Statistics 582, Problem Set 3, Solutions

Wellner; 1/25/2018

1. (vdV *Asymp. Stat.*, problem 14, page 83): Suppose that we observe a random sample from the distribution of (X, Y) in the following *errors in variables* model:

$$\begin{aligned} X &= Z + e \\ Y &= \alpha + \beta Z + f \end{aligned}$$

where (e, f) is bivariate normally distributed with mean 0 and covariance matrix $\sigma^2 I$ and is independent from the unobservable variable Z. In analogy to Example 5.26, construct a system of estimating equations for $\theta = (\alpha, \beta)$ based on a conditional likelihood, and study the properties of the corresponding estimators $\widehat{\theta}_n = (\widehat{\alpha}_n, \widehat{\beta}_n)$. In particular, what is the limiting distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ under some reasonable assumptions about existence of moments of Z?

Hint: Condition on the unobserved $(Z_1, \ldots, Z_n) = (z_1, \ldots, z_n)$ and treat the z_i's as parameters in the model. Then the joint density $p(x_i, y_i) \equiv p(x_i, y_i; \alpha, \beta, z_i, \sigma^2)$ of (X_i, Y_i) is Gaussian. From there, proceed as follows:

(a) Show that the log-likelihood of the data $(X_1, Y_1), \ldots, (X_n, Y_n)$ is given by

$$l_n(\alpha, \beta, \underline{z}, \sigma^2) = -n \log(2\pi\sigma^2) - \frac{1}{\sigma^2} \left\{ \sum_{i=1}^n (Y_i - \alpha - \beta z_i)^2 + \sum_{i=1}^n (X_i - z_i)^2 \right\}.$$

(b) Maximize $l_n(\alpha, \beta, \underline{z}, \sigma^2)$ for fixed α, $\beta, \sigma^2)$ as a function of $\underline{z}$ to find

$$\begin{aligned} l_n^{prof,1}(\alpha, \beta, \sigma^2) &\equiv l_n(\alpha, \beta, \hat{\underline{z}}(\alpha, \beta), \sigma^2) \\ &= -n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2(1+\beta^2)} \sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2. \end{aligned}$$

(c) Maximize $l_n^{prof,1}(\alpha, \beta, \sigma^2)$ with respect to α for fixed β, σ^2 to find

$$\begin{aligned} l_n^{prof,2}(\beta, \sigma^2) &= l_n^{prof,1}(\hat{\alpha}(\beta), \beta, \sigma^2) \\ &= -n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2(1+\beta^2)} \sum_{i=1}^n (Y_i - \overline{Y} - \beta(X_i - \overline{X}))^2. \end{aligned}$$

(d) Maximize $l_n^{prof,2}(\beta, \sigma^2)$ with respect to β to find $\hat{\beta}$. You should find that:

$$\begin{aligned} \hat{z}_i &\equiv \hat{z}_i(\alpha, \beta) = X_i + \frac{(Y_i - \alpha - \beta X_i)\beta}{1+\beta^2}, \quad i = 1, \ldots, n, \\ \hat{\alpha} &\equiv \hat{\alpha}(\beta) = \overline{Y} - \beta \overline{X}, \\ \hat{\beta} &= \frac{S_{YY} - S_{XX} + \sqrt{(S_{YY} - S_{XX})^2 + 4S_{XY}^2}}{2S_{XY}} \end{aligned}$$

where

$$S_{XX} = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \qquad S_{YY} = n^{-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2,$$
$$S_{XY} = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

(e) Now study the asymptotic properties of $(\hat{\alpha}_n, \hat{\beta}_n)$.
This is sometimes called *orthogonal regression*, since the fitted line minimizes orthogonal distances from observed points to the regression line. If instead of equal variances for the errors we had assumed $\sigma_\epsilon^2/\sigma_f^2 = \delta$, then the solution becomes

$$\begin{aligned}
\hat{z}_i &\equiv \hat{z}_i(\alpha,\beta) = X_i + \frac{(Y_i - \alpha - \beta X_i)\beta}{\delta + \beta^2}, \quad i = 1, \ldots, n, \\
\hat{\alpha} &\equiv \hat{\alpha}(\beta) = \overline{Y} - \beta\overline{X}, \\
\hat{\beta} &= \frac{S_{YY} - \delta S_{XX} + \sqrt{(S_{YY} - \delta S_{XX})^2 + 4\delta S_{XY}^2}}{2S_{XY}}.
\end{aligned}$$

This is called *Deming regression*. For semiparametric (*structural*) versions of these models, see Bickel and Ritov (1987), van der Vaart (1996), and Bickel, Klaassen, Ritov, and W (1993, 1998), Example 4.5.2, pages 127-128 and 135 - 139.

Solution: (a) Conditioning on $Z_i = z_i$'s, the joint density of (X_i, Y_i) is given by

$$p(x_i, y_i; \alpha, \beta, z_i, \sigma^2) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}(x_i - z_i)^2 + (y_i - \alpha - \beta z_i)^2\right).$$

It follows that the log-likelihood for all the data is given by

$$l_n(\alpha, \beta, \underline{z}, \sigma^2) = -n\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left\{\sum_{i=1}^{n}(Y_i - \alpha - \beta z_i)^2 + \sum_{i=1}^{n}(X_i - z_i)^2\right\}.$$

(b) Differentiating the log-likelihood with respect to z_i yields

$$\frac{\partial}{\partial z_i} l_n(\alpha, \beta, \underline{z}, \sigma^2) = -\frac{2}{2\sigma^2}\{-(Y_i - (\alpha + \beta z_i))\beta - (X_i - z_i)\} = 0$$

if

$$z_i = \hat{z}_i = \hat{z}_i(\alpha, \beta) = \frac{(Y_i - \alpha)\beta + X_i}{1 + \beta^2} = X_i + \frac{(Y_i - \alpha - \beta X_i)\beta}{1 + \beta^2}.$$

Upon noting that

$$\begin{aligned}
Y_i - (\alpha + \beta\hat{z}_i) &= \frac{Y_i - \alpha - \beta X_i}{1 + \beta^2}, \quad \text{and} \\
X_i - \hat{z}_i &= -\frac{Y_i - \alpha - \beta X_i}{1 + \beta^2},
\end{aligned}$$

we find that

$$\begin{aligned}
l_n^{prof,1}(\alpha, \beta, \sigma^2) &\equiv l_n(\alpha, \beta, \hat{\underline{z}}(\alpha, \beta), \sigma^2) \\
&= -n\log(2\pi\sigma^2) - \frac{1}{2\sigma^2(1 + \beta^2)}\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2.
\end{aligned}$$

(c) Differentiating $l_n^{prof,1}$ with respect to α yields

$$\frac{\partial}{\partial \alpha} l_n^{prof,1}(\alpha, \beta, \sigma^2) = \frac{1}{\sigma^2(1+\beta^2)} \sum_{i=1}^n (Y_i - \alpha - \beta X_i)(-1) = 0$$

if $\alpha = \hat{\alpha}(\beta) = \overline{Y}_n - \beta \overline{X}_n$. Plugging this back into $l_n^{prof,1}$ yields

$$\begin{aligned} l_n^{prof,2}(\beta, \sigma^2) &= l_n^{prof,1}(\hat{\alpha}(\beta), \beta, \sigma^2) \\ &= -n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2(1+\beta^2)} \sum_{i=1}^n (Y_i - \overline{Y} - \beta(X_i - \overline{X}))^2 \\ &= -n \log(2\pi\sigma^2) - \frac{n}{2\sigma^2(1+\beta^2)} \left\{ S_{YY} - 2\beta S_{XY} + \beta^2 S_{XX} \right\}. \end{aligned}$$

(d) Differentiating $l_n^{prof,2}$ with respect to β yields

$$\begin{aligned} \frac{\partial}{\partial \beta} l_n^{prof,2}(\beta, \sigma^2) &= \frac{-n}{\sigma^2(1+\beta^2)} \left\{ \{-S_{XY} + \beta S_{XX}\} - \frac{\beta}{(1+\beta^2)} \left\{ S_{YY} - 2\beta S_{XY} + \beta^2 S_{XX} \right\} \right\} \\ &= \frac{n}{\sigma^2(1+\beta^2)^2} \left\{ \beta^2 S_{XY} + \beta(S_{XX} - S_{YY}) - S_{XY} \right\} = 0 \end{aligned}$$

if

$$\beta = \hat{\beta} = \frac{S_{YY} - S_{XX} + \sqrt{(S_{YY} - S_{XX})^2 + 4S_{XY}^2}}{2S_{XY}}$$

as claimed.
Although it was not required or mentioned in the problem statement, it is also of interest to estimate the variance parameter $\sigma_e^2 = \sigma_f^2 \equiv \sigma^2$. Thus we form

$$l_n^{prof,3}(\sigma^2) \equiv l_n^{prof,2}(\hat{\beta}, \sigma^2) = -n \log(2\pi\sigma^2) - \frac{n}{2\sigma^2(1+\hat{\beta}^2)} \{S_{YY} - 2\hat{\beta} S_{XY} + \hat{\beta}^2 S_{XX}\}.$$

Differentiating with respect to σ^2 yields

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} l_n^{prof,3}(\sigma^2) &= -\frac{n}{\sigma^2} + \frac{n}{2}(\sigma^2)^{-2} \frac{S_{YY} - 2\hat{\beta} S_{XY} + \hat{\beta}^2 S_{XX}}{(1+\hat{\beta}^2)} \\ &\equiv -\frac{n}{(\sigma^2)^2} \left\{ \sigma^2 - 2^{-1} S_p^2 \right\} \end{aligned}$$

and hence

$$\hat{\sigma}^2 = 2^{-1} S_p^2 = 2^{-1} \frac{S_{YY} - 2\hat{\beta} S_{XY} + \hat{\beta}^2 S_{XX}}{(1+\hat{\beta}^2)}.$$

(e) To study the properties of these estimators, first note that

$$\begin{aligned} S_{XX} &= n^{-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2 \\ &= n^{-1} \left\{ \sum_{i=1}^n (Z_i - \overline{Z}_n)^2 + \sum_{i=1}^n (e_i - \overline{e}_n)^2 + 2 \sum_1^n (Z_i - \overline{Z})(e_i - \overline{e}_n) \right\} \\ &\to_p \sigma_Z^2 + \sigma^2 + 0, \end{aligned}$$

while

$$
\begin{aligned}
S_{YY} &= n^{-1}\sum_{i=1}^{n}(Y_i-\overline{Y}_n)^2 = n^{-1}\sum_{i=1}^{n}\{\beta(Z_i-\overline{Z}_n)+(f_i-\overline{f}_n)\}^2 \\
&= n^{-1}\left\{\sum_{i=1}^{n}\beta^2(Z_i-\overline{Z}_n)^2+\sum_{i=1}^{n}(f_i-\overline{f}_n)^2+2\beta\sum_{i=1}^{n}(Z_i-\overline{Z}_n)(f_i-\overline{f}_n)\right\} \\
&\to_p \beta^2\sigma_Z^2+\sigma^2.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
S_{XY} &= n^{-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)(Y_i-\overline{Y}_n) \\
&= n^{-1}\sum_{i=1}^{n}\{\beta(Z_i-\overline{Z}_n)+(f_i-\overline{f}_n)\}\{(Z_i-\overline{Z}_n)+(e_i-\overline{e}_n)\} \\
&\to_p \beta\sigma_Z^2+0.
\end{aligned}
$$

Thus we find (by application of the Mann-Wald theorem) that

$$
\begin{aligned}
(S_{YY}-S_{XX})^2+4S_{XY}^2 &\to_p (\beta^2-1)^2\sigma_Z^4+4\beta^2\sigma_Z^4 \\
&= (\beta^4+2\beta^2+1)\sigma_Z^4=(\beta^2+1)^2\sigma_Z^4
\end{aligned}
$$

and hence

$$
\hat{\beta}_n \to_p \frac{(\beta^2-1)\sigma_Z^2+(\beta^2+1)\sigma_Z^2}{2\beta\sigma_Z^2}=\beta.
$$

This, in turn, yields

$$
\hat{\alpha}=\overline{Y}_n-\hat{\beta}_n\overline{X}_n \to_p \alpha+\beta E(Z)-\beta E(Z)=\alpha
$$

and

$$
\begin{aligned}
\hat{\sigma}_n^2 &\to_p \frac{1}{1+\beta^2}\{\beta^2\sigma_Z^2+\sigma^2-2\beta^2\sigma_Z^2+\beta^2(\sigma_Z^2+\sigma^2)\} \\
&= \frac{1+\beta^2}{1+\beta^2}\frac{\sigma^2}{2}=\frac{\sigma^2}{2}.
\end{aligned}
$$

This inconsistency of the MLE of σ^2 is not too surprising in view of the previous observed inconsistency of the MLE of the variance in the Neyman - Scott model(s).

To find the limiting distribution of $\hat{\theta}=(\hat{\alpha},\hat{\beta})$ we will use an “M-theorem”. Let

$$
\mathbb{M}_n(\theta)=\mathbb{M}_n(\alpha,\beta)\equiv(1+\beta^2)^{-1}\mathbb{P}_n(Y-\alpha-\beta X)^2\equiv\mathbb{P}_n m_\theta(X,Y)
$$

and

$$
M(\theta)=M(\alpha,\beta)\equiv(1+\beta^2)^{-1}P(Y-\alpha-\beta X)^2\equiv P m_\theta(X,Y)
$$

This fits the context of vdV Theorem 5.23 (or "Theorem" on page 10 of the M- and Z-theorem handout) with

$$m_\theta(X,Y) \equiv m(X,Y;\alpha,\beta) = (Y-\alpha-\beta X)^2/(1+\beta^2).$$

Thus we compute

$$\begin{aligned}
\dot m_\theta(X,Y) &= \begin{pmatrix} -2(Y-\alpha-\beta X)/(1+\beta^2) \\ -2(Y-\alpha-\beta X)(X+\beta(Y-\alpha))/(1+\beta^2)^2 \end{pmatrix} \\
&= \begin{pmatrix} -2(Y-\alpha-\beta X)/(1+\beta^2) \\ -\frac{2}{1+\beta^2}e(X,Y) - 2\frac{\beta}{(1+\beta^2)^2}e(X,Y)^2 \end{pmatrix}
\end{aligned}$$

where $e(X,Y) \equiv e(X,Y;\alpha,\beta) \equiv Y-\alpha-\beta X$. This yields

$$P_0(\dot m_{\theta_0}\dot m_{\theta_0}^T) = \frac{4}{(1+\beta_0^2)^2}\begin{pmatrix} P_0 e_0^2 & \frac{P_0 e_0^2(X+\beta_0(Y-\alpha_0))}{1+\beta_0^2} \\ \frac{P_0 e_0^2(X+\beta_0(Y-\alpha_0))}{1+\beta_0^2} & \frac{P_0(e_0^2(X+\beta_0(Y-\alpha_0))^2)}{(1+\beta_0^2)^2} \end{pmatrix}$$

where $e_0(X,Y) \equiv e(X,Y;\alpha_0,\beta_0)$. To compute the matrix $V_0 \equiv V_{\theta_0}$ in the quadratic expansion of M we first compute

$$\ddot m_\theta(X,Y) = \begin{pmatrix} \frac{2}{(1+\beta^2)} & \frac{2X}{(1+\beta^2)} + \frac{4\beta(Y-\alpha-\beta X)}{(1+\beta^2)^2} \\ \frac{2X}{(1+\beta^2)} + \frac{4\beta(Y-\alpha-\beta X)}{(1+\beta^2)^2} & \ddot m_{\beta\beta}(X,Y) \end{pmatrix}$$

where

$$\begin{aligned}
&\ddot m_{\beta\beta}(X,Y) \\
&= \frac{2}{(1+\beta^2)^2}\left\{X(X+\beta(Y-\alpha)) + \frac{4\beta}{1+\beta^2}e(X,Y)(X+\beta(Y-\alpha)) - (Y-\alpha)e(X,Y)\right\} \\
&= \frac{2}{(1+\beta^2)^2}\left\{X^2(1+\beta^2) + \frac{4\beta^2 e^2(X,Y)}{1+\beta^2} + \{X\beta + 4\beta X - (Y-\alpha)\}e(X,Y)\right\}.
\end{aligned}$$

Then I calculate

$$V_{\theta_0} = \frac{2}{(1+\beta_0^2)}\begin{pmatrix} 1 & E(Z) \\ E(Z) & E(Z^2) + \sigma^2\left(1+\frac{4\beta_0^2}{1+\beta_0^2}\right) \end{pmatrix}.$$

This non-singular since

$$\det(V_{\theta_0}) = \frac{4}{(1+\beta_0^2)^2}\left\{Var(Z) + \sigma^2\left(1+\frac{4\beta_0^2}{1+\beta_0^2}\right)\right\} > 0.$$

Thus the M-theorem yields

$$\sqrt{n}(\hat\theta_n - \theta_0) \to_d N_2(0, V_{\theta_0}^{-1}P_0(\dot m_{\theta_0}\dot m_{\theta_0}^T)V_{\theta_0}^{-1}).$$

2. (a) (vdV *Asymp. Stat.*, problem 15, page 83): In Example 5.27, for what point is the least squares estimator consistent if we drop the the condition that $E(e|X) = 0$? Derive an (implicit) solution in terms of the function $E(e|X)$. Is it necessarily θ_0 if $E(e) = 0$?

(b) Investigate the limit distribution of $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ under the assumption $E(e|X) = 0$.

Solution: (a) Let $e_\theta(X,Y) \equiv Y - f_\theta(X)$. We expect that

$$\begin{aligned} \hat{\theta}_n &= \text{argmin}_\theta \mathbb{P}_n (Y - f_\theta(X))^2 \equiv \text{argmin}_\theta \mathbb{M}_n(\theta) \\ &= \text{argmin}_\theta n^{-1} \sum_{i=1}^n (Y_i - f_\theta(X_i))^2 \\ \rightarrow_{a.s.} &\ \text{argmin}_\theta M(\theta) \equiv P(Y - f_\theta(X))^2 = P(e_\theta^2(X) \equiv \theta^* \end{aligned}$$

where $(X,Y) \sim P$ on $\mathbb{R} \times \mathbb{R}$ (or even on $\mathbb{R}^d \times \mathbb{R}$) and $\theta^* \equiv \theta^*(P)$. Note that

$$M(\theta) = P(Y - f_\theta(X))^2 = EVar(Y|X) + E\{(E(Y|X) - f_\theta(X))^2\}$$

where the first term depends on the true joint distribution P of (X,Y) but not on θ. Thus we expect the minimizer θ^* of $\theta \mapsto M(\theta)$ to satisfy

$$0 = \dot{M}(\theta^*) = -2E\{(E(Y|X) - f_{\theta^*}(X))\dot{f}_{\theta^*}(X)\}. \tag{1}$$

Now suppose that $Y = f_{\theta_0}(X) + e$ where $E(e|X) \neq 0$. Then $E(Y|X) = f_{\theta_0}(X) + E(e|X)$, and (1) becomes

$$0 = E\{(E(e|X) + f_{\theta_0}(X) - f_{\theta^*}(X))\dot{f}_{\theta^*}(X)\}.$$

Since $E(e|X) \neq 0$, it is clear that $\theta^*(P) \neq \theta_0$ in general. When $E(e|X) = 0$, then (1) reduces to

$$0 = E\{(f_{\theta_0}(X) - f_{\theta^*}(X))\dot{f}_{\theta^*}(X)\}$$

which clearly holds for $\theta^* = \theta_0$.

Now assume that $E(e|X) = 0$ holds. If we assume that $\theta_0 \equiv \theta_0(P)$ satisfies

$$\sup_{\theta:\ |\theta - \theta_0| \geq \epsilon} M(\theta) < M(\theta_0)$$

for every $\epsilon > 0$ and

$$\sup_{\theta \in \Theta} |\mathbb{M}_n(\theta) - M(\theta)| \rightarrow_p 0, \tag{2}$$

then by Theorem 5.7 (discussed in Stat 581) that any $\hat{\theta}_n$ satisfying $\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\theta_0) - o_p(1)$ satisfies $\hat{\theta}_n \rightarrow_p \theta_0$. Note that

$$\mathbb{M}_n(\theta) - M(\theta) = (\mathbb{P}_n - P)(m_\theta(X,Y))$$

where $m_\theta(X,Y) \equiv (Y - f_\theta(X))^2$, so (2) holds if $\{m_\theta(X,Y) :\ \theta \in \Theta\}$ is a Glivenko-Cantell class for P.

To investigate asymptotic normality we need to consider the hypotheses of Theorem 5.23 for the functions $m_\theta(X,Y)$. Now

$$\dot{m}_\theta(X,Y) = -2(Y - f_\theta(X))\dot{f}_\theta(X),$$

so a natural candidate for $\dot{m}$ in Theorem 5.23 is

$$\sup_{\theta:|\theta-\theta_0|\le\delta}|\dot{m}_\theta(X,Y)| = 2\sup_{\theta:\ |\theta-\theta_0|\le\delta}|Y-f_\theta(X)|\cdot|\dot{f}_\theta(X)| \equiv \dot{m}(X,Y).$$

Thus we will assume that $P\dot{m}^2(X,Y)<\infty$. Furthermore, assume that

$$M(\theta) = Pm_\theta(X,Y) = M(\theta_0) + \frac{1}{2}(\theta-\theta_0)^2 V_{\theta_0} + o(\|\theta-\theta_0\|^2)$$

in a neighborhood of θ_0 where $V_{\theta_0}<0$ (and θ_0 maximizes $M(\theta)$ so that the derivative $\dot{M}(\theta_0)=0$). By Theorem 5.23 it $\mathbb{M}_n(\hat{\theta}_n)\ge\sup_\theta\mathbb{M}_n(\theta)-o_p(n^{-1})$ and $\hat{\theta}_n\to_p\theta_0$, then

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n-\theta_0) &= -V_{\theta_0}^{-1}\sqrt{n}\mathbb{P}_n\dot{m}_\theta\dot{m}_\theta(X,Y)+o_p(1)\\ &\to_d N(0, V_{\theta_0}^{-1}P(\dot{m}_\theta^2)V_{\theta_0}^{-1}).\end{aligned}$$

3. Consider the zero-inflated Poisson distribution p_θ as described in Example 3 of the handout on M- and Z- theorems. Suppose that $X_1,\ldots,X_n$ i.i.d. p_θ are observed.
(a) Set up alternative estimating equations for $\theta=(\gamma,\lambda)$ where $\gamma\in[0,1]$ and $\lambda>0$ based on $g_1(x)=x$ and $g_2(x)=x^2$. Express your alternative estimator $\hat{\theta}_n=(\hat{\gamma}_n,\hat{\lambda}_n)$ of θ explicitly in terms of the first and second moments, $\overline{X}_n$ and $\overline{X^2}_n$, of the data, and show that your estimators are consistent when the model holds.
(b) Use Huber's Z-theorem to show that $\sqrt{n}(\hat{\theta}_n-\theta_0)\to_d N_2(0,\Sigma)$ and give the form of Σ.
(c) What happens if the X_i's are i.i.d. $p\notin\mathcal{P}=\{p_\theta:\ \theta\in\Theta\}$? Describe the parameter $\theta(P)$ to which $\hat{\theta}_n$ converges in probability and use Huber's theorem to establish a limit theorem for $\sqrt{n}(\hat{\theta}_n-\theta(P))$ in this case.

Solution: (a) For $X\sim p_\theta$ we compute

$$\begin{aligned}E_\theta X &= (1-\gamma)\lambda,\\ E_\theta X^2 &= (1-\gamma)(\lambda+\lambda^2).\end{aligned}$$

Thus the method of moments estimators $\hat{\theta}_n$ of θ based on $g_1(x)=x$ and $g_2(x)=x^2$ are given by

$$\Psi_n(\hat{\theta}_n)=\mathbb{P}_n\psi(X;\hat{\theta}_n)=0$$

where

$$\begin{aligned}\psi_{\theta,1}(x) &= \psi_1(x;\theta)\equiv x-t_1(\theta)=x-(1-\gamma)\lambda,\\ \psi_{\theta,2}(x) &= \psi_2(x;\theta)\equiv x^2-t_2(\theta)=x^2-(1-\gamma)(\lambda+\lambda^2).\end{aligned}$$

The population versions of the estimating equations, $\Psi(\theta)=P\psi(X;\theta)=0$ can be rewritten as

$$E_PX=(1-\gamma)\lambda, \quad \text{and} \quad E_PX^2=(1-\gamma)(\lambda+\lambda^2),$$

and hence we find that

$$\frac{E_PX^2}{E_PX}=1+\lambda, \quad \text{or} \quad \lambda=\lambda(P)=\frac{E_PX^2}{E_PX}-1.$$

This leads to

$$(1-\gamma)\lambda = E_P(X) \quad \text{or} \quad \gamma = 1 - \frac{[E_P(X)]^2}{E_P(X^2)-E_P(X)} = \frac{Var_P(X)-E_P(X)}{E_P(X^2)-E_P(X)}.$$

The corresponding estimators become

$$\hat{\gamma}_n = 1 - \frac{\overline{X}_n^2}{\overline{X^2}_n - \overline{X}_n} = \frac{S_n^2 - \overline{X}_n}{\overline{X^2}_n - \overline{X}_n}, \quad \text{and}$$

$$\hat{\lambda}_n = \frac{\overline{X^2}_n}{\overline{X}_n} - 1.$$

When the model holds, these are consistent:

$$\hat{\lambda}_n \to_p \frac{(1-\gamma)(\lambda+\lambda^2)}{(1-\gamma)\lambda} - 1 = \lambda, \quad \text{and}$$

$$\hat{\gamma}_n \to_p \frac{(1-\gamma)[1+\gamma\lambda]-(1-\gamma)\lambda}{(1-\gamma)[\lambda+\lambda^2]-(1-\gamma)\lambda} = \frac{\gamma(1-\gamma)\lambda^2}{(1-\gamma)\lambda^2} = \gamma.$$

(b) If the models holds, we know $X \sim P_{\theta_0} \in \mathcal{P}$. Then with $P_0 \equiv P_{\theta_0}$ and $E_0 h(X) \equiv E_{\theta_0} h(X)$,

$$E_0 X^4 = (1-\gamma)E_{Poiss(\lambda)}X^4 = (1-\gamma)(\lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4) < \infty,$$

so $X^2 \in L_2(P_0)$ and the covariance matrix $\Sigma_0 \equiv Cov_0(X, X^2)$ is well defined. By the multivariate CLT it follows that

$$\sqrt{n}(\Psi_n - \Psi)(\theta_0) = \sqrt{n}\begin{pmatrix} \overline{X}_n - E_0(X) \\ \overline{X^2}_n - E_0 X^2 \end{pmatrix} \to_d N_2(0, \Sigma_0).$$

so A2 holds. As noted in class on 21 January, condition A3 holds trivially for method of moment estimators. To verify A4 we write

$$\Psi(\theta) = P\psi(X;\theta) = P\begin{pmatrix} X - t_1(\theta) \\ X^2 - t_2(\theta) \end{pmatrix},$$

and hence

$$\dot{\Psi}(\theta_0) \quad = \quad -\dot{t}(\theta) = \begin{pmatrix} \lambda & -(1-\gamma) \\ \lambda+\lambda^2 & -(1-\gamma)(1+2\lambda) \end{pmatrix}$$

where $\det(-\dot{\Psi}(\theta_0)) = (1-\gamma)\lambda^2 > 0$ if $0 \le \gamma < 1$. Thus we conclude from Huber's theorem that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d -\dot{\Psi}_0^{-1} Z \sim N_2(0, B\Sigma_0 B^T)$$

where $B \equiv -\dot{\Psi}_0^{-1}$.

(c) As can be seen from (a), when $P \notin \mathcal{P}$, if we assume that $E_P(X^2) < \infty$, then

$$\hat{\gamma}_n \to_p (Var_P(X) - E_P(X))/(E_P(X^2) - E_P(X)), \quad \text{and}$$

$$\hat{\lambda}_n \to_p \frac{E_P(X^2)}{E_P(X)} - 1 \equiv \lambda(P).$$

Note that this $\lambda(P) \ge 0$ for any distribution on $\{0,1,2,\ldots\}$ with equality holding if P is Bernoulli(p) for any $p \in (0,1]$. Supposing that $\theta_0(P)$ is well-defined, Huber's theorem holds if $E_P(X^4) < \infty$. Then $\sqrt{n}(\Psi_n - \Psi)(\theta_0) \to_d N_2(0, \Sigma_P)$ where $\Sigma_P = Cov_P(X, X^2)$ and $\dot{\Psi}(\theta_0)$ remains as in (b) but with $\theta_0 = \theta(P)$. Thus Huber's theorem yields $\sqrt{n}(\hat{\theta}_n - \theta(P)) \to_d N_2(0, B_P \Sigma_P B_P^T)$.

4. Let X be a random variable with distribution function F having finite first moment: $E|X| < \infty$.
(a) Show that $f(b) \equiv E|X - b|$ is minimized by $b =$ any median of the distribution F of X. [A median m of F is any value satisfying $F(m) = P(X \le m) \ge 1/2$ and $1 - F(m-) = P(X \ge m) \ge 1/2$; see Lehmann and Casella, TPE, page 62, problems 1.7 and 1.8.]
(b) For $0 < \tau < 1$, let $\rho_\tau(x) = x(\tau - 1_{(-\infty,0)}(x))$. Consider minimizing

$$M_\tau(\theta) = E\rho_\tau(X - \theta)$$

with respect to θ. Show that the solution $\theta_0 = \theta_0(F)$ is given by the $\tau-$th quantile of F: $\theta_0(F) = F^{-1}(\tau)$.

(a) Suppose that m is a median of F. From Lehmann and Casella problem 1.7, it follows that $m_0 \le m \le m_1$ so that the set of medians is a closed interval.
This is easily proved as follows: suppose that $\mathcal{M}$ is the set of medians of F. Note that $\mathcal{M}$ is always non-empty since $m_0 \equiv \inf\{x : \ F(x) \ge 1/2\} \in \mathcal{M}$. If $\mathcal{M} = \{m_0\}$, then $[m_0, m_0] = \{m_0\}$ is closed. If If $a, b \in \mathcal{M}$ with $a < b$, then if $c \in (a, b)$ we have $P(X \le c) \ge P(X \ge a) \ge 1/2$ (since $a \in \mathcal{M}$), and $P(X \ge c) \ge P(X \ge b) \ge 1/2$ (since $b \in \mathcal{M}$). Thus $c \in \mathcal{M}$ and hence $(a, b) \subset \mathcal{M}$. Let $(m_0, m_1) = \cup_{a,b\in\mathcal{M}}(a, b)$ be the union of all the open intervals contained in $\mathcal{M}$. Then if $m \in (m_0, m_1)$

$$\begin{aligned} 1/2 &\le P(X \le m) = E1\{X \le m\} \to E1\{X < m_1\} \le E1\{X \le m_1\} = P(X \le m_1), \quad \text{and} \\ 1/2 &\le P(X \ge m) = E1\{X \ge m\} \to E1\{X \ge m_1\} = P(X \ge m_1) \end{aligned}$$

as $m \nearrow m_1$ by the dominated convergence theorem. Thus $m_1 \in \mathcal{M}$. Similarly,

$$\begin{aligned} 1/2 &\le P(X \le m) = E1\{X \le m\} \to E1\{X \le m_0\} \le P(X \le m_0), \quad \text{and} \\ 1/2 &\le P(X \ge m) = E1\{X \ge m\} \to E1\{X > m_0\} \le P(X \ge m_0) \end{aligned}$$

as $m \searrow m_0$ by the dominated convergence theorem. Thus $m_0 \in \mathcal{M}$ and we conclude that $[m_0, m_1] \subset \mathcal{M}$. On the other hand $\mathcal{M} \subset [m_0, m_1]$ with $m_0 \equiv \inf\{x : F(x) \ge 1/2\}$ and $m_1 \equiv \inf\{x : F(x) > 1/2\}$.
Now suppose that $c > m_1$ and $m \in [m_0, m_1]$. By examining the graphs of $|x - c|$ and $|x - m|$ we see that

$$\begin{aligned} |x - c| - |x - m| &= (m - c)1_{[x\ge c]} + (c - m)1_{[x\le m]} + \{(c - x) - (x - m)\}1_{[m<x<c]} \\ &= (c - m)\left\{1_{[x\le m]} - 1_{[x\ge c]}\right\} + (c + m - 2x)1_{[m<x<c]} \\ &= (c - m)\left\{1_{[x\le m]} - 1_{[x\ge c]}\right\} + 2(c - x)1_{[m<x<c]} - (c - m)1_{[m<x<c]} \\ &= (c - m)\left\{1_{[x\le m]} - 1_{[x>m]}\right\} + 2(c - x)1_{[m<x<c]}. \end{aligned}$$

Replacing x by X and taking expectations across the identity with respect to X yields

$$\begin{aligned} E|X - c| - E|X - m| &= (c - m)\{P(X \le m) - P(X > m)\} + 2E\{(c - X)1_{[m<X<c]}\} \\ &\ge 0 + 0 = 0 \end{aligned}$$

since m is a median of F implies that $P(X \le m) - P(X > m) \ge 0$ and $c > m_1 \ge m_0$ implies that $E\{(c - X)1_{[m<X<c]}\} = E\{(c - X)1_{[m_1<X<c]}\} > 0$. Similarly, if $c < m_0$,

$$|x - c| - |x - m| = (m - c)(1_{[x\ge m]} - 1_{[x<m]}) + 2(x - c)1_{[c<x<m]},$$

and taking expectations yields

$$\begin{aligned} E|X-c| - E|X-m| &= (m-c)\{P(X \geq m) - P(X < m)\} + 2E\{(X-c)1_{[c<X<m]}\} \\ &\geq 0. \end{aligned}$$

Thus $E|X-b|$ is minimized by any median of the distribution F of X.
(b) First rewrite $M_\tau(\theta)$ as

$$\begin{aligned} M_\tau(\theta) &= E\rho_\tau(X-\theta) = (\tau-1)\int_{(-\infty,\theta]}(x-\theta)dF(x) + \tau\int_{(\theta,\infty)}(x-\theta)dF(x) \\ &= (\tau-1)\int_{(-\infty,\theta]}(x-\theta)f(x)dx + \tau\int_{(\theta,\infty)}(x-\theta)f(x)dx \end{aligned}$$

if F is absolutely continuous with density f. Thus in this case

$$\begin{aligned} M_\tau'(\theta) &= -(\tau-1)\int_{(-\infty,\theta]}dF(x) - \tau\int_{(\theta,\infty)}dF(x) \\ &= (1-\tau)F(\theta) - \tau(1-F(\theta)) = F(\theta) - \tau \\ &= 0 \end{aligned}$$

if θ is a $\tau-$quantile of F, $\theta = \theta_0(F) = F^{-1}(\tau)$. If F is an arbitrary distribution, then an argument as in (a) shows that any $\tau-$th quantile of F minimizes $M_\tau(\theta)$. The "check function" $\rho_\tau(x)$ is the key trick in the formulation of quantile regression: see e.g. Koenker (2005), *Quantile Regression*, especially his Figure 1.2, page 6.

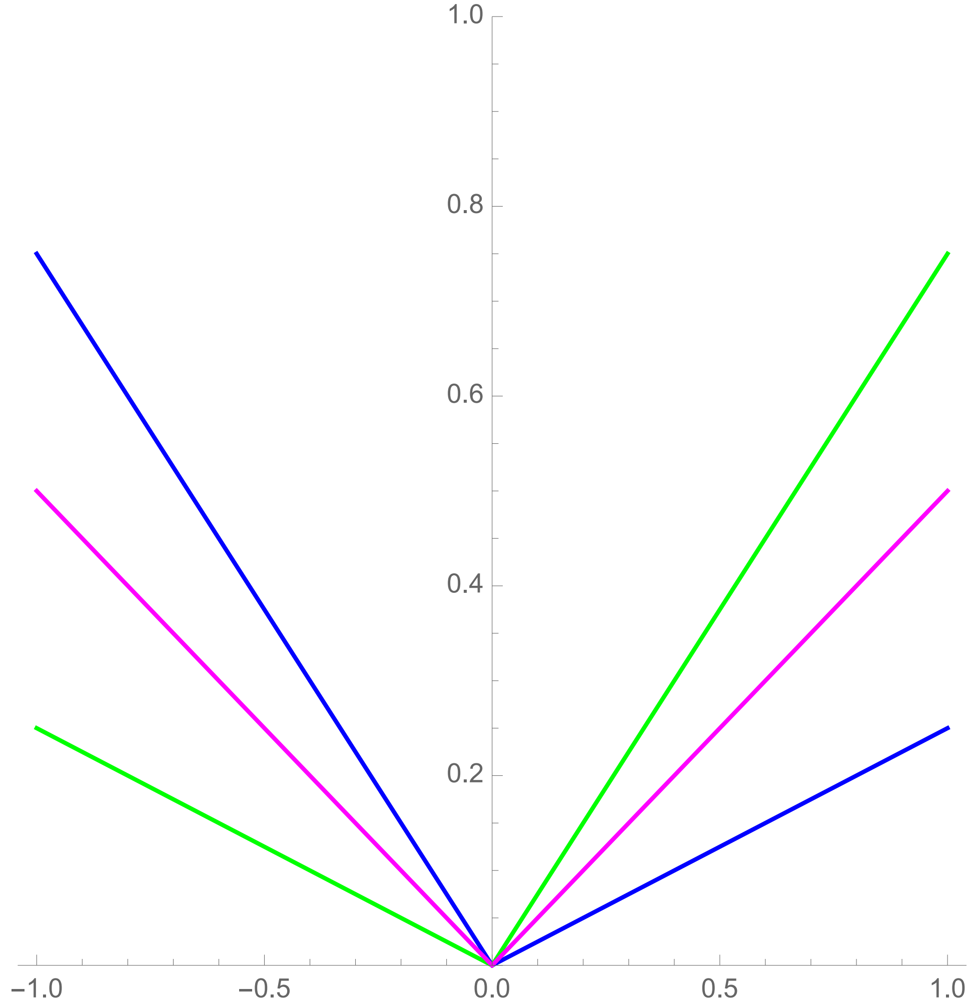

Figure 1: Plots of $\rho_\tau(x)$, $\tau \in \{1/4, 1/2, 3/4\}$; green $=$ 3/4, magenta $=$ 1/2, blue $=$ 1/4.