{\dot{g}'\Delta}{\sigma_g} \leq \Phi^{-1}(\lambda)) \end{cases}$$

Hence, assuming that if quantile is set-valued we take leftmost value of its set (these cases are of measure 0 w.r.t. Π):

$$Q^{\lambda}_{\mathcal{N}(h, I_{\theta_0}^{-1})}[\delta(\Delta)] = \begin{cases} 1 \text{ if } P_{\mathcal{N}(h, I_{\theta_0}^{-1})}(\frac{\dot{g}'\Delta}{\sigma_g} \leq \Phi^{-1}(\lambda)) < \lambda \\ 0 \text{ if } P_{\mathcal{N}(h, I_{\theta_0}^{-1})}(\frac{\dot{g}'\Delta}{\sigma_g} \leq \Phi^{-1}(\lambda)) \geq \lambda \end{cases} \tag{7}$$

Optimal Asymptotic Treatment Rule for Quantile Case: Proof Sketch (2)

Note that since $\Delta \sim \mathcal{N}(h, I_{\theta_0}^{-1})$,
$P_{\mathcal{N}(h, I_{\theta_0}^{-1})}(\frac{\dot{g}'\Delta}{\sigma_g} \leq \Phi^{-1}(\lambda)) = \Phi(\Phi^{-1}(\lambda) - \frac{\dot{g}'h}{\sigma_g}) \geq \lambda$ if $\dot{g}'h \leq 0$

Hence, $Q^{\lambda}_{\mathcal{N}(h, I_{\theta_0}^{-1})}(\delta(\Delta)) = 0$ for $\dot{g}'h \leq 0$.

Symmetrically, $Q^{\lambda}_{\mathcal{N}(h, I_{\theta_0}^{-1})}(1 - \delta(\Delta)) = 0$ for $\dot{g}'h > 0$.

As a result, quantile of loss is zero for every $\dot{g}'h$.

Example of Quantiles (Continued)

Now, focus on $\delta^*(\Delta) = I\{\frac{\dot{g}'\Delta}{\sigma_g} > \Phi^{-1}(\lambda)\}$.

Assume that we have a consistent estimator $\hat{\sigma}_g$ of σ_g and consistent estimator $\hat{\theta}_n$ of θ_0 which is best regular:

$$\sqrt{n}(\hat{\theta}_n - \theta_0 - \frac{h}{\sqrt{n}}) \to_{P^n_{\theta_0 + \frac{h}{\sqrt{n}}}} \mathcal{N}(0, I_{\theta_0}^{-1}) \text{ for any } h \in R^k$$

In that case, by Lemma 3 of Hirano and Porter (2009):

$$\frac{\sqrt{n} g(\hat{\theta}_n)}{\hat{\sigma}_g} \to_{P^n_{\theta_0 + \frac{h}{\sqrt{n}}}} \mathcal{N}(\frac{\dot{g}'h}{\sigma_g}, 1) \text{ for any } h \in R^k \tag{8}$$

Example of Quantiles (Continued)

By (10), it follows that δ^* has a finite sample counterpart $\delta_n^* = I\{\frac{\sqrt{n}g(\hat{\theta}_n)}{\hat{\sigma_g}} > \Phi^{-1}(\lambda)\}$ in a sense that $\delta_n^* \to_{P^n_{\theta_0+\frac{h}{\sqrt{n}}}} \delta^*$ for any $h \in R^k$.

Finally, by adjusting Lemma 1 of Hirano and Porter (2009) for quantiles it follows that

$$lim_{n\to\infty} B_n(\delta_n^*, \hat{\theta}_n, \pi) = B_\infty(\delta^*, \theta_0, \pi)$$

Thus, δ_n^* is an asymptotically optimal rule minimizing Bayesian λ-quantile of loss.

Minmax Quantile

- Since under δ^* quantile of loss is zero for every h, δ^* is also minmax optimal;
- By following previous steps and applying Lemma 4 of Hirano and Porter (2009) it also follows that:

$$lim_{n\to\infty} sup_h Q^{\lambda}_{P^n_{\theta_0+\frac{h}{\sqrt{n}}}} \left(L(\delta_n^*, \theta_0 + \frac{h}{\sqrt{n}}) \right) = sup_h Q^{\lambda}_{\mathcal{N}(h, I_{\theta_0}^{-1})} (L_\infty(\delta^*, h))$$

Whats next?

- Semiparametric setting;
- Empirical application.