ROBUST HYPOTHESIS TESTING VIA Lq-LIKELIHOOD

Yichen Qin and Carey E. Priebe

University of Cincinnati and Johns Hopkins University

Supplementary Material

S1 Assumptions and Explanations

In the manuscript, we have made use of the following assumptions.

Assumption 1 For any $q \in (0,1]$, f satisfies the following regularity conditions:

1. θ_0 is an interior point in Θ_0.
2. $\sup_{\theta \in \Theta_0} \|\frac{1}{n}\sum_{i=1}^{n} \tilde{\psi}(X_i;\theta,q) - \mathbb{E}\tilde{\psi}(X;\theta,q)\| \xrightarrow{p} 0$ as $n \to \infty$, where $\|\cdot\|$ represents the ℓ_2-norm.
3. $\max_{1\le k\le p} \mathbb{E}_{\theta_0}|\tilde{\psi}_k(X_i;\theta_0,q)|^3,\ k = 1,...,p$ is upper bounded by a constant, where $\tilde{\psi}_k$ is the k-th element of $\tilde{\psi}$.
4. The smallest eigenvalue of $\mathbf{A}$ is bounded away from zero.
5. Let b_{jk} be the j-th row, k-th column element in $\mathbf{B}$, then b_{jk}^2 for $j,k = 1,...,p$ are upper bounded by a constant.

6. The second order partial derivatives of $\tilde{\psi}(x;\theta,q)$ are dominated by an integrable functions with respect to the true distribution of X for all θ in a neighborhood of θ_0.

Assumption 2 For any $\varepsilon \in (0,1)$, the gross error model h is such that $\mathbb{E}_h[f''_\theta(X;\theta^*_{\varepsilon,1})/f(X;\theta^*_{\varepsilon,1})]$ is positive definite, where $\theta^*_{\varepsilon,1} = \arg\max_\theta \mathbb{E}_h[\log f(X;\theta)]$.

Assumption 3 $\mathbb{E}_f[f'_\theta(X;\theta_0)f'_\theta(X;\theta_0)^T(f(X;\theta_0)^{-2q} - f(X;\theta_0)^{-q-1})]$ is negative definite for any $q \in (0,1)$.

Assumption 4 There exists a constant $q^{**} \in (0,1)$, such that $\lambda_j(\mathbf{A}_{\varepsilon,q}\mathbf{B}^{-1}_{\varepsilon,q})$ are monotonic function in q for any $q \in (q^{**},1)$.

Remark on Assumption 1: The regularity conditions in Assumption 1 are based on the regularity conditions introduced in Ferrari and Yang (2010) pp. 760, B1, B3, C1 through C4, with some modifications. In particular, their regularity conditions are based on ψ and for the case of $q_n \to 1$, whereas ours are based on $\tilde{\psi}$ and for the case of the fixed q.

Remark on Assumption 2: For an exponential family (i.e., $f = m(x)\exp\{\theta^T T(x) - A(\theta)\}$, with the sufficient statistic $T(X)$), Assumption 2 becomes that $\mathbb{E}_h\{(T(X) - A'_\theta(\theta^*_{\varepsilon,1}))(T(X) - A'_\theta(\theta^*_{\varepsilon,1}))^T\} - A''_\theta(\theta^*_{\varepsilon,1})$ is positive definite. Note that

$$\mathbb{E}_h\{(T(X) - A'_\theta(\theta^*_{\varepsilon,1}))(T(X) - A'_\theta(\theta^*_{\varepsilon,1}))^T\}$$

$$= \mathrm{Cov}_h T(X) + (\mathbb{E}_h T(X) - A'_\theta(\theta^*_{\varepsilon,1}))(\mathbb{E}_h T(X) - A'_\theta(\theta^*_{\varepsilon,1}))^T,$$

where the second term on the right is positive semidefinite. This assumption implies either that the covariance of $T(X)$ under h, $\mathrm{Cov}_h T(X)$, is "larger" than the covariance of $T(X)$ under $f(x;\theta^*_{\varepsilon,1})$, $A''_\theta(\theta^*_{\varepsilon,1})$, or that $(\mathbb{E}_h T(X) - A'_\theta(\theta^*_{\varepsilon,1}))(\mathbb{E}_h T(X) - A'_\theta(\theta^*_{\varepsilon,1}))^T$ has large positive eigenvalues. Note that if $h = f(x;\theta^*_{\varepsilon,1})$, then $\mathbb{E}_h\{(T(X) - A'_\theta(\theta^*_{\varepsilon,1}))(T(X) - A'_\theta(\theta^*_{\varepsilon,1}))^T\} = A''_\theta(\theta^*_{\varepsilon,1})$.

More specifically, suppose the gross error model h contains a true null model $N(0,1)$ and a contamination model $N(0,\sigma^2)$, i.e., $h(x) = (1-\varepsilon)\varphi(x;0,1) + \varepsilon\varphi(x;0,\sigma^2)$. Here $\varphi(x;0,1)$ represents the pdf of the standard normal distribution and $\varphi(x;0,\sigma^2)$ represents the pdf of the contamination distribution. In this case, Assumption 2 is equivalent to $\sigma^2 > 1$. To see this, consider a one-dimensional problem where $\mathbb{E}_h[f''_\theta(X;\theta^*_{\varepsilon,1})/f(X;\theta^*_{\varepsilon,1})]$ is a scalar and $f = \varphi(x;\theta,1)$ is the pdf of a normal distribution with an unknown mean θ and a known variance, we have

$$f'_\theta(x) = f(x)(x-\theta)$$

$$f''_\theta(x) = f(x)((x-\theta)^2 - 1)$$

Plugging them into $\mathbb{E}_h[f''_\theta(X;\theta^*_{\varepsilon,1})/f(X;\theta^*_{\varepsilon,1})]$, where $\theta^*_{\varepsilon,1} = 0$ (due to the symmetric contamination $N(0,\sigma^2)$), we have

$$\begin{aligned}\mathbb{E}_h[f''_\theta(X)/f(X)] &= \mathbb{E}_h[X^2 - 1] \\ &= \varepsilon(\sigma^2 - 1)\end{aligned}$$

So, if $\sigma^2 > 1$, for any $\varepsilon \in (0,1)$, we have $\mathbb{E}_h[f''_\theta(X)/f(X)] > 0$. To sum up, for a normal mixture gross error model $h(x) = (1-\varepsilon)\varphi(x;0,1)+\varepsilon\varphi(x;0,\sigma^2)$, the range of values of σ^2 satisfying Assumption 2 is $\sigma^2 \in (1,+\infty)$.

Remark on Assumption 3: When f belongs to the exponential family, Assumption 3 becomes that $\mathbb{E}_f(T(X) - A'_\theta(\theta_0))(T(X) - A'_\theta(\theta_0))^T(f^{2(1-q)} - f^{1-q})$ is negative definite for any $q \in (0,1)$. In this case, $\mathbb{E}_f(T(X) - A'_\theta(\theta_0))(T(X) - A'_\theta(\theta_0))^T$ is $\mathrm{Cov}T(X)$. This assumption implies that the weighted version of the covariance of $T(X)$, using the weight $f^{2(1-q)}$, is "smaller" than the weighted version of the covariance of $T(X)$ using the weight f^{1-q}.

More specifically, let us continue with the same example from the previous remark. Suppose $f = \varphi(x;\theta,1)$ is the pdf of a normal distribution with an unknown mean θ and a known variance. Consider a one-dimensional problem where $\mathbb{E}_f[f'_\theta(X;\theta_0)f'_\theta(X;\theta_0)^T(f(X;\theta_0)^{-2q} - f(X;\theta_0)^{-q-1})]$ is a scalar and $\theta_0 = 0$ represent the the true mean. Therefore, we have

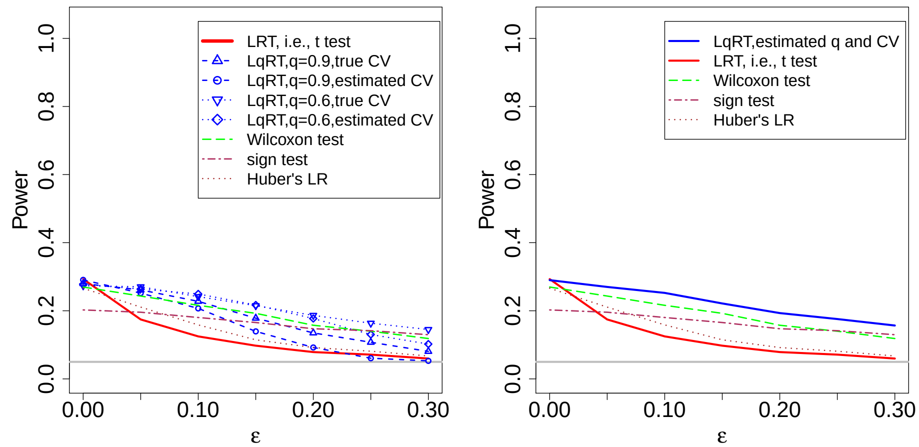
$$
\begin{aligned}
&\mathbb{E}_f\Big[f'_\theta(X)f'_\theta(X)^T\big(f(X)^{-2q} - f(X)^{-q-1}\big)\Big]\\
=&\mathbb{E}_f\Big[\Big(\frac{f'_\theta(X)}{f(X)}\Big)^2\big(f(X)^{2(1-q)} - f(X)^{1-q}\big)\Big]\\
=&\mathbb{E}_f\Big[X^2\big(f(X)^{2(1-q)} - f(X)^{1-q}\big)\Big]\\
=&\int_{-\infty}^{+\infty} x^2\Big[\frac{1}{\sqrt{2\pi}}\exp\Big\{-\frac{x^2}{2}\Big\}\Big]^{1+2(1-q)}dx - \int_{-\infty}^{+\infty} x^2\Big[\frac{1}{\sqrt{2\pi}}\exp\Big\{-\frac{x^2}{2}\Big\}\Big]^{1+(1-q)}dx\\
=&(2\pi)^{q-1}(3-2q)^{-\frac{3}{2}} - (2\pi)^{\frac{1}{2}(q-1)}(2-q)^{-\frac{3}{2}}
\end{aligned}
$$

For any $q \in (0,1)$, we have $q-1 < (q-1)/2 < 0$ and $3-2q > 2-q > 0$. Therefore, $(2\pi)^{q-1} < (2\pi)^{\frac{1}{2}(q-1)}$ and $(3-2q)^{-\frac{3}{2}} < (2-q)^{-\frac{3}{2}}$. Hence we have $\mathbb{E}_f\big[f'_\theta(X)f'_\theta(X)^T(f(X)^{-2q} - f(X)^{-q-1})\big] < 0$ and f satisfies Assumption 3.

S2 Additional Simulation Studies

S2.1 Simulation on Testing the Mean of the Normal Distribution Under Different Alternative Models

We extend the simulation from Section 4.1 in the manuscript to test the power of LqRT for different alternative hypotheses. We simulated data using $h(x;\theta,\varepsilon)$ with $\theta = 0.2, 0.5$, and 0.8 and compared the powers of these tests. The results in Figures 1, 2, and 3 indicate similar phenomena as



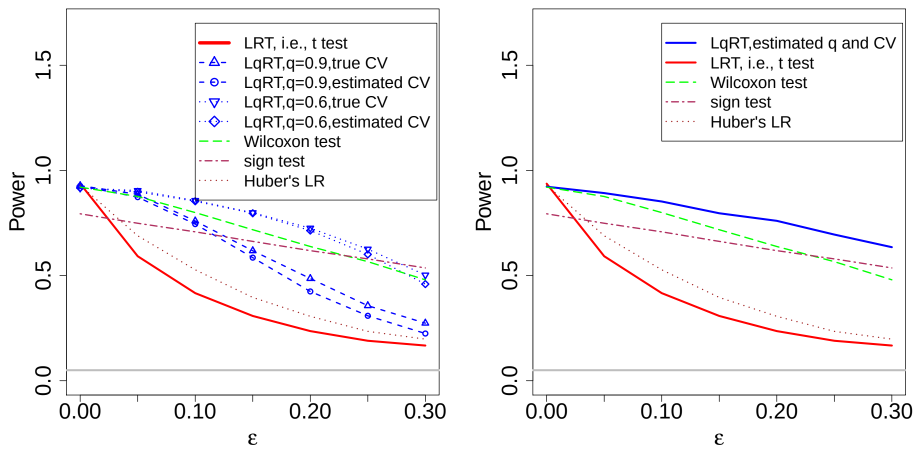
(a) Fixed q (b) Estimated q

Figure 1: Comparison of the powers for the LqRT, LRT, Wilcoxon test, sign test, and HLRT under different levels of heavy-tail contamination when testing for the mean of the normal distribution ($H_0 : \theta = 0$, $H_1 : \theta \neq 0$). The mean of the data generating process is $\theta = 0.2$.

shown in the manuscript.

S2.2 Simulation on Testing the Mean of the Normal Distribution Under Point Mass Contamination

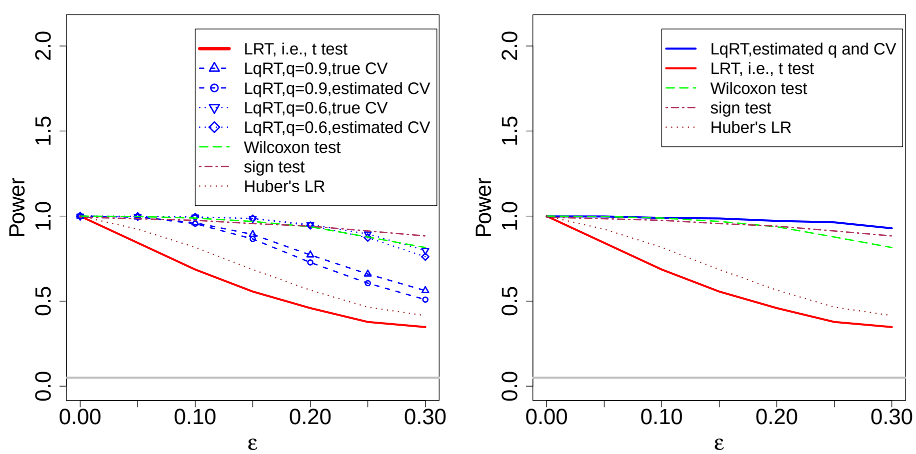
We investigate the proposed method in a point mass contamination setting. We adopted the setup of Section 4.1 in the manuscript and changed the data generating process to $h(x; \theta, \varepsilon) = (1 - \varepsilon)\varphi(x; \theta, 1) + \varepsilon\varphi(x; -5, 0.0001)$. The results are presented in Figure 4. In the left panel, as ε increases, the sizes of the LRT, the LqRT with $q = 0.9$, the Wilcoxon test, and the sign test all increase above 0.05. However, the sizes of LqRT with $q = 0.6$ and LqRT with estimated q are controlled at 0.05. With regard to the power, we



(a) Fixed q

(b) Estimated q

Figure 2: Comparison of the powers for the LqRT, LRT, Wilcoxon test, sign test, and HLRT under different levels of heavy-tail contamination when testing for the mean of the normal distribution ($H_0 : \theta = 0,\ H_1 : \theta \neq 0$). The mean of the data generating process is $\theta = 0.5$.



(a) Fixed q

(b) Estimated q

Figure 3: Comparison of the powers for the LqRT, LRT, Wilcoxon test, sign test, and HLRT under different levels of heavy-tail contamination when testing for the mean of the normal distribution ($H_0 : \theta = 0,\ H_1 : \theta \neq 0$). The mean of the data generating process is $\theta = 0.8$.

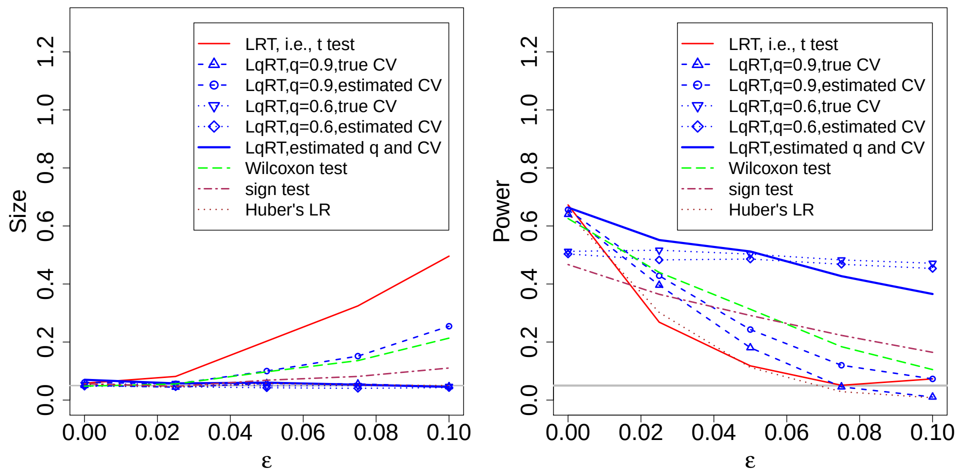


Figure 4: Comparison of powers and sizes for the LqRT with fixed q ($q = 0.9, 0.6$), the LqRT with estimated q and estimated critical value, LRT, Wilcoxon test, sign test, and HLRT under different levels of point mass contamination when testing for the mean of the normal distribution ($H_0 : \theta = 0$, $H_1 : \theta \neq 0$). The powers are calculated using the data generating process with mean $\theta = 0.34$.

observe the same phenomena as shown in the manuscript.

For Figure 4, we report the estimated q in each iteration for calculating the average power and size in Figures 5 and 6, respectively. The estimated q gradually decreases as the contamination becomes more serious.

Comparing Figures 4, 5, and 6, the LqRT with estimated q again combines the advantages of the LqRTs with fixed qs in terms of power. The size is also better controlled at 5% for the LqRT with estimated q.

Lastly, in Figure 7, we report estimated q for calculating the average size shown in Figure 9 in the manuscript.

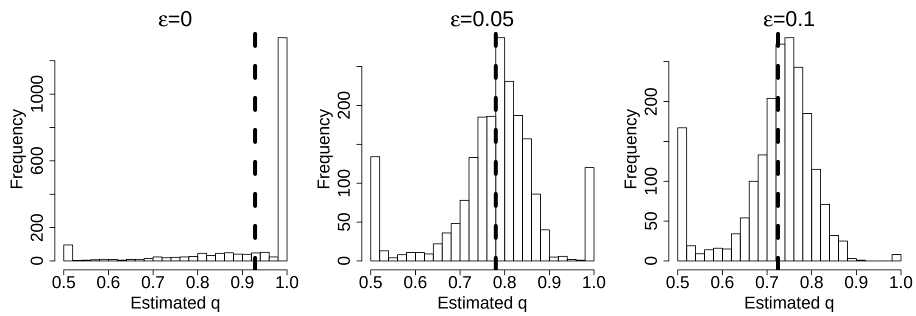


Figure 5: Histogram of the estimated q of the LqRT at different levels of point mass contamination when testing for the mean of the normal distribution. These estimated qs are obtained when calculating the powers (i.e., the right panel of Figure 4). The mean estimated q is indicated by a vertical dashed line.

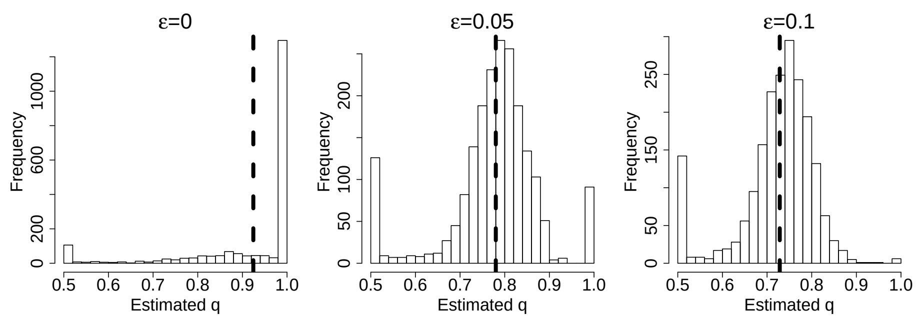


Figure 6: Histogram of the estimated q of the LqRT at different levels of point mass contamination when testing for the mean of the normal distribution. These estimated qs are obtained when calculating the sizes (i.e., the left panel of Figure 4). The mean estimated q is indicated by a vertical dashed line.

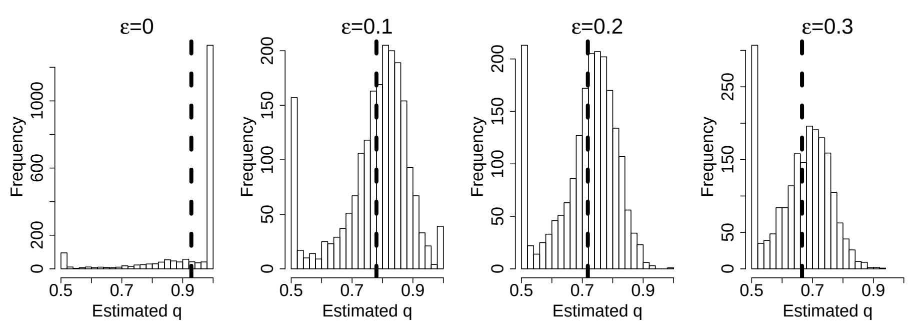


Figure 7: Histogram of the estimated q of the LqRT at different levels of heavy-tail contamination when testing for the mean of the normal distribution. These estimated qs are obtained when calculating the sizes (i.e., the left panel of Figure 9 in the manuscript). The mean estimated q is indicated by a vertical dashed line.

S3 Proofs of Theorems

S3.1 Sketch of Proof of Consistency of BCMLqE

Proof.

$$\begin{aligned}
&\frac{\partial}{\partial\theta}[\mathbb{E}_{\theta_0} L_q(f(X;\theta)) - C(\theta,q)]\Big|_{\theta=\theta_0} \\
&= \big[\mathbb{E}_{\theta_0}[f'_\theta(X;\theta)f(X;\theta)^{-q}] - \int f(x;\theta)^{1-q} f'_\theta(x;\theta)dx\big]\Big|_{\theta=\theta_0} \\
&= \big[\int f'_\theta(x;\theta)f(x;\theta)^{-q}f(x;\theta_0)dx - \int f(x;\theta)^{1-q} f'_\theta(x;\theta)dx\big]\Big|_{\theta=\theta_0} \\
&= 0,
\end{aligned}$$

and the consistency of the BCMLqE follows. □

S3.2 Proof of Theorem 3

Proof. When $q = 1$, we have $C(\theta,1) = 1$, $c(\theta,1) = 0$, and $c'(\theta,1) = 0$. Hence,

$$\begin{aligned}
\psi(x;\theta,1) &= \frac{f'_\theta(x;\theta)}{f(x;\theta)}, \\
\psi'(x;\theta,1) &= \frac{f''_\theta(x;\theta)}{f(x;\theta)} - \frac{f'_\theta(x;\theta)f'_\theta(x;\theta)^T}{f(x;\theta)^2}, \\
\mathbf{A}_{\varepsilon,1} &= \mathbb{E}_h[\psi(X;\theta^*_{\varepsilon,1},1)\psi(X;\theta^*_{\varepsilon,1},1)^T] = \mathbb{E}_h\frac{f'_\theta(X;\theta^*_{\varepsilon,1})f'_\theta(X;\theta^*_{\varepsilon,1})^T}{f(X;\theta^*_{\varepsilon,1})^2},
\end{aligned}$$

$$\mathbf{B}_{\varepsilon,1} = -\mathbb{E}_h \psi'(X;\theta^*_{\varepsilon,1},1) = -\mathbb{E}_h[\frac{f''_\theta(X;\theta^*_{\varepsilon,1})}{f(X;\theta^*_{\varepsilon,1})} - \frac{f'_\theta(X;\theta^*_{\varepsilon,1})f'_\theta(X;\theta^*_{\varepsilon,1})^T}{f(X;\theta^*_{\varepsilon,1})^2}].$$

Therefore,

$$\mathbf{A}_{\varepsilon,1} - \mathbf{B}_{\varepsilon,1} = \mathbb{E}_h[\frac{f''_\theta(X;\theta^*_{\varepsilon,1})}{f(X;\theta^*_{\varepsilon,1})}].$$

By Assumption 2, we know that $\mathbf{A}_{\varepsilon,1} - \mathbf{B}_{\varepsilon,1}$ is positive definite and that $\lambda_j(\mathbf{A}_{\varepsilon,1} - \mathbf{B}_{\varepsilon,1}) > 0$. Since $\mathbf{B}_{\varepsilon,1}$ is positive definite, we know $\lambda_j(\mathbf{A}_{\varepsilon,1}\mathbf{B}^{-1}_{\varepsilon,1}) > 1$. When f belongs to the exponential family, we have $f(x;\theta) = m(x)\exp\{\theta^T T(x) - A(\theta)\}$, $f'_\theta(x;\theta) = f(x;\theta)(T(x) - A'_\theta(\theta))$ and $f''_\theta(X;\theta) = f(x;\theta)[(T(x) - A'_\theta(\theta))(T(x) - A'_\theta(\theta))^T - A''_\theta(\theta)]$, such that

$$\mathbb{E}_h[\frac{f''_\theta(X;\theta^*_{\varepsilon,1})}{f(X;\theta^*_{\varepsilon,1})}] = \mathbb{E}_h[(T(x) - A'_\theta(\theta^*_{\varepsilon,1}))(T(x) - A'_\theta(\theta^*_{\varepsilon,1}))^T - A''_\theta(\theta^*_{\varepsilon,1})].$$

□

S3.3 Proof of Theorem 4

Proof.

$$\psi(x;\theta,q) = \frac{f'_\theta(x;\theta)}{f(x;\theta)} f(x;\theta)^{1-q},$$
$$\psi'(x;\theta,q) = \frac{f''_\theta(x;\theta)}{f(x;\theta)} f(x;\theta)^{1-q} - q\frac{f'_\theta(x;\theta)f'_\theta(x;\theta)^T}{f(x;\theta)^2} f(x;\theta)^{1-q},$$

$$
\begin{aligned}
c(\theta,q) &= \int f'_\theta(x;\theta)f(x;\theta)^{1-q}dx = \mathbb{E}_f f'_\theta(X;\theta)f(X;\theta)^{-q},\\
c'(\theta,q) &= \int f''_\theta(x;\theta)f(x;\theta)^{1-q} + f'_\theta(x;\theta)f'_\theta(x;\theta)^T(1-q)f(x;\theta)^{-q}dx\\
&= \mathbb{E}_f f''_\theta(X;\theta)f(X;\theta)^{-q} + \mathbb{E}_f\frac{f'_\theta(x;\theta)f'_\theta(x;\theta)^T}{f(x;\theta)^2}(1-q)f(x;\theta)^{1-q},\\
\mathbf{A}_{\varepsilon,q} &= \mathbb{E}_h[(\psi(X;\theta^*_{\varepsilon,q},q)-c(\theta^*_{\varepsilon,q},q))(\psi(X;\theta^*_{\varepsilon,q},q)-c(\theta^*_{\varepsilon,q},q))^T]\\
&= \mathbb{E}_h[(\psi(X;\theta^*_{\varepsilon,q},q)-c(\theta^*_{\varepsilon,q},q))\psi(X;\theta^*_{\varepsilon,q},q)^T] - \{\mathbb{E}_h[\psi(X;\theta^*_{\varepsilon,q},q)-c(\theta^*_{\varepsilon,q},q)]\}c(\theta^*_{\varepsilon,q},q)^T\\
&= \mathbb{E}_h[(\psi(X;\theta^*_{\varepsilon,q},q)-c(\theta^*_{\varepsilon,q},q))\psi(X;\theta^*_{\varepsilon,q},q)^T]\\
&= \mathbb{E}_h[\psi(X;\theta^*_{\varepsilon,q},q)\psi(X;\theta^*_{\varepsilon,q},q)^T] - c(\theta^*_{\varepsilon,q},q)\mathbb{E}_h\psi(X;\theta^*_{\varepsilon,q},q)^T,\\
\mathbf{B}_{\varepsilon,q} &= -\mathbb{E}_h[\psi'(X;\theta^*_{\varepsilon,q},q)-c'(\theta^*_{\varepsilon,q},q)]\\
&= c'(\theta^*_{\varepsilon,q},q)-\mathbb{E}_h[\psi'(X;\theta^*_{\varepsilon,q},q)].
\end{aligned}
$$

We further have

$$
\begin{aligned}
\mathbf{A}_{\varepsilon,q}-\mathbf{B}_{\varepsilon,q} &= \mathbb{E}_h[\psi\psi^T] - c\mathbb{E}_h[\psi^T] - c' + \mathbb{E}_h[\psi']\\
&= \mathbb{E}_h\frac{f'_\theta f'^T_\theta}{f^2}f^{2(1-q)} - c\mathbb{E}_h[\frac{f'^T_\theta}{f}f^{1-q}] - c' + \mathbb{E}_h[\frac{f''_\theta}{f}f^{1-q} - q\frac{f'_\theta f'^T_\theta}{f^2}f^{1-q}]\\
&= \mathbb{E}_h\frac{f'_\theta f'^T_\theta}{f^2}f^{2(1-q)} - \mathbb{E}_f[f'_\theta f^{-q}]\mathbb{E}_h[\frac{f'^T_\theta}{f}f^{1-q}] - \mathbb{E}_f[f''_\theta f^{-q}] - (1-q)\mathbb{E}_f[\frac{f'_\theta f'^T_\theta}{f^2}f^{1-q}]\\
&\quad + \mathbb{E}_h[\frac{f''_\theta}{f}f^{1-q} - q\frac{f'_\theta f'^T_\theta}{f^2}f^{1-q}]\\
&= \mathbb{E}_h\frac{f'_\theta f'^T_\theta}{f^2}f^{2(1-q)} - \mathbb{E}_f[f'_\theta f^{-q}]\mathbb{E}_h[\frac{f'^T_\theta}{f}f^{1-q}] - \mathbb{E}_f[f''_\theta f^{-q}] + (q-1)\mathbb{E}_f[\frac{f'_\theta f'^T_\theta}{f^2}f^{1-q}]\\
&\quad + \mathbb{E}_h[\frac{f''_\theta}{f}f^{1-q} - q\frac{f'_\theta f'^T_\theta}{f^2}f^{1-q}].
\end{aligned}
$$

When $\varepsilon = 0$, $h = f$, we have $\theta^*_{0,q} = \theta_0$, and

$$\begin{aligned}\mathbf{A}_{0,q} - \mathbf{B}_{0,q} &= \mathbb{E}_f \frac{f'_\theta f'^T_\theta}{f^2} f^{2(1-q)} - \mathbb{E}_f[f'_\theta f^{-q}]\mathbb{E}_f[\frac{f'^T_\theta}{f} f^{1-q}] - \mathbb{E}_f[\frac{f'_\theta f'^T_\theta}{f^2} f^{1-q}] \\ &= \mathbb{E}_f \frac{f'_\theta f'^T_\theta}{f^2}(f^{2(1-q)} - f^{1-q}) - \mathbb{E}_f[f'_\theta f^{-q}]\mathbb{E}_f[f'_\theta f^{1-q}]^T.\end{aligned}$$

We know $\mathbb{E}_f[f'_\theta f^{-q}]\mathbb{E}_f[f'_\theta f^{1-q}]^T$ is always positive semidefinite. By Assumption 3, we know that $\mathbf{A}_{0,q} - \mathbf{B}_{0,q}$ is negative definite, which means $\lambda_j(\mathbf{A}_{0,q}\mathbf{B}^{-1}_{0,q}) < 1$.

We denote $\lambda_j(\varepsilon, q) = \lambda_j(\mathbf{A}_{\varepsilon,q}\mathbf{B}^{-1}_{\varepsilon,q})$. Suppose that for any $\varepsilon \in (0, E]$, $\lambda_j(\varepsilon, 1) > 1$. For any $q \in [Q, 1)$ where $Q \in (q^{**}, 1)$, $\lambda_j(0, q) < 1$. We can define $g(\varepsilon) = \frac{1-Q}{E}\varepsilon + Q$. Since $\lambda_j(E, g(E)) > 1$ and $\lambda_j(0, g(0)) < 1$, by the intermediate value theorem, there exists an $\tilde{\varepsilon} \in (0, E)$, such that $\lambda_j(\tilde{\varepsilon}, g(\tilde{\varepsilon})) = 1$. Furthermore, for any $\varepsilon \in (0, \tilde{\varepsilon})$, we have $\lambda_j(\varepsilon, g(\varepsilon)) < 1$. Therefore, for any $\varepsilon \in (0, \tilde{\varepsilon})$, we know that $\lambda_j(\varepsilon, g(\varepsilon)) < 1 < \lambda_j(\varepsilon, 1)$. By Assumption 4, we know $\lambda_j(\varepsilon, q)$ is increasing in q for any $q \in (g(\varepsilon), 1)$.

We let $L = \lambda_j(\varepsilon, g(\varepsilon))$ and $U = \lambda_j(\varepsilon, 1)$ so that $L < 1 < U$. Here, L and U are introduced as lower and upper bounds of $\lambda_j(\varepsilon, q)$ for $q \in (g(\varepsilon), 1)$. If $1 - L = |L - 1| < |U - 1| = U - 1$, by Assumption 4, we know that for any $q \in (g(\varepsilon), 1)$ where $\varepsilon \in (0, \tilde{\varepsilon})$, we have $|\lambda_j(\varepsilon, q) - 1| < |\lambda_j(\varepsilon, 1) - 1|$.

On the other hand, if $1 - L = |L - 1| \geq |U - 1| = U - 1$, since we know

$\lambda_j(\varepsilon, q)$ is increasing in q for any $q \in (g(\varepsilon), 1)$, there exists a $g^*(\varepsilon) \in (g(\varepsilon), 1)$, such that $|\lambda_j(\varepsilon, g^*(\varepsilon)) - 1| < |U - 1|$. Therefore, for any $q \in (g^*(\varepsilon), 1)$ where $\varepsilon \in (0, \tilde{\varepsilon})$, we have $|\lambda_j(\varepsilon, q) - 1| < |\lambda_j(\varepsilon, 1) - 1|$. Note that, in the statement of the theorem, we refer to $g(\varepsilon)$ or $g^*(\varepsilon)$ as q^*.

To sum up, there exists an $\tilde{\varepsilon} \in (0, 1)$, such that for any arbitrary $\varepsilon \in (0, \tilde{\varepsilon})$, there exists a $q^* \in (q^{**}, 1)$ and for any $q \in (q^*, 1)$, we have $|\lambda_j(\varepsilon, q) - 1| < |\lambda_j(\varepsilon, 1) - 1|$.

When f belongs to the exponential family, we have

$$\mathbb{E}_f \frac{f'_\theta f'^T_\theta}{f^2}(f^{2(1-q)} - f^{1-q}) = \mathbb{E}_f (T(X) - A'_\theta(\theta_0))(T(X) - A'_\theta(\theta_0))^T (f^{2(1-q)} - f^{1-q}).$$

□

S3.4 Proof of Theorem 5

We will prove the simple null hypothesis case. To start, we note that $t_{\min} = 0$ and $t_{\min} = \infty$. Then we can rewrite our test statistic as $D_q = \frac{1}{2}(\hat{\theta}_q - \theta_0)^T n\mathbf{B}(\theta^*)(\hat{\theta}_q - \theta_0)$, where θ^* is between $\hat{\theta}_q$ and θ_0. Since we know $\mathbf{B}(\theta)$ is continuous in θ, we conclude that our test statistic and $\hat{\theta}_q$ have the same breakdown points.

Bibliography

Ferrari, D. and Yang, Y. (2010). Maximum Lq-likelihood estimation. *Annals of Statistics*, 38:753–783.