Statistics 581, Final Exam Solutions

Wellner; 12/13/2010

1. (40 points) **Define** the following terms. In each case, provide an appropriate (brief) context for your definition.
   (a) The *Kullback - Leibler* divergence (or information) between a probability measure P and another (sub-)probability measure Q on the same measurable space $(\mathcal{X}, \mathcal{A})$.
   (b) The *Hellinger distance* between two probability measures P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$.
   (c) The inverse or quantile function F^{-1} corresponding to a d.f. F on R.
   (d) An $n^{1/4}-$consistent preliminary estimator of θ in a parametric model.
   (e) A one-step estimator a (vector-)parameter θ in a regular parametric model.

   Solution: See Course Notes, Chapters 1-4.

2. (40 points) **State** *four* of the following five results, providing an appropriate (brief) context for your statements:
   (a) A result about the finite-dimensional limiting distributions of the sample quantile process $\{\sqrt{n}(\mathbb{F}_n^{-1}(t) - F^{-1}(t)) : \ 0 < t < 1\}$ specifying the assumption(s) carefully.
   (b) The asymptotic behavior of the likelihood ratio statistic $2 \log \lambda_n$ (assuming the MLE's $\hat{\theta}_n$ exist) for testing a simple null hypothesis $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$ under a fixed alternative $\theta \neq \theta_0$ in a regular parametric model.
   (c) The asymptotic behavior of the likelihood ratio statistic $2 \log \lambda_n$ (assuming the MLE's $\hat{\theta}_n$ exist) for testing a simple null hypothesis $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$ under a sequence of local alternatives $\theta_n = \theta_0 + tn^{-1/2}$ in a regular parametric model.
   (d) A relationship between squared Hellinger distance $H^2(P, Q)$ and the Hellinger affinity $\rho(P, Q) = \int \sqrt{pq} d\mu$.
   (e) A result relating $H^2(P^n, Q^n)$ to $H^2(P, Q)$ where P^n is the measure corresponding to the product density $\prod_{i=1}^n p(x_i)$ of $X_1, \ldots, X_n$ i.i.d. P (and similarly for Q^n).

   Solution: See Course Notes, Chapters 1-4.

Do either problem 3 or problem 4:

3. (40 points) Let $X_1, \ldots, X_n$ be i.i.d. $P_\theta = \text{Normal}(\theta, 1)$.
(a) Give the Hodges superefficient estimator T_n of θ (with superefficiency at $\theta = 0$).
(b) What is the limiting distribution of $\sqrt{n}(T_n - \theta)$ as a function of θ?
(c) What is the limiting distribution of $\sqrt{n}(T_n - \theta_n)$ when sampling from $\theta = \theta_n$ when $\theta_n = cn^{-1/2}$?
(d) Does the limit distribution in (c) depend on c? Is T_n a locally regular estimator of θ at $\theta = 0$?
(e) What is the limit of $E_{\theta_n}\{[\sqrt{n}(T_n - \theta_n)]^2\}$ when $\theta_n = cn^{-1/2}$ as in (c)? For what values of c does the limiting risk of T_n exceed the (limiting) risk of $\overline{X}_n$?

Solution: See Course Notes, Chapter 3.

4. (40 points)
(a) **State** the Glivenko-Cantelli theorem. Then **prove** that it holds *if it holds* for the case of i.i.d. Uniform$(0,1)$ random variables.
(b) **Prove** the Glivenko-Cantelli theorem for i.i.d. Uniform$(0,1)$ random variables: if $\xi_1, \ldots, \xi_n, \ldots$ are i.i.d. Uniform$(0,1)$ with empirical distribution function

$$\mathbb{G}_n(t) = \frac{1}{n}\sum_{i=1}^n 1_{[0,t]}(\xi_i)\,, \qquad \text{then} \qquad \sup_{0\le t\le 1} |\mathbb{G}_n(t) - t| \to_{a.s.} 0\,.$$

Solution: See Course Notes, Chapter 2.

Do either problem 5 or problem 6:

5. (40 points)
Suppose that $\underline{X}, \underline{X}_1, \ldots, \underline{X}_n$ are i.i.d. $\text{Mult}_k(1, \underline{p})$, so that $\underline{N}_n \equiv \sum_{i=1}^n \underline{X}_i \sim \text{Mult}_k(n, \underline{p})$. Thus

$$
\begin{aligned}
P_{\underline{p}}(\underline{X} = \underline{x}) &= \prod_{j=1}^k p_j^{x_j} \quad \text{for } \; x_i \in \{0,1\}, \; \sum_1^k x_i = 1\,, \\
P_{\underline{p},n}(\underline{N}_n = \underline{m}) &= \frac{n!}{\prod_{j=1}^k m_j!} \prod_{j=1}^k p_j^{m_j} \quad \text{for } \; m_i \geq 0, \text{ integers } \sum_{j=1}^k m_j = n\,.
\end{aligned}
$$

(a) Compute $K(P_{\underline{q}}, P_{\underline{p}}) \equiv K(\underline{q}, \underline{p})$ for vectors $\underline{q}$, $\underline{p}$ with $\sum p_j = \sum q_j = 1$.
(b) Evaluate $K(\widehat{\underline{p}}, \underline{p})$ where $\widehat{\underline{p}} = n^{-1}\underline{N}_n$. Relate this to the log-likelihood $\log L_n(\underline{p}|\underline{N}_n)$.
(c) Use the result of (b) to show, without using any calculus, that the MLE of $\underline{p}$ is $\widehat{\underline{p}} = \underline{N}/n$.

Solution: (a) First,

$$
\log \frac{p_{\underline{q}}(\underline{x})}{p_{\underline{p}}(\underline{x})} = \log \prod_{j=1}^k \frac{q_j^{x_j}}{p_j^{x_j}} = \sum_{j=1}^k x_j \log(\frac{q_j}{p_j})\,.
$$

Thus

$$
K(\underline{q}, \underline{p}) = \sum_{j=1}^k q_j \log \frac{q_j}{p_j}\,.
$$

(b) From A it follows that

$$
K(\widehat{\underline{p}}, \underline{p}) = \sum_{j=1}^k \widehat{p}_j \log \frac{\widehat{p}_j}{p_j} = -\sum_{j=1}^k \widehat{p}_j \log \frac{p_j}{\widehat{p}_j}\,.
$$

Now

$$
\begin{aligned}
\log L_n(\underline{p}|\underline{N}_n) &= \sum_{j=1}^k N_j \log p_j + \log\left(\frac{n!}{\prod N_j!}\right) \\
&= n \sum_{j=1}^k \widehat{p}_j \log p_j + \log\left(\frac{n!}{\prod N_j!}\right) \\
&= n \sum_{j=1}^k \widehat{p}_j \log\left(\frac{p_j}{\widehat{p}_j}\right) + n \sum_{j=1}^k \widehat{p}_j \log \widehat{p}_j + \log\left(\frac{n!}{\prod N_j!}\right) \\
&= -K(\widehat{\underline{p}}, \underline{p}) + \text{terms constant in } \underline{p}\,.
\end{aligned}
$$

Even more neatly, as several of you noted,

$$
\log \frac{L_n(\widehat{\underline{p}}|\underline{N}_n)}{L_n(\underline{p}|\underline{N}_n)} = n \sum_{j=1}^k \{\widehat{p}_j \log \widehat{p}_j - \widehat{p}_j \log p_j\} = nK(\widehat{\underline{p}}, \underline{p})\,.
$$

(c) Since $K(\underline{\widehat{p}}, \underline{p}) \ge 0$ with equality if and only if $\underline{p} = \underline{\widehat{p}}$, we see from the identity in B that $L_n(\underline{p}|\underline{N}_n)$ is maximized by $\underline{p} = \underline{\widehat{p}}$.

6. (40 points).
Suppose that $P = P_0 = N(0,1)$, $Q = P_\theta = N(\theta, 1)$ on $(\mathbb{X}, \mathcal{A}) = (\mathbb{R}, \mathcal{B})$.
(a) Compute $K(P,Q) = K(P_0, P_\theta)$.
(b) Compute $H^2(P,Q) = 1 - \rho(P,Q)$ and $\rho(P,Q) = \int \sqrt{p(x)q(x)}dx$. [It might be easiest to compute $\rho(P,Q)$ first recalling that if $Z \sim N(0,1)$ then $E\exp(tZ) = \exp(t^2/2)$.]
(c) Compute $d_{TV}(P,Q) = 1 - \eta(P,Q)$ and $\eta(P,Q) = \int p(x) \wedge q(x) dx$. [It might be easiest to compute $\eta(P,Q)$ first.]
(d) Show in general that $K(P,Q) \ge 2H^2(P,Q)$, thereby strengthening the fact $K(P,Q) \ge 0$ that we proved in class. [Hint: write both $K(P,Q)$ and $H^2(P,Q)$ in terms of $Y = (p(X)/q(X))^{1/2}$ and use the inequality $\log(1+x) \ge x/(1+x)$ for $x \ge 0$. You will need to relate $E_Q Y$ and $E_Q Y^2$ to $H^2(P,Q)$.]
(e) Use the results of (a) and (d) to find a lower bound for $K(P^n, Q^n)$ in terms of $H^2(P,Q)$ or $\rho(P,Q)$; here P^n and Q^n are the probability distributions of $X_1, \ldots, X_n$ i.i.d. as P and Q^n respectively.

Solution: (a) Now $p(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ and $q(x) = (2\pi)^{-1/2}\exp(-(x-\theta)^2/2)$, so

$$\frac{p(x)}{q(x)} = \exp(-x^2/2 + (x-\theta)^2/2) = \exp(-\theta x + \theta^2/2),$$
$$\log\frac{p}{q}(x) = -\theta x + \theta^2/2,$$

and it follows that

$$K(P,Q) = E_P\left(\log\frac{p}{q}\right) = -\theta E_P(X) + \theta^2/2 = \theta^2/2.$$

(b) We compute $\rho(P,Q)$ first:

$$\begin{aligned}\rho(P,Q) &= \int \sqrt{p(x)q(x)}dx = \int \frac{1}{\sqrt{2\pi}}\exp(-x^2/4)\exp(-(x-\theta)^2/4)dx \\ &= \int \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\exp((2\theta x - \theta^2)/4)\,dx \\ &= \exp(-\theta^2/4)E_P\exp(\theta X/2) = \exp(-\theta^2/4)\exp(\theta^2/8) = \exp(-\theta^2/8).\end{aligned}$$

Hence

$$H^2(P,Q) = 1 - \rho(P,Q) = 1 - \exp(-\theta^2/8).$$

(c) We compute $\eta(P,Q)$ first. Since $\phi(x) \ge \phi(x-\theta)$ if and only if

$$\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \ge \frac{1}{\sqrt{2\pi}}\exp(-(x-\theta)^2/2)$$

or equivalently, if and only if

$$1 \ge \exp(\theta x - \theta^2/2) \qquad \text{iff} \qquad \theta x - \theta^2/2 \le 0$$

and, for $\theta > 0$, this holds if and only if $x \le \theta/2$. Hence it follows that

$$\begin{aligned}
\int p(x) \wedge q(x) dx &= \int \phi(x) \wedge \phi(x-\theta) dx \\
&= \int_{-\infty}^{\theta/2} \phi(x-\theta) dx + \int_{\theta/2}^{\infty} \phi(x) dx \\
&= \int_{-\infty}^{-\theta/2} \phi(y) dy + 1 - \Phi(\theta/2) \\
&= \Phi(-\theta/2) + 1 - \Phi(\theta/2)\,, \qquad \text{if } \ \theta > 0 \\
&= 2\Phi(-|\theta|/2), \quad \text{if } \theta > 0.
\end{aligned}$$

When $\theta < 0$, $\theta x - \theta^2/2 < 0$ if and only if $x \ge \theta/2$, and this yields

$$\begin{aligned}
\int p(x) \wedge q(x) dx &= \int_{\theta/2}^{\infty} \phi(x-\theta) dx + \int_{-\infty}^{\theta/2} \phi(x) dx \\
&= \Phi(\theta/2) + 1 - \Phi(-\theta/2)\,, \qquad \text{if } \ \theta > 0 \\
&= 2\Phi(-|\theta|/2), \quad \theta < 0.
\end{aligned}$$

It follows that $\eta(P,Q) = 2\Phi(-|\theta|/2)$ for all θ, and hence

$$d_{TV}(P,Q) = 1 - \eta(P,Q) = 1 - 2\Phi(-|\theta|/2).$$

The following Figure shows $K(P_0, P_\theta)$, $H^2(P_0, P_\theta)$, and $d_{TV}(P_0, P_\theta)$ as functions of θ.

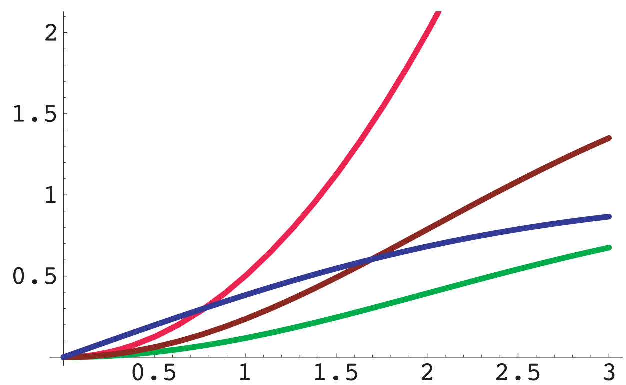


Figure 1: $K(P_0, P_\theta)$ (red), $H^2(P_0, P_\theta)$ (green), $d_{TV}(P_0, P_\theta)$ (blue), and $2H^2(P_0, P_\theta)$ (burgundy) as functions of θ

(d) Let $Y \equiv \sqrt{p/q}$. Then

$$\begin{aligned}
K(P,Q) &= 2\int p \log(\sqrt{p/q}) d\mu = 2E_P \log Y = 2E_P \log(1 + (Y-1)) \\
&\ge 2E_P \frac{Y-1}{1+Y-1} \qquad \text{using } \ \log(1+x) \ge \frac{x}{1+x}, \\
&= 2E_P \frac{Y-1}{Y} = 2\int p \left\{ \sqrt{\frac{p}{q}} - 1 \right\} \sqrt{\frac{q}{p}} d\mu
\end{aligned}$$

$$
\begin{aligned}
&= 2\int p\left\{1-\sqrt{\frac{q}{p}}\right\}d\mu = 2\left\{1-\int \sqrt{pq}\,d\mu\right\} \\
&= 2H^2(P,Q).
\end{aligned}
$$

Here is a another way of organizing the argument with a slightly different choice of Y, as follows. Let $Y \equiv \sqrt{p/q}-1$. Then $p/q = (1+Y)^2$ and it follows that

$$
\begin{aligned}
K(P,Q) &= \int p\log(p/q)d\mu = \int (p/q)\log(p/q)q\,d\mu \\
&= 2\int (p/q)\log(p/q)^{1/2}dQ = 2\int (1+Y)^2\log(1+Y)dQ \\
&\geq 2\int (1+Y)^2\frac{Y}{1+Y}dQ \\
&= 2\int Y(1+Y)dQ = 2\left\{\int Y\,dQ + \int Y^2 dQ\right\}.
\end{aligned}
$$

But we also have

$$
\begin{aligned}
\int Y\,dQ &= \int \sqrt{pq}\,d\mu - 1 = -H^2(P,Q)\,, \\
\int Y^2 dQ &= \int [\sqrt{p/q}-1]^2 q\,d\mu = \int [\sqrt{p}-\sqrt{q}]^2 d\mu = 2H^2(P,Q)\,.
\end{aligned}
$$

By combining the results of these last two displays it follows that

$$
K(P,Q) \geq 2\left\{2H^2(P,Q) - H^2(P,Q)\right\} = 2H^2(P,Q)\,.
$$

(e) By (d), and (c) of the next problem,

$$
K(P^n,Q^n) \geq 2H^2(P^n,Q^n) = 2\{1-\rho(P^n,Q^n)\} = 2\{1-\rho(P,Q)^n\}
$$

in general. For the particular case we began with $\rho(P,Q) = \exp(-\theta^2/8)$, and thus we conclude that in this case

$$
K(P^n,Q^n) \geq 2\{1-\exp(-n\theta^2/8)\}\,.
$$

Another lower bound follows from the identity $K(P^n,Q^n) = nK(P,Q)$: we conclude from this together with the inequality from (d) that

$$
K(P^n,Q^n) = nK(P,Q) \geq n2H^2(P,Q) = 2n(1-\exp(-\theta^2/8)).
$$

7. (72 points). (Bivariate normal means) Suppose that (X_i, Y_i), $i = 1, \ldots, n$ are i.i.d. bivariate normal $N((\mu, \nu)^T, \Sigma)$ where

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix}$$

and density

$$p_\theta(x, y) = \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left(\frac{(x-\mu)^2}{\sigma^2} - 2\rho \frac{(x-\mu)}{\sigma}\frac{(y-\nu)}{\tau} + \frac{(y-\nu)^2}{\tau^2} \right) \right).$$

Suppose that σ^2, τ^2, and ρ are known, and $\theta = (\mu, \nu) \in \mathbb{R}^2$.

(a) Find the score functions for μ and ν for $n = 1$.

(b) Find the information matrix for $\theta = (\mu, \nu)$ (for $n = 1$), and compute $I^{-1}(\theta)$.

(c) Find the efficient score function $l_1^* = l_\mu^*$ for μ with $\nu = \theta_2$ as a nuisance parameter.

(d) Find $I_{11\cdot 2}$ and relate this to $E_\theta(l_1^*)^2$ and the matrix $I^{-1}(\theta)$ computed in (b).

(e) What is the information and information bound for estimation of μ when $\nu = \nu_0$ is known?

(f) Show the connection between $\dot{l}_1 = \dot{l}_\mu$ and $l_1^* = l_\mu^*$ and $I_{11} = I_{\mu,\mu}$ and $I_{11\cdot 2} = I_{\mu\mu\cdot\nu}$ geometrically for small values of ρ and for values of ρ near 1.

(g) What is the MLE $(\hat{\mu}_n, \hat{\nu}_n)$ of (μ, ν) when both μ and ν are unknown? What is the MLE $\hat{\hat{\mu}}_n$ of μ when $\nu = \nu_0$ is known?

(h) Identify the (efficient) influence functions and give the resulting asymptotically linear representation of $\sqrt{n}((\hat{\mu}_n, \hat{\nu}_n) - (\mu, \nu))$ where $(\hat{\mu}_n, \hat{\nu}_n)$ is the MLE found in (g). What is the (asymptotic) linear representation of $\sqrt{n}(\hat{\hat{\mu}}_n - \mu)$, assuming that $\nu = \nu_0$ is known? What are the resulting limiting distributions in both cases?

Solution: (a) Now

$$\begin{aligned} \log p_\theta(x, y) &= -\frac{1}{2(1-\rho^2)} \left\{ \frac{(x-\mu)^2}{\sigma^2} - 2\rho \frac{(x-\mu)}{\sigma}\frac{(y-\nu)}{\tau} + \frac{(y-\nu)^2}{\tau^2} \right\} \\ &\quad - \log(2\pi) - \frac{1}{2}\log(\sigma^2\tau^2(1-\rho^2)), \end{aligned}$$

and therefore the scores $\dot{l}_\mu$ and $\dot{l}_\nu$ for μ and ν are given by:

$$\begin{aligned} \dot{l}_\mu(x, y) &= \frac{1}{2(1-\rho^2)} \left\{ 2\frac{(x-\mu)}{\sigma^2} - 2\frac{\rho}{\sigma}\frac{(y-\nu)}{\tau} \right\} \\ &= \frac{1}{1-\rho^2} \left\{ \frac{x-\mu}{\sigma^2} - \frac{\rho}{\sigma}\frac{y-\nu}{\tau} \right\} \\ \dot{l}_\nu(x, y) &= \frac{1}{1-\rho^2} \left\{ \frac{y-\nu}{\tau^2} - \frac{\rho}{\tau}\frac{x-\mu}{\sigma} \right\}. \end{aligned}$$

(b) Computing further derivatives of the scores computed in (a) yields:

$$\ddot{l}_{\mu,\mu}(x, y) = -\frac{1}{\sigma^2(1-\rho^2)},$$

$$\ddot{l}_{\nu,\nu}(x,y) = -\frac{1}{\tau^2(1-\rho^2)},$$
$$\ddot{l}_{\mu,\nu}(x,y) = \frac{\rho}{\sigma\tau(1-\rho^2)},$$

and therefore

$$I(\theta) = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma^2} & -\frac{\rho}{\sigma\tau} \\ -\frac{\rho}{\sigma\tau} & \frac{1}{\tau^2} \end{pmatrix}$$

The corresponding inverse inverse information matrix is:

$$\begin{aligned} I^{-1}(\theta) &= (1-\rho^2)\begin{pmatrix} \frac{1}{\tau^2} & \frac{\rho}{\sigma\tau} \\ \frac{\rho}{\sigma\tau} & \frac{1}{\sigma^2} \end{pmatrix}\frac{1}{\frac{1}{\sigma^2\tau^2}(1-\rho^2)} \\ &= \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix}. \end{aligned}$$

(c) The efficient score function l_1^* for estimation of μ with ν as a nuisance parameter is

$$\begin{aligned} \dot{l}_1^*(x,y) &= \dot{l}_1(x,y) - I_{12}I_{22}^{-1}\dot{l}_2(x,y) \\ &= \frac{1}{1-\rho^2}\left\{\frac{x-\mu}{\sigma^2} - \frac{\rho}{\sigma}\frac{y-\nu}{\tau} - \left(\frac{-\rho}{\sigma\tau}\right)\left(\frac{y-\nu}{\tau^2} - \frac{\rho}{\tau}\frac{x-\mu}{\sigma}\right)\right\} \\ &= \frac{1}{1-\rho^2}\left\{\left(\frac{x-\mu}{\sigma^2} - \rho^2\frac{x-\mu}{\sigma^2}\right) - \frac{\rho}{\sigma}\left(\frac{y-\nu}{\tau} - \frac{y-\nu}{\tau}\right)\right\} \\ &= \frac{x-\mu}{\sigma^2}. \end{aligned}$$

(d) Direct computation yields

$$\begin{aligned} I_{11\cdot 2} &= I_{11} - I_{12}I_{22}^{-1}I_{21} = \frac{1}{1-\rho^2}\left\{\frac{1}{\sigma^2} - \left(\frac{-\rho}{\sigma\tau}\right)\tau^2\left(\frac{-\rho}{\sigma\tau}\right)\right\} \\ &= \frac{1}{1-\rho^2}\frac{1}{\sigma^2}\{1-\rho^2\} = \frac{1}{\sigma^2}. \end{aligned}$$

Note that

$$I_{11\cdot 2} = E_\theta(l_1^*)^2 = E_\theta\left(\frac{X-\mu}{\sigma^2}\right)^2 = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2} = 1/I^{11}(\theta)$$

where

$$I^{-1}(\theta) = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix}$$

and $I^{11}(\theta) = I_{11\cdot 2}^{-1} = \sigma^2$.
(e) When $\nu = \nu_0$ is known, the information for μ is $I_{11} = \frac{1}{(1-\rho^2)\sigma^2}$, and the information bound for estimating μ is $I_{11}^{-1} = (1-\rho^2)\sigma^2$. Note that when ν is known, this means it should be possible to estimate μ with variance $(1-\rho^2)\sigma^2$ in this case, which is considerable smaller than σ^2 if ρ is close to ± 1.

(f) $l_1^* = l_\mu^*$ is the projection of $\dot{l}_1 = \dot{l}_\mu$ on the orthogonal complement $[\dot{l}_2]^\perp$ of the linear span of $\dot{l}_2$, $[\dot{l}_2]$: hence

$$\dot{l}_1 = (\dot{l}_1 - I_{12}I_{22}^{-1}\dot{l}_2) + I_{12}I_{22}^{-1}\dot{l}_2 \equiv l_1^* + I_{12}I_{22}^{-1}\dot{l}_2$$

where l_1^* and $I_{12}I_{22}^{-1}\dot{l}_2$ are orthogonal. It follows that

$$I_{11} = E_\theta \dot{l}_1^2 = E_\theta(l_1^*)^2 + I_{12}I_{22}^{-1}I_{21} = I_{11\cdot 2} + I_{12}^{-1}I_{22}^{-1}I_{21}$$

where $I_{11\cdot 2} \equiv I_{11} - I_{12}I_{22}^{-1}I_{21}$. When these terms are evaluated in this particular model, the above identity becomes:

$$\frac{1}{(1-\rho^2)\sigma^2} = \frac{1}{\sigma^2} + \frac{\rho^2}{\sigma^2(1-\rho^2)}.$$

When $\rho = 0$ the information for μ when ν is known, $I_{11} = 1/(\sigma^2(1-\rho^2))$ is the same as the first term on the right side, $I_{11\cdot 2}$ the information for μ when ν is unknown (and the second term on the right side is zero).
When ρ is close to 1, the information I_{11} for μ when ν is known becomes very large and approximately the same as the second term on the right side, while the first term on the right side, $I_{11\cdot 2}$, stays fixed at $1/\sigma^2$. See the figures on the last page of the solution set.
(g) When μ and ν are both unknown it is easily seen that $(\hat{\mu}_n, \hat{\nu}_n) = (\overline{X}_n, \overline{Y}_n)$. This is most easily seen in terms of the efficient score equations

$$0 = \sum_{i=1}^n l_\mu^*(X_i, Y_i) = \frac{1}{\sigma^2}\sum_1^n (X_i - \mu),$$
$$0 = \sum_{i=1}^n l_\nu^*(X_i, Y_i) = \frac{1}{\tau^2}\sum_{i=1}^n (Y_i - \nu),$$

which yield $\hat{\mu}_n = \overline{X}_n$ and $\hat{\nu}_n = \overline{Y}_n$.
When $\nu = \nu_0$ is known, $\hat{\hat{\mu}}_n$ is the solution of the (one) score equation

$$0 = \sum_{i=1}^n \dot{l}_\mu(X_i, Y_i) = \frac{1}{1-\rho^2}\left\{\sum_1^n \frac{X_i - \mu}{\sigma^2} - \frac{\rho}{\sigma}\sum_1^n \frac{Y_i - \nu_0}{\tau}\right\},$$

or, equivalently

$$0 = \overline{X} - \mu - \rho\frac{\sigma}{\tau}(\overline{Y}_n - \nu_0),$$

and this yields

$$\hat{\hat{\mu}}_n = \overline{X}_n - \rho\frac{\sigma}{\tau}(\overline{Y}_n - \nu) = \overline{X - \frac{\rho\sigma}{\tau}(Y - \nu_0)}.$$

Note that $Var_\theta(X - (\rho\sigma/\tau)(Y - \nu_0)) = \sigma^2(1-\rho^2) < \sigma^2$ if $\rho \neq 0$.
(h) From Theorem 4.1.2,

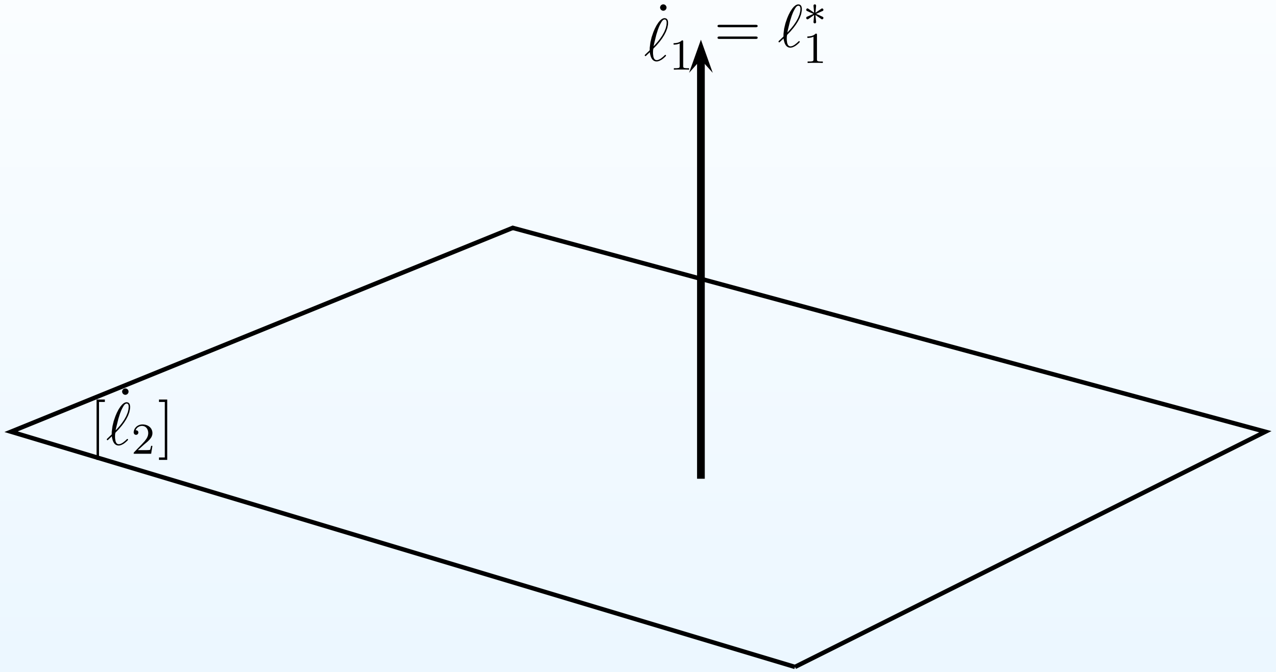
$$\sqrt{n}((\hat{\mu}_n, \hat{\nu}_n)^T - (\mu, \nu)^T) \to_d \underline{D} \sim N_2(0, I^{-1}(\theta))$$

where $I^{-1}(\theta)$ is as was computed in (b) above. In particular

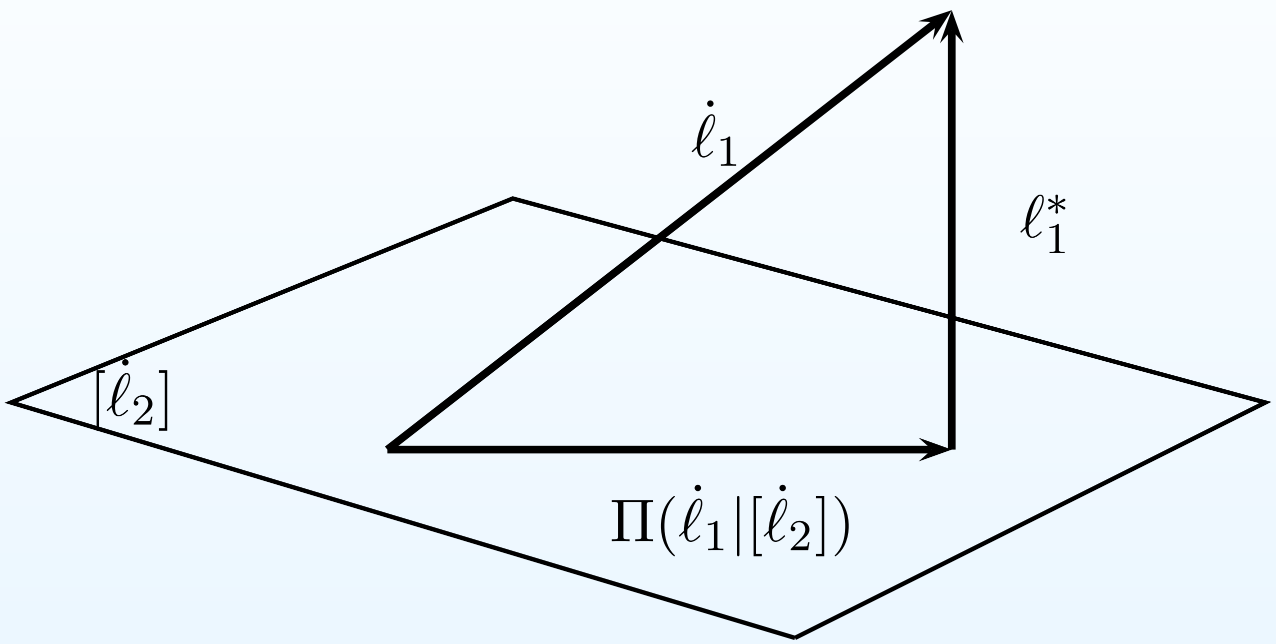
$$\sqrt{n}(\hat{\mu}_n - \mu) \to_d D_1 \sim N(0, \sigma^2).$$

Similarly, also from Theorem 4.1.2, when $\nu = \nu_0$ is known,

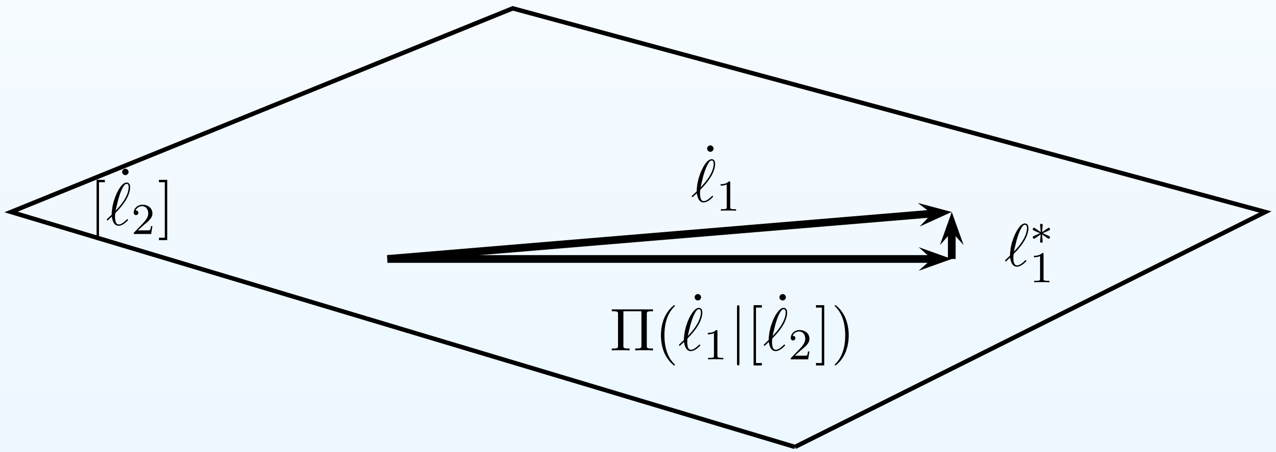
$$\sqrt{n}(\hat{\hat{\mu}}_n - \mu) \to_d I_{11}^{-1}Z_1 \sim N(0, (1-\rho^2)\sigma^2).$$



$$\rho = 0$$



ρ moderate



ρ close to 1