The Annals of Statistics
1974, Vol. 2, No. 2, 225–247

FISHER INFORMATION AND THE PITMAN ESTIMATOR OF A LOCATION PARAMETER[1]

BY SIDNEY C. PORT AND CHARLES J. STONE

University of California, Los Angeles

In this paper we consider estimation of the location parameter $\theta \in R^d$ based on a random sample from $(\theta + X, Y)$, where X is a d-dimensional random vector, Y is a random element of some measure space $\mathscr{Y}$, and (X, Y) has a known distribution. We first define the Fisher information $\mathscr{I}(\theta + X, Y)$ and the inverse information $\mathscr{I}^-(\theta + X, Y)$ under *no* regularity conditions. The properties of these quantities are investigated. Supposing that $E|X|^\delta < \infty$ for some $\delta > 0$ we show that for n sufficiently large the Pitman estimator $\hat{\theta}_n$ of θ based on a random sample of size n is well defined, unbiased, and its covariance, which is independent of θ, satisfies the inequality $n \operatorname{Cov} \hat{\theta}_n \geqq \mathscr{I}^-(\theta + X, Y)$. Moreover, $\lim_{n\to\infty} n \operatorname{Cov} \hat{\theta}_n = \mathscr{I}^-(\theta + X, Y)$ and $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ is asymptotically normal with mean zero and covariance $\mathscr{I}^-(\theta + X, Y)$.

1. Introduction. In this paper we consider estimation of the location parameter $\theta \in R^d$ based on a random sample from $(\theta + X, Y)$, where X is a d-dimensional random vector, Y is a random element of some measure space $\mathscr{Y}$ having distribution μ_Y, and (X, Y) has a known distribution. Such a model arises in the estimation of regression coefficients of a stochastic process.

In Section 2 we define the *Fisher information* $\mathscr{I}(\theta + X, Y)$ as a nonnegative function on R^d. To do so we let W be independent of (X, Y) and have the standard normal distribution on R^d. For $\sigma > 0$, $\mathscr{I}(\theta + X + \sigma W, Y)$ is defined classically as follows: Let $f_\sigma(x \mid y)$ denote the continuous version of the conditional density of $X + \sigma W$ given that $Y = y$, let $f_\sigma'(x \mid y)$ denote the derivative of $f_\sigma(x \mid y)$ with respect to x (that is, the vector of partial derivatives), and let (,) denote the usual inner product on R^d. Then

$$\mathscr{I}(\theta + X + \sigma W, Y)(e) = \int\int \frac{(f_\sigma'(x \mid y), e)^2}{f_\sigma(x \mid y)}\, dx \mu_Y(dy)\,, \qquad e \in R^d\,.$$

After showing that $\mathscr{I}(\theta + X + \sigma W, Y)$ is non-decreasing as $\sigma \to 0$, we define

$$\mathscr{I}(\theta + X, Y) = \lim_{\sigma\to 0} \mathscr{I}(\theta + X + \sigma W, Y)\,.$$

We set $\mathscr{I}(\theta + X) = \mathscr{I}(\theta + X, 0)$. If $d = 1$ then $\mathscr{I}(\theta + X)(e) = \mathscr{I}e$, $-\infty < e < \infty$, where the positive constant $\mathscr{I}$ agrees with the definition of Huber [3]. There Huber showed that

$$\mathscr{I} = \int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)}\right)^2 f(x)\, dx$$

Received October 1972; revised May 1973.

[1] Research supported by the National Science Foundation under grant GP-33431X.

AMS 1970 *subject classifications*. Primary 62F10; Secondary 62F20.

Key words and phrases. Location parameter, Fisher information, Cramér–Rao inequality, Pitman estimator, asymptotic normality.

if X has an absolutely continuous density f such that the indicated integral is finite and $\mathscr{I} = +\infty$ otherwise.

Set

$$V = \{e : \mathscr{I}(\theta + X, Y)(e) < \infty\} .$$

Then V is a subspace of R^d and there is a positive definite symmetric linear transformation $A : V \to V$ such that

$$\mathscr{I}(\theta + X, Y)(e) = (e, Ae) , \qquad e \in V .$$

In Section 3 we define the *inverse Fisher information* $\mathscr{I}^-(\theta + X, Y)$ as the nonnegative definite symmetric linear transformation from R^d to itself given by

$$\begin{aligned} \mathscr{I}^-(\theta + X, Y)e &= A^{-1}e , & e \in V , \\ &= 0 , & e \in V^\perp , \end{aligned}$$

where $V^\perp$ is the orthogonal complement of V. If $V = R^d$, we can write the inverse information as $\mathscr{I}^{-1}(\theta + X, Y)$. Various properties of inverse Fisher information are obtained in Section 3. In particular

$$\lim_{\sigma \to 0} \mathscr{I}^{-1}(\theta + X + \sigma W, Y) = \mathscr{I}^-(\theta + X, Y) .$$

In Section 4 we study estimators of θ based on samples from $(\sigma + X + \sigma W, Y)$. These results are used mainly as tools in Section 5 and in Stone [8].

Consider a random sample

$$(\theta + X_1, Y_1), \cdots, (\theta + X_n, Y_n)$$

of size n from $(\theta + X, Y)$. For $e \in R^d$ set $\mu = (e, \theta)$. If T_n is an estimator of μ based on this sample, its maximum mean square risk is defined as $\sup_\theta E_\theta(T_n - \mu)^2$. Let $M_n(e)$ denote the infimum of this quantity as T_n runs over all such estimators. If $M_n(e) < \infty$, the Pitman estimator P_n of μ exists, is unbiased and invariant (i.e., $P_n - \mu$ independent of θ), and has mean square risk $M_n(e)$. If T_n is an invariant estimator of μ based on the same sample and having finite mean square risk, then

$$P_n = T_n - E_0[T_n \mid X_i - X_1, Y_i, 1 \leqq i \leqq n] .$$

If $\tilde{\theta}_n$ is an estimator of θ based on the sample of size n from $(\theta + X, Y)$, its maximum total mean square risk is defined as $\sup_\theta E_\theta|\tilde{\theta}_n - \theta|^2$. Let M_n denote the infimum of this quantity as $\tilde{\theta}_n$ runs over all such estimators. If $M_n < \infty$, the Pitman estimator $\hat{\theta}_n$ of θ exists, is unbiased and invariant (i.e., $\hat{\theta}_n - \theta$ is independent of θ), and has total mean square risk M_n. If $\tilde{\theta}_n$ is any invariant estimator of θ based on the same sample and having finite total mean square risk, then

$$\hat{\theta}_n = \tilde{\theta}_n - E_0[\tilde{\theta}_n \mid X_i - X_1, Y_i, 1 \leqq i \leqq n] .$$

For discussions of Pitman estimators and their minimax properties in various levels of generality, see Pitman [6], Girshick and Savage [2], Blackwell and Girshick [1], Kudô [5], and Kiefer [4].

In Section 5 we show that $nM_n(e) \geqq (e, \mathscr{I}^-(\theta + X, Y)e)$ and consequently that $nM_n \geqq \text{trace}\, \mathscr{I}^-(\theta + X, Y)$. It follows that if $M_n < \infty$, then $n \operatorname{Cov} \hat{\theta}_n \geqq \mathscr{I}^-(\theta + X, Y)$. If

$$(1.1) \qquad E|X|^\delta < \infty \qquad \text{for some } \delta > 0 ,$$

then $M_n < \infty$ and the Pitman estimator $\hat{\theta}_n$ exists for n sufficiently large. In Theorem 5.2 we show that if (1.1) holds, then $\mathscr{L}(n^{\frac{1}{2}}(\hat{\theta}_n - \theta)) \to N(0, \mathscr{I}^-(\theta + X, Y))$ and $n \operatorname{Cov} \hat{\theta}_n \to \mathscr{I}^-(\theta + X, Y)$ as $n \to \infty$. Observe that the results of Theorem 5.2 are shown to hold without making any assumptions concerning existence or smoothness of densities. But the mild assumption (1.1) is necessary.

Further results on these topics are covered in [8].

2. Fisher information. In this section we will define and study the properties of the Fisher information $\mathscr{I}(\theta + X, Y)$ on θ contained in $(\theta + X, Y)$. Here θ is an unknown constant in R^d, X is a d-dimensional random vector, Y is a random element of some measure space $\mathscr{Y}$, and (X, Y) has a known distribution. The Fisher information $\mathscr{I}(\theta + X)$ of θ contained in $\theta + X$ is defined by setting $\mathscr{I}(\theta + X) = \mathscr{I}(\theta + X, 0)$. Let μ_Y denote the distribution of Y and let $F(\cdot \mid y)$ denote the (regular) conditional distribution of X given $Y = y$.

Let W be independent of (X, Y) and have the standard normal distribution on R^d. For $\sigma > 0$, σW has the density φ_σ given by

$$\varphi_\sigma(x) = \frac{1}{\sigma^d (2\pi)^{d/2}} \exp[-|x|^2/2\sigma^2] .$$

Let $F_\sigma(\cdot \mid y)$ denote the conditional distribution of $X + \sigma W$ given $Y = y$. Then $F_\sigma(\cdot \mid y)$ has the density

$$f_\sigma(x \mid y) = \int F(dz \mid y) \varphi_\sigma(x - z) .$$

This density is differentiable with respect to x and

$$f_\sigma'(x \mid y) = \int F(dz \mid y) \varphi_\sigma'(x - z) .$$

Here

$$\varphi_\sigma'(x) = \frac{-x}{\sigma^2} \varphi_\sigma(x) .$$

(By the derivative of a function on R^d, we mean the d-dimensional vector of its partial derivatives.) The distribution of $(\theta + X + \sigma W, Y)$ is absolutely continuous with respect to the product measure $dx\mu_y(dy)$ on $R^d \times \mathscr{Y}$ and has the density

$$f_\sigma(x, y; \theta) = f_\sigma(x - \theta \mid y) .$$

We define $\mathscr{I}(\theta + X + \sigma W, Y)$ to be the nonnegative function on R^d given by

$$(2.1) \qquad \mathscr{I}(\theta + X + \sigma W, Y)(e) = E\left[\frac{((\partial/\partial\theta) f_\sigma(X, Y; \theta), e)^2}{f_\sigma(X, Y; \theta)}\right] .$$

Then $\mathscr{I}(\theta + X + \sigma W, Y)$ is given independently of θ by

$$\mathscr{I}(\theta + X + \sigma W, Y)(e) = E\left[\frac{(f_\sigma'(X \mid Y), e)}{f_\sigma(X \mid Y)}\right]^2 = \iint \frac{(f_\sigma'(x \mid y), e)^2}{f_\sigma(x \mid y)} dx\mu_Y(dy) .$$

LEMMA 2.1. $\mathscr{I}(\theta + X + \sigma W, Y)(e) \leqq \sigma^{-2}|e|^2$.

PROOF. By applying Schwarz's inequality to

$$(f_\sigma'(x \mid y), e) = (\int F(dz \mid y)\varphi_\sigma'(x - z), e)$$
$$= \int F(dz \mid y)(\varphi_\sigma(x - z))^{\frac{1}{2}} \frac{(\varphi_\sigma'(x - z), e)}{(\varphi_\sigma(x - z))^{\frac{1}{2}}},$$

we conclude that

$$\frac{(f_\sigma'(x \mid y), e)^2}{f_\sigma(x \mid y)} \leqq \int F(dz \mid y) \frac{(\varphi_\sigma'(x - z), e)^2}{\varphi_\sigma(x - z)}.$$

By integrating on x and noting that

$$\int \frac{(\varphi_\sigma'(x), e)^2}{\varphi_\sigma(x)} dx = \sigma^{-2}|e|^2,$$

we see that the lemma is valid.

LEMMA 2.2. *There is a positive definite symmetric linear transformation $A(\sigma)$ from R^d onto itself such that*

$$\mathscr{I}(\theta + X + \sigma W, Y)(e) = (e, A(\sigma)e).$$

PROOF. The transformation $A(\sigma)$ is given more explicitly by

$$(e_1, A(\sigma)e_2) = \int\int \frac{(f_\sigma'(x \mid y), e_1)(f_\sigma'(x \mid y), e_2)}{f_\sigma(x \mid y)} dx \mu_Y(dy).$$

It is well defined by Lemma 2.1 and Schwarz's inequality and is easily seen to be positive definite.

LEMMA 2.3. *If Z is independent of (X, Y), then*

$$\mathscr{I}(\theta + X + \sigma W, Y, Z) = \mathscr{I}(\theta + X + \sigma W, Y).$$

PROOF. By hypotheses $F(\cdot \mid y, z) = F(\cdot \mid y)$. Thus $f_\sigma(x \mid y, z) = f_\sigma(x \mid y)$, from which the result follows.

Let $\mathscr{B}_Y$ be the σ-field of events determined by Y.

LEMMA 2.4. *If Z is a d-dimensional random vector that is $\mathscr{B}_Y$ measurable, then*

$$\mathscr{I}(\theta + X + Z + \sigma W, Y) = \mathscr{I}(\theta + X + \sigma W, Y).$$

PROOF. There is a measurable function $g: \mathscr{Y} \to R^d$ such that $z = g(Y)$. Thus the conditional density of $X + Z + \sigma W$ given $Y = y$ is just $f_\sigma(x - g(y) \mid y)$, from which the result follows.

The next result is close to results found in Section 5a.4 of Rao [7].

LEMMA 2.5. *Suppose $\mathscr{B}_Z \subseteqq \mathscr{B}_Y$. Then*

$$E\left[\frac{f_\sigma'(X \mid Y)}{f_\sigma(X \mid Y)}\middle| X, Z\right] = \frac{g_\sigma'(X \mid Z)}{g_\sigma(X \mid Z)}, \tag{2.2}$$

where $g_\sigma(\cdot\,|\,z)$ *is the conditional density of* $X + \sigma W$ *given* $Z = z$, *and*

$$\mathscr{I}(\theta + X + \sigma W, Z) \leqq \mathscr{I}(\theta + X + \sigma W, Y)\,. \tag{2.3}$$

PROOF. By assumption Z is a random element of some measure space $\mathscr{X}$. In order to prove (2.2) it suffices to show that if A is a measurable subset of $\mathscr{X}$ and $h: R^d \to R$ is continuously differentiable and has compact support, then

$$E\left[h(X)1_A(Z)\frac{f_\sigma'(X\,|\,Y)}{f_\sigma(X\,|\,Y)}\right] = E\left[h(X)1_A(Z)\frac{g_\sigma'(X\,|\,Z)}{g_\sigma(X\,|\,Z)}\right]. \tag{2.4}$$

The right side of (2.4) equals $-Eh'(X)1_A(Z)$. Since $B_Z \subseteqq B_Y$ there is a measurable subset B of Y such that $1_A(Z) = 1_B(Y)$. Using this observation, we conclude that the left side of (2.4) also equals $-Eh'(X)1_A(Z)$. Thus (2.4) holds and hence so does (2.2). It follows from (2.2) that

$$E\left[\frac{(g_\sigma'(X\,|\,Z), e)}{g_\sigma(X\,|\,Z)}\right]^2 = E\left[\frac{(g_\sigma'(X\,|\,Z), e)(f_\sigma'(X\,|\,Y), e)}{g_\sigma(X\,|Z)f_\sigma(X\,|\,Y)}\right].$$

Consequently

$$\begin{aligned} 0 &\leqq E\left[\frac{(g_\sigma'(X\,|\,Z), e)}{g_\sigma(X\,|\,Z)} - \frac{(f_\sigma(X\,|\,Y), e)}{f_\sigma(X\,|\,Y)}\right]^2 \\ &= \mathscr{I}(\theta + X + \sigma W, Z) + \mathscr{I}(\theta + X + \sigma W, Y) - 2\mathscr{I}(\theta + X + \sigma W, Z) \\ &= \mathscr{I}(\theta + X + \sigma W, Y) - \mathscr{I}(\theta + X + \sigma W, Z)\,, \end{aligned}$$

which completes the proof of (2.3).

THEOREM 2.1. $\mathscr{I}(\theta + X + \sigma W)(e)$ *is a continuous function of the distribution of* X.

PROOF. Let F denote the distribution of X and let f_σ denote the density of $X + \sigma W$. Then

$$f_\sigma(x) = \int \varphi_\sigma(x - z)F(dz)\,,$$

$$(f_\sigma'(x), e) = \int (\varphi_\sigma'(x - z), e)F(dz)\,,$$

and

$$\mathscr{I}(\theta + X + \sigma W) = \int \frac{(f_\sigma'(x), e)^2}{f_\sigma(x)}\,dx\,.$$

LEMMA 2.6. *For* $0 \leqq N < \infty$

$$\int \frac{(\int_{|z|\geqq N} (\varphi_\sigma'(x - z), e)F(dz))^2}{f_\sigma(x)}\,dx \leqq \sigma^{-2}|e|^2 \int_{|z|\geqq N} F(dz)\,.$$

PROOF. We can suppose that

$$\int_{|z|\geqq N} F(dz) > 0\,,$$

since otherwise the result is trivially true. It follows as in the proof of Lemma 2.1 that for any probability distribution G on R^d

$$\int \frac{(\int (\varphi_\sigma'(x - z), e)G(dz))^2}{\int \varphi_\sigma(x - z)G(dz)}\,dx \leqq \sigma^{-2}|e|^2\,.$$

The lemma follows easily by setting $G(dz) = F(dz)/\int_{|z|\geqq N} F(dz)$.

LEMMA 2.7. *For fixed $\sigma > 0$, $N \geqq 0$, and $\varepsilon > 0$, we can choose $M \geqq N$ independently of F such that*

$$\int_{|x| \geqq M} \frac{(\int_{|z| \leqq N} (\varphi_\sigma'(x-z), e) F(dz))^2}{f_\sigma(x)} \leqq \varepsilon |e|^2, \qquad e \in R^d.$$

PROOF. There are positive constants K and c such that

$$\varphi_\sigma(x) \leqq K e^{-c|x|^2}, \qquad x \in R^d,$$

and

$$|(\varphi_\sigma'(x), e)| \leqq K|e|\,|x| \varphi_\sigma(x), \qquad x \in R^d.$$

For $|x| \geqq N$

$$\left|\int_{|z| \leqq N} (\varphi_\sigma'(x-z), e) F(dz)\right| \leqq K|e|(|x| + N) f_\sigma(x).$$

Also

$$\begin{aligned} \left|\int_{|z| \leqq N} (\varphi_\sigma'(x-z), e) F(dz)\right| &\leqq K^2 |e| \int_{|z| \leqq N} |x - z| e^{-c|x-z|^2} F(dz) \\ &\leqq K^2 |e| (|x| + N) e^{-c(|x|-N)^2}. \end{aligned}$$

Consequently

$$\frac{|\int_{|z| \leqq N} (\varphi_\sigma'(x-z), e) F(dz)|^2}{f_\sigma(z)} \leqq K^3 |e|^2 (|x| + N)^2 e^{-c(|x|-N)^2}$$

from which the lemma follows easily.

LEMMA 2.8. *For $\sigma > 0$, $M \geqq 0$, and $e \in R^d$*

$$\int_{|x| \leqq M} \frac{(f_\sigma'(x), e)^2}{f_\sigma(x)} dx$$

is a continuous function of F.

PROOF. This is a direct consequence of the definition of weak convergence of distributions.

Theorem 2.1 follows easily from Lemmas 2.6—2.8.

COROLLARY 2.1. *Let X_n, X be d-dimensional random vectors and suppose that the conditional distribution of X_n given $Y = y$ converges weakly as $n \to \infty$ to the conditional distribution of X given $Y = y$ for almost all y* (w.r.t. μ_Y). *Then*

$$\lim_{n \to \infty} \mathscr{I}(\theta + X_n + \sigma W, Y) = \mathscr{I}(\theta + X + \sigma W, Y).$$

PROOF. This result follows easily from Lemma 2.1 and Theorem 2.1.

LEMMA 2.9. *For each e, $\mathscr{I}(\theta + X + \sigma W, Y)(e)$ is a continuous and non-increasing function of σ (equivalently, it is non-decreasing as σ decreases).*

PROOF. Let W_1 and W_2 be independent standard normal random variables on R^d such that (W_1, W_2) is independent of (X, Y). Let $0 < \sigma_0 < \sigma$. Then

$$\begin{aligned} \mathscr{I}(\theta + X + \sigma W, Y) &= \mathscr{I}(\theta + X + \sigma_0 W_1 + (\sigma^2 - \sigma_0^2)^{\frac{1}{2}} W_2, Y) \\ &\leqq \mathscr{I}(\theta + X + \sigma_0 W_1 + (\sigma^2 - \sigma_0^2)^{\frac{1}{2}} W_2, Y, W_2) \\ &= \mathscr{I}(\theta + X + \sigma_0 W_1, Y, W_2) \\ &= \mathscr{I}(\theta + X + \sigma_0 W_1, Y). \end{aligned}$$

This shows that $\mathscr{I}(\theta + X + \sigma W, Y)(e)$ is non-increasing in σ. Now the conditional distribution of $X + (\sigma^2 - \sigma_0^2)^{\frac{1}{2}} W_2$ given $Y = y$ is continuous in σ for $\sigma > \sigma_0$. Thus it follows from Corollary 2.1 that

$$\mathscr{I}(\theta + X + \sigma_0 W_1 + (\sigma^2 - \sigma_0^2)^{\frac{1}{2}} W_2, Y)(e)$$

is continuous in σ for $\sigma > \sigma_0$. This shows that $\mathscr{I}(\theta + X + \sigma W, Y)(e)$ is continuous in σ for $\sigma > \sigma_0$ and hence for all $\sigma > 0$.

It follows from Lemma 2.9 that

$$\lim_{\sigma \to 0} \mathscr{I}(\theta + X + \sigma_0 W + \sigma W, Y) = \mathscr{I}(\theta + X + \sigma_0 W, Y), \qquad \sigma_0 > 0. \tag{2.5}$$

Thus Lemma 2.9 allows to define in general the *Fisher information* $\mathscr{I}(\theta + X, Y)$ on θ contained in $(\theta + X, Y)$ as

$$\mathscr{I}(\theta + X, Y) = \lim_{\sigma \to 0} \mathscr{I}(\theta + X + \sigma W, Y). \tag{2.6}$$

We see from (2.5) that the new definition agrees with the definition given in (2.1) when the latter is applicable. Clearly $0 \leqq \mathscr{I}(\theta + X, Y) \leqq \infty$ and, by Lemma 2.2

$$\mathscr{I}(\theta + X, Y)(e) > 0, \qquad e \neq 0. \tag{2.7}$$

THEOREM 2.2. (i) *If Z is independent of (X, Y), then*

$$\mathscr{I}(\theta + X, Y, Z) = \mathscr{I}(\theta + X, Y).$$

(ii) *If Z is a d-dimensional random vector that is $\mathscr{B}_Y$ measurable, then*

$$\mathscr{I}(\theta + X + Z, Y) = \mathscr{I}(\theta + X, Y).$$

(iii) *If $\mathscr{B}_Z \subseteqq \mathscr{B}_Y$, then*

$$\mathscr{I}(\theta + X, Z) \leqq \mathscr{I}(\theta + X, Y).$$

PROOF. These results follow immediately from Lemmas 2.3—2.5.

THEOREM 2.3. *The set*

$$V = \{e : \mathscr{I}(\theta + X, Y)(e) < \infty\}$$

is a subspace of R^d, on which $\mathscr{I}(\theta + X, Y)$ is a positive definite quadratic form.

PROOF. Since

$$\begin{aligned}\mathscr{I}(\theta + X + \sigma W, Y)(a_1 e_1 + a_2 e_2) \\ \leqq 2a_1^2 \mathscr{I}(\theta + X + \sigma W, Y)(e_1) + 2a_2^2 \mathscr{I}(\theta + X + \sigma W, Y)(e_2),\end{aligned}$$

V is a subspace of R^d. The remainder of the theorem follows from Lemma 2.2.

THEOREM 2.4. *Let X_n, X be d-dimensional random vectors and suppose that the conditional distribution of X_n given $Y = y$ converges weakly as $n \to \infty$ to the conditional distribution of X given $Y = y$ for almost all y* (w.r.t. μ_Y). *Then*

$$\liminf_{n \to \infty} \mathscr{I}(\theta + X_n, Y) \geqq \mathscr{I}(\theta + X, Y).$$

PROOF. By Corollary 2.1

$$\liminf_{n\to\infty} \mathscr{I}(\theta + X_n, Y) \geqq \lim_{n\to\infty} \mathscr{I}(\theta + X_n + \sigma W, Y) = \mathscr{I}(\theta + X + \sigma W, Y) .$$

The desired result now follows by letting $\sigma \to 0$.

COROLLARY 2.2. *Let W_n, $n \geqq 1$, be any sequence of d-dimensional random vectors such that W_n is independent of (X, Y) and $W_n \to 0$ in probability as $n \to \infty$. Then*

$$\lim_{n\to\infty} \mathscr{I}(\theta + X + W_n, Y) = \mathscr{I}(\theta + X, Y) .$$

THEOREM 2.5. *Suppose $\mathscr{B}_{Y_n}$ is non-decreasing in n and $\mathscr{B}(Y_1, Y_2, \cdots) = \mathscr{B}_Y$. Then*

$$\lim_{n\to\infty} \mathscr{I}(\theta + X, Y_n) = \mathscr{I}(\theta + X, Y) .$$

PROOF. We start with a special case of the result.

LEMMA 2.10. *Suppose $\mathscr{B}_{Y_n}$ is non-decreasing in n and $\mathscr{B}(Y_1, Y_2, \cdots) = \mathscr{B}_Y$. Then*

$$\lim_{n\to\infty} \mathscr{I}(\theta + X + \sigma W, Y_n) = \mathscr{I}(\theta + X + \sigma W, Y) .$$

PROOF. Let $F_n(\cdot \mid y_n)$ denote the conditional distribution of X given $Y_n = y_n$ and set

$$f_{\sigma,n}(x \mid y_n) = \int \varphi_\sigma(x - z) F_n(dz \mid y_n) .$$

Now by a standard martingale argument $F_n(\cdot \mid Y_n)$ converges weakly to $F(\cdot \mid Y)$ as $n \to \infty$ with probability one. It follows from Theorem 2.1 that with probability one

$$\lim_{n\to\infty} \int \frac{(f'_{\sigma,n}(x \mid Y_n), e)^2}{f_{\sigma,n}(x \mid Y_n)} dx = \int \frac{(f_\sigma'(x \mid Y), e)^2}{f_\sigma(x \mid Y)} dx .$$

Thus by Lemma 2.1

$$\begin{aligned} \lim_{n\to\infty} \mathscr{I}(\theta + X + \sigma W, Y_n)(e) &= \lim_{n\to\infty} E \int \frac{(f'_{\sigma,n}(x \mid Y_n), e)^2}{f_{\sigma,n}(x \mid Y_n)} dx \\ &= E \int \frac{(f_\sigma'(x \mid Y), e)^2}{f_\sigma(x \mid Y)} dx \\ &= \mathscr{I}(\theta + X + \sigma W, Y) . \end{aligned}$$

This completes the proof of Lemma 2.10.

It follows from Lemma 2.10 that

$$\liminf_{n\to\infty} \mathscr{I}(\theta + X, Y_n) \geqq \lim_{n\to\infty} \mathscr{I}(\theta + X + \sigma W, Y_n) = \mathscr{I}(\theta + X + \sigma W, Y) .$$

By letting $\sigma \to 0$ we conclude that

$$\liminf_{n\to\infty} \mathscr{I}(\theta + X, Y_n) \geqq \mathscr{I}(\theta + X, Y) . \tag{2.8}$$

By hypothesis $\mathscr{B}_{Y_n} \subseteqq \mathscr{B}_Y$, so according to Theorem 2.2 (iii)

$$\mathscr{I}(\theta + X, Y_n) \leqq \mathscr{I}(\theta + X, Y) . \tag{2.9}$$

The conclusion of the theorem follows immediately from (2.8) and (2.9).

COROLLARY 2.3. *$\mathscr{I}(\theta + X, Y_1, \cdots, Y_n) \uparrow \mathscr{I}(\theta + X, Y_1, Y_2, \cdots)$ as $n \to \infty$.*

THEOREM 2.6. *Suppose X_n, X are d-dimensional random vectors, Z_n, Z are k-dimensional random vectors, and that the conditional distribution of (X_n, Z_n) given $Y = y$ converges weakly as $n \to \infty$ to that of (X, Z) given $Y = y$ for almost all y* (w.r.t. μ_Y). *Then*

$$\liminf_{n\to\infty} \mathscr{I}(\theta + X_n, Z_n, Y) \geqq \mathscr{I}(\theta + X, Z, Y) .$$

PROOF. We begin with a special case.

LEMMA 2.11. *Suppose X_n, X are d-dimensional random vectors, Z_n, Z are k-dimensional random vectors and that the distribution of (X_n, Z_n) converges weakly as $n \to \infty$ to that of (X, Z). Then*

$$\liminf_{n\to\infty} \mathscr{I}(\theta + X_n + \sigma W, Z_n) \geqq \mathscr{I}(\theta + X + \sigma W, Z) .$$

PROOF. It is easy to construct functions $\rho_i : R^k \to R^k$ having the following properties:

(i) each ρ_i takes on only countably many values;
(ii) if $i > j$ and $\rho_i(z_1) = \rho_i(z_2)$, then $\rho_j(z_1) = \rho_j(z_2)$;
(iii) $\lim_{i\to\infty} \rho_i(z) = z$ for all z;
(iv) $P(Z \in \partial\{z : \rho_i(z) = c\}) = 0$ for all i and c.

It follows from the hypotheses of Lemma 2.11 that if c is a possible value of $\rho_i(Z)$, then the conditional distribution of X_n given $\rho_i(Z_n) = c$ converges weakly as $n \to \infty$ to the conditional distribution of X given $\rho_i(Z) = c$. Thus by Corollary 2.1

$$\lim_{n\to\infty} \mathscr{I}(\theta + X_n + \sigma W, \rho_i(Z_n)) = \mathscr{I}(\theta + X + \sigma W, \rho_i(Z)) .$$

Consequently by Lemma 2.5

$$\begin{aligned}\liminf_{n\to\infty} \mathscr{I}(\theta + X_n + \sigma W, Z_n) &\geqq \lim_{n\to\infty} \mathscr{I}(\theta + X + \sigma W, \rho_i(Z_n)) \\ &= \mathscr{I}(\theta + X + \sigma W, \rho_i(Z)) .\end{aligned}$$

Since

$$\lim_{i\to\infty} \mathscr{I}(\theta + X + \sigma W, \rho_i(Z)) = \mathscr{I}(\theta + X + \sigma W, Z)$$

by Theorem 2.5, the lemma is valid.

LEMMA 2.12. *Under the hypothesis of Theorem* 2.6

$$\liminf_{n\to\infty} \mathscr{I}(\theta + X_n + \sigma W, Z_n, Y) \geqq \mathscr{I}(\theta + X + \sigma W, Z, Y) .$$

PROOF. Let $\mathscr{I}(\theta + X + \sigma W, Z \mid Y = y)$ denote the information in $(\theta + X + \sigma W, Z)$ when (Y, Z) is chosen independently of W according to its conditional distribution given $Y = y$. Then by Lemma 2.11

$$\liminf_{n\to\infty} \mathscr{I}(\theta + X_n + \sigma W, Z_n \mid Y = y) \geqq \mathscr{I}(\theta + X + \sigma W, Z \mid Y = y)$$

for almost all y (w.r.t. μ_Y). Since

$$\mathscr{I}(\theta + X_n + \sigma W, Z_n, Y) = \int \mathscr{I}(\theta + X_n + \sigma W, Z_n \mid Y = y)\mu_Y(dy)$$

and

$$\mathscr{I}(\theta + X + \sigma W, Z, Y) = \int \mathscr{I}(\theta + X + \sigma W, Z \mid Y = y)\mu_Y(dy)\,,$$

the conclusion of the lemma follows from Fatou's lemma.

Theorem 2.6 follows easily from Lemma 2.12.

THEOREM 2.7. (i) *If* $c \neq 0$, *then* $\mathscr{I}(\theta + cX, Y)(ce) = \mathscr{I}(\theta + X, Y)(e)$.
(ii) *If* $Q\colon R^d \to R^d$ *is an orthogonal linear transformation, then*

$$\mathscr{I}(\theta + QX, Y)(Qe) = \mathscr{I}(\theta + X, Y)(e)\,.$$

PROOF. We will prove (ii), the proof of (i) being similar. It suffices to show that

$$\mathscr{I}(\theta + QX + \sigma W, Y)(Qe) = \mathscr{I}(\theta + X + \sigma W, Y)(e)\,. \tag{2.10}$$

Let $g_\sigma(\cdot \mid y)$ be the conditional density of $QX + \sigma W$ given $Y = y$. Then $g_\sigma(\cdot \mid y)$ is also the conditional density of $Q(X + \sigma W)$ given $Y = y$. Consequently

$$g_\sigma(x \mid y) = f_\sigma(Q^{-1}x \mid y)$$

and

$$g_\sigma'(x \mid y) = (Q^{-1})'f_\sigma'(Q^{-1}x \mid y)\,,$$

where $(Q^{-1})'$ is the transpose of Q^{-1}. Therefore

$$\begin{aligned}
\mathscr{I}(\theta + QX + \sigma W, Y)(Qe) &= \iint \frac{(g_\sigma'(x \mid y), Qe)^2}{g_\sigma(x \mid y)}\, dx\mu_Y(dy) \\
&= \iint \frac{((Q^{-1})'f_\sigma'(Q^{-1}x \mid y, Qe)^2}{f_\sigma(Q^{-1}x \mid y)}\, dx\mu_Y(dy) \\
&= \iint \frac{((Q^{-1})'f_\sigma'(x \mid y), Qe)^2}{f_\sigma(x \mid y)}\, dx\mu_Y(dy) \\
&= \iint \frac{(f_\sigma'(x \mid y), e)^2}{f_\sigma(x \mid y)}\, dx\mu_Y(dy) \\
&= \mathscr{I}(\theta + X + \sigma W, Y)(e)\,.
\end{aligned}$$

This verifies (2.10), which completes the proof of the theorem.

Consider a random vector X in R^d having a density f which is differentiable in the sense that there is an R^d-valued function f' such that

$$\int \left|\frac{f'(x)}{f(x)}\right|^2 f(x)\, dx < \infty \tag{2.11}$$

and

$$\int f'(x)\psi(x)\, dx = -\int f(x)\psi'(x)\, dx\,, \qquad \psi \in C_c^1\,. \tag{2.12}$$

Here C_c^1 denotes the continuously differentiable functions on R^d having compact support. Note that (2.11) implies that $|f'|$ is integrable on R^d and hence integrable on compacts. If $d = 1$, this last fact and (2.12) are together equivalent to the fact that f is absolutely continuous. For general d, (2.12) certainly holds if f is continuously differentiable. When (2.11) and (2.12) hold the "classical"

Fisher information $\mathscr{I}_C(\theta + X)$ is defined as

$$\mathscr{I}_C(\theta + X)(e) = \int \left(\frac{(f'(x), e)}{f(x)}\right)^2 f(x)\, dx\,, \qquad e \in R^d\,.$$

We will show that $\mathscr{I}(\theta + X) = \mathscr{I}_C(\theta + X)$ in this case. To do so we first show that in general $\mathscr{I}(\theta + X)$ agrees with the definition of Huber [3] for $d = 1$ and is a natural extension of Huber's definition for $d > 1$.

THEOREM 2.8. *For* $e \in R^d$

$$\mathscr{I}(\theta + X)(e) = \sup_\psi \frac{[E(\psi'(X), e)]^2}{E\psi^2(X)}\,, \tag{2.13}$$

where the sup *extends over all* $\psi \in C_c^1$ *such that* $E\psi^2(X) > 0$.

PROOF. We observe first that if X has a density f such that (2.11) and (2.12) hold, then

$$\mathscr{I}_C(\theta + X)(e) = \sup_\psi \frac{[E(\psi'(X), e)]^2}{E\psi^2(X)}\,. \tag{2.14}$$

That (2.14) holds with equality replaced by "$\geqq$" follows from Schwarz's inequality. To obtain the desired equality we choose $\psi \in C_c^1$ to approximate $(f'(\cdot), e)/f(\cdot)$ in $\mathscr{L}^2(\text{dist}\ X)$.

It follows from (2.14) that

$$\begin{aligned}\mathscr{I}(\theta + X + \sigma W)(e) &= \mathscr{I}_C(\theta + X + \sigma W)(e) \\ &= \sup_\psi \frac{[E(\psi'(X + \sigma W), e)]^2}{E\psi^2(X + \sigma W)}\,.\end{aligned} \tag{2.15}$$

For fixed e and ψ

$$\lim_{\sigma\to 0} E(\psi'(X + \sigma W), e) = E(\psi'(X), e)$$

and

$$\lim_{\sigma\to 0} E\psi^2(X + \sigma W) = E\psi^2(X)\,.$$

Since

$$\lim_{\sigma\to 0} \mathscr{I}(\theta + X + \sigma W)(e) = \mathscr{I}(\theta + X)(e)\,,$$

we see that

$$\mathscr{I}(\theta + X)(e) \geqq \sup_\psi \frac{[E(\psi'(X), e)]^2}{E\psi^2(X)}\,. \tag{2.16}$$

On the other hand, by a convexity argument as in Huber [3, page 86–7], we conclude that

$$\frac{[E(\psi'(X + \sigma W), e)]^2}{E\psi^2(X + \sigma W)} \leqq \sup_\psi \frac{[E(\psi'(X), e)]^2}{E\psi^2(X)}\,.$$

Consequently by (2.15)

$$\mathscr{I}(\theta + X + \sigma W)(e) \leqq \sup_\psi \frac{[E(\psi'(X), e)]^2}{E\psi^2(X)}\,.$$

By letting $\sigma \to 0$ we conclude that

$$\mathcal{I}(\theta + X)(e) \leqq \sup_\psi \frac{[E(\psi'(X), e)]^2}{E\psi^2(X)}. \tag{2.17}$$

The conclusion of the theorem follows from (2.16) and (2.17).

COROLLARY 2.4. *Suppose that* X *has a density* f *such that* (2.11) *and* (2.12) *hold. Then* $\mathcal{I}(\theta + X) = \mathcal{I}_C(\theta + X)$.

PROOF. This result follows immediately from (2.13) and (2.14).

Suppose that for almost all y (w.r.t. μ_Y) the conditional distribution $F(\cdot \mid y)$ of X given that $Y = y$ has a density $f(\cdot \mid y)$ that satisfies (2.11) and (2.12). Then the "classical" Fisher information $\mathcal{I}_C(\theta + X, Y)$ is defined as

$$\mathcal{I}_C(\theta + X, Y)(e) = \iint \left(\frac{(f'(x \mid y), e)}{f(x \mid y)}\right)^2 f(x \mid y)\, dx \mu_Y(dy). \tag{2.18}$$

By conditioning on Y and applying Corollary 2.4, we obtain the next result.

COROLLARY 2.5. *If the conditions stated just prior to* (2.18) *hold, then* $\mathcal{I}(\theta + X, Y) = \mathcal{I}_C(\theta + X, Y)$.

It might appear that our definition of information

$$\mathcal{I}(\theta + X, Y) = \lim_{\sigma \to 0} \mathcal{I}_C(\theta + X + \sigma W, Y)$$

is arbitrary in that it depends on W having a standard normal distribution. This is not really true. For let W_n be d-dimensional random vectors which converge to zero in probability as $n \to \infty$ and are independent of (X, Y). Suppose that $(X + W_n, Y)$ satisfy the assumptions of Corollary 2.5. We could attempt to define a "new" measure of information

$$\mathcal{I}_N(\theta + X, Y) = \lim_{n \to \infty} \mathcal{I}_C(\theta + X + W_n, Y).$$

But it follows from Corollary 2.5 that

$$\mathcal{I}_C(\theta + X + W_n, Y) = \mathcal{I}(\theta + X + W_n, Y).$$

Thus by Corollary 2.2

$$\begin{aligned} \mathcal{I}_N(\theta + X, Y) &= \lim_{n \to \infty} \mathcal{I}_C(\theta + X + W_n, Y) \\ &= \lim_{n \to \infty} \mathcal{I}(\theta + X + W_n, Y) \\ &= \mathcal{I}(\theta + X, Y). \end{aligned}$$

We now wish to define the Fisher information $\mathcal{I}(\theta + X_1, Y_1, \cdots, \theta + Y_n, Y_n)$ of θ contained in $(\theta + X_1, Y_1, \cdots, \theta + X_n, Y_n)$. To do so let $u(x_1, y_1, \cdots, x_n, y_n)$ be an *invariant* function i.e. a measurable R^d-valued function such that for all θ

$$u(\theta + x_1, y_1, \cdots, \theta + x_n, y_n) = \theta + u(x_1, y_1, \cdots, x_n, y_n)$$

(for example, $u(x_1, y_1, \cdots, x_n, y_n) = x_1$). Set

$$U_n = u(X_1, Y_1, \cdots, X_n, Y_n).$$

Then

$$u(\theta + X_1, Y_1, \cdots, \theta + X_n, Y_n) = \theta + U_n .$$

We define

$$\begin{aligned}\mathscr{I}(\theta + X_1, Y_1, \cdots, \theta + X_n, Y_n) \\ = \mathscr{I}(\theta + U_n, X_1 - U_n, Y_1, \cdots, X_n - U_n, Y_n) .\end{aligned}$$

This is a reasonable definition since the observations

$$(\theta + X_1, Y_1, \cdots, \theta + X_n, Y_n)$$

and

$$(\theta + U_n, X_1 - U_n, Y_1, \cdots, X_n - U_n, Y_n)$$

are equivalent to each other. Moreover, it follows easily from Theorem 2.2 that the definition is independent of the choice of the invariant function u (for $u(X_1 - V_n, Y_1, \cdots, X_n - V_n, Y_n) = U_n - V_n$).

THEOREM 2.9. *If* (X_i, Y_i), $1 \leqq i \leqq n$, *are independent and each* X_i *is a random d-dimensional vector, then*

$$\mathscr{I}(\theta + X_1, Y_1, \cdots, \theta + X_n, Y_n) = \sum_{i=1}^n \mathscr{I}(\theta + X_i, Y_i) .$$

PROOF. Let W_i, $1 \leqq i \leqq n$, be independent standard normal random vectors on R^d such that W_i, $1 \leqq i \leqq n$, is independent of (X_i, Y_i), $1 \leqq i \leqq n$. Then

$$\begin{aligned}&\mathscr{I}_C(\theta + X_1 + \sigma W_1, Y_1, X_2 - X_1 \\ &\quad + \sigma W_2 - \sigma W_1, Y_2, \cdots, X_n - X_1 + \sigma W_n - \sigma W_1, Y_n) \\ &\quad = \sum_{i=1}^n \mathscr{I}_C(\theta + X_i + \sigma W_i, Y_i) ,\end{aligned} \tag{2.19}$$

since both sides of (2.19) equal

$$\mathscr{I}_C(\theta + X_1 + \sigma W_1, Y_1, \cdots, \theta + X_n + \sigma W_n, Y_n) ,$$

where the last quantity is defined "classically". It follows from (2.19) and Corollary 2.5 that

$$\begin{aligned}&\mathscr{I}(\theta + X_1 + \sigma W_1, Y_1, X_2 - X_1 \\ &\quad + \sigma W_2 - \sigma W_1, Y_2, \cdots, X_n - X_1 + \sigma W_n - \sigma W_1, Y_n) \\ &\quad = \sum_{i=1}^n \mathscr{I}(\theta + X_i + \sigma W_i, Y_i) .\end{aligned} \tag{2.20}$$

The desired result follows from (2.20), Theorem 2.2 and Theorem 2.6.

3. Inverse Fisher information. According to Theorem 2.3

$$V = \{e : \mathscr{I}(\theta + X, Y)(e) < \infty\}$$

is a subspace of R^d, on which $\mathscr{I}(\theta + X, Y)$ is a positive definite quadratic form. Thus there is a positive definite symmetric linear transformation $A : V \to V$ such that

$$\mathscr{I}(\theta + X, Y)(e) = (e, Ae) , \qquad e \in V .$$

The *inverse Fisher information* $\mathscr{I}^-(\theta + X, Y)$ on θ contained in $(\theta + X, Y)$ is defined to be the nonnegative definite symmetric linear transformation defined by

$$\mathscr{I}^{-}(\theta + X, Y)e = A^{-1}e\,, \qquad e \in V\,,$$
$$= 0\,, \qquad e \in V^{\perp}\,,$$

where $V^{\perp}$ is the orthogonal complement of V. Note that if $V = R^d$, then A is a positive definite symmetric linear transformation from R^d to itself,

$$\mathscr{I}(\theta + X, Y)(e) = (e, Ae)\,, \qquad e \in R^d\,,$$

and

$$\mathscr{I}^{-}(\theta + X, Y) = A^{-1}\,.$$

In this case we can write $\mathscr{I}^{-}(\theta + X, Y)$ as $\mathscr{I}^{-1}(\theta + X, Y)$. If $d = 1$, we can think of $\mathscr{I}^{-}(\theta + X, Y)$ as a nonnegative number.

If A and B are nonnegative definite symmetric linear transformations we write $A \leqq B$ to signify that $B - A$ is nonnegative definite. If $A \leqq B$ and A and B are invertible, then $B^{-1} \leqq A^{-1}$ (see Exercise 9 on page 56 of Rao [7]).

Let $A(\sigma)$ be as in Lemma 2.2. It follows from Lemma 2.9 that $A(\sigma)$ is non-increasing in σ, i.e. that if $\sigma_2 \geqq \sigma_1$, then $A(\sigma_2) \leqq A(\sigma_1)$. Thus

$$\mathscr{I}^{-1}(\theta + X + \sigma W, Y) = A^{-1}(\sigma)$$

is non-increasing in σ. This implies that

$$\lim_{\sigma\to 0} \mathscr{I}^{-1}(\theta + X + \sigma W, Y) = \lim_{\sigma\to 0} A^{-1}(\sigma)$$

exists as a nonnegative definite symmetric linear transformation.

THEOREM 3.1. $\mathscr{I}^{-}(\theta + X, Y) = \lim_{\sigma\to 0} \mathscr{I}^{-1}(\theta + X + \sigma W, Y)$.

PROOF. Set

$$A^{-} = \lim_{\sigma\to 0} A^{-1}(\sigma)\,.$$

Let U denote the nullspace of A^{-} and $U^{\perp}$ the orthogonal complement of U. It is easily seen that A^{-} maps $U^{\perp}$ onto itself. Thus, when restricted to $U^{\perp}$, A^{-} is positive definite and symmetric. Since, for $\lambda > 0$,

$$(A^{-} + \lambda I)^{-1}e = \lambda^{-1}e\,, \qquad e \in U\,,$$

we see that

$$\lim_{\lambda\to\infty} (e, (A^{-} + \lambda I)^{-1}e) = \infty\,, \qquad e \in U \text{ and } e \neq 0\,. \tag{3.1}$$

For $\lambda > 0$, $(A^{-} + \lambda I)$ maps $U^{\perp}$ onto itself. Thus $(A^{-} + \lambda I)^{-1}$ maps $U^{\perp}$ onto itself. Let $B\colon U^{\perp} \to U^{\perp}$ denote the inverse of the transformation $A^{-}\colon U^{\perp} \to U^{\perp}$. Then B is positive definite and symmetric. Moreover

$$\lim_{\lambda\to 0} (A^{-} + \lambda I)^{-1}e = Be\,, \qquad e \in U^{\perp}\,.$$

Thus

$$\lim_{\lambda\to 0} (e, (A^{-} + \lambda I)^{-1}e) = (e, Be) < \infty\,, \qquad e \in U^{\perp}\,. \tag{3.2}$$

Now $(e, (A^{-1}(\sigma) + \lambda I)^{-1}e)$ is monotonic in both σ and λ, so that

$$\begin{aligned}\mathscr{I}(\theta + X, Y)(e) &= \lim_{\sigma\to 0} (e, A(\sigma)e) \\ &= \lim_{\sigma\to 0} [\lim_{\lambda\to 0} (e, (A^{-1}(\sigma) + \lambda I)^{-1}e)] \\ &= \lim_{\lambda\to 0} [\lim_{\sigma\to 0} (e, (A^{-1}(\sigma) + \lambda I)^{-1}e)] \\ &= \lim_{\lambda\to 0} (e, (A^{-} + \lambda I)^{-1}e) .\end{aligned}$$

Therefore (3.1) and (3.2) imply

$$\mathscr{I}(\theta + X, Y)(e) = \infty , \qquad e \in U \text{ and } e \neq 0 , \tag{3.3}$$

and

$$\mathscr{I}(\theta + X, Y)(e) = (e, Be) < \infty , \qquad e \in U^{\perp} . \tag{3.4}$$

It follows easily from (3.3) and (3.4) that $V = U^{\perp}$ and hence that $A = B$. Consequently A^{-1} is the restriction of A^{-} to V. Thus $\mathscr{I}^{-}(\theta + X, Y) = A^{-}$, which completes the proof of the theorem.

COROLLARY 3.1. (i) $\mathscr{I}^{-}(\theta + cX, Y) = c^2\mathscr{I}^{-}(\theta + X, Y)$. (ii) *If $Q : R^d$ is an orthogonal linear transformation, then*

$$(Qe, \mathscr{I}^{-}(\theta + QX, Y)Qe) = (e, \mathscr{I}^{-}(\theta + X, Y)e) .$$

THEOREM 3.2. $\mathscr{I}(\theta + X_1, Y_1) \geqq \mathscr{I}(\theta + X_2, Y_2)$ *if and only if* $\mathscr{I}^{-}(\theta + X_1, Y_1) \leqq \mathscr{I}^{-}(\theta + X_2, Y_2)$.

PROOF. Set $A_i^{-} = \mathscr{I}^{-}(\theta + X_i, Y_i)$. If $A_1^{-} \leqq A_2^{-}$, then $A_1^{-} + \lambda I \leqq A_2^{-} + \lambda I$ for $\lambda > 0$ and hence $(A_1^{-} + \lambda I)^{-1} \geqq (A_2^{-} + \lambda I)^{-1}$; thus (see the proof of Theorem 3.1)

$$\begin{aligned}\mathscr{I}(\theta + X_1, Y_1)(e) &= \lim_{\lambda\to 0} (e, (A_1^{-} + \lambda I)^{-1}e) \\ &\geqq \lim_{\lambda\to 0} (e, (A_2^{-} + \lambda I)^{-1}e) \\ &= \mathscr{I}(\theta + X_2, Y_2)(e) .\end{aligned}$$

Suppose conversely that $\mathscr{I}(\theta + X_1, Y_1) \geqq \mathscr{I}(\theta + X_2, Y_2)$. Set

$$V_i = \{e : \mathscr{I}(\theta + X_i, Y_i)(e) < \infty\} .$$

Then $V_1 \subseteqq V_2$ and hence $V_1^{\perp} \supseteqq V_2^{\perp}$. Clearly $\mathscr{I}^{-}(\theta + X_1, Y_1) = \mathscr{I}^{-}(\theta + X_2, Y_2) = 0$ on $V_2^{\perp}$, so it suffices to show that $\mathscr{I}^{-}(\theta + X_1, Y_1) \leqq \mathscr{I}^{-}(\theta + X_2, Y_2)$ on V_2. In doing so we can, with no loss of generality, assume that $V_2 = R^d$.

Let $B_1 : V_1 \to V_1$ be the positive definite symmetric linear transformation such that

$$\mathscr{I}(\theta + X_1, Y_1)(e_1) = (e_1, B_1 e_1) , \qquad e_1 \in V_1 .$$

Let $B_2 : R^d \to R^d$ be the positive definite symmetric linear transformation such that

$$\mathscr{I}(\theta + X_2, Y_2)(e) = (e, B_2 e) \qquad e \in R^d .$$

We know that

$$(e_1, B_1 e_1) \geqq (e_1, B_2 e_1) , \qquad e_1 \in V_1 .$$

To complete the proof of the theorem we need to show that

$$(e_1, B_1^{-1}e_1) \leqq (e_1 + e_2, B_2^{-1}(e_1 + e_2)) , \qquad e_1 \in V_1 \text{ and } e_2 \in V_1^{\perp} . \tag{3.5}$$

But (3.5) follows immediately from the following result on matrices.

LEMMA 3.1. *Consider the partitioned positive definite symmetric matrix*

$$B_2 = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} .$$

If $B_1 \geqq B_{22}$, *then*

$$\begin{pmatrix} 0 & 0 \\ 0 & B_1^{-1} \end{pmatrix} \leqq B_2^{-1} . \tag{3.6}$$

PROOF. Let $C = B_2^{-1}$ be partitioned as is B_2. Then $0 \leqq B_{22}^{-1} - B_1^{-1} = C_{22} - B_1^{-1} - C_{21}C_{11}^{-1}C_{12}$. Thus

$$0 \leqq \begin{pmatrix} I & 0 \\ C_{21}C_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} C_{11} & 0 \\ 0 & C_{22} - B_1^{-1} - C_{21}C_{11}^{-1}C_{12} \end{pmatrix} \begin{pmatrix} I & C_{11}^{-1}C_{12} \\ 0 & I \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} - B_1^{-1} \end{pmatrix} ,$$

from which (3.6) follows.

For $0 < d_1 < d$, let θ_1 denote the first d_1 coordinates of θ, θ_2 the last $d - d_1$ coordinates of θ, and write $\theta = (\theta_1, \theta_2)$. Similarly we write $X = (X_1, X_2)$ and $e = (e_1, e_2)$. It is straightforward to show that

$$\mathscr{I}(\theta_1 + X_1, X_2, Y)(e_1) = \mathscr{I}(\theta + X, Y)(e_1, 0) \tag{3.7}$$

and consequently that

$$\mathscr{I}(\theta_1 + X_1, Y)(e_1) \leqq \mathscr{I}(\theta + X, Y)(e_1, 0) . \tag{3.8}$$

The next theorem gives the corresponding inequality for inverse Fisher information.

THEOREM 3.3. *For* $e_1 \in R^{d_1}$

$$(e_1, \mathscr{I}^{-}(\theta_1 + X_1, Y)e_1) \leqq ((e_1, 0), \mathscr{I}^{-}(\theta + X, Y)(e_1, 0)) . \tag{3.9}$$

PROOF. It follows from (iv) (b) of page 270 Rao [7] that

$$\mathscr{I}(\theta_1 + X_1 + \sigma W_1, Y)(e_1) \leqq \mathscr{I}(\theta + X + \sigma W, Y)(e_1, e_2) . \tag{3.10}$$

Let $A(\sigma)$ and $B(\sigma)$ denote the positive definite transformations corresponding to $\mathscr{I}(\theta + X + \sigma W, Y)$ and $\mathscr{I}(\theta_1 + X_1 + \sigma W_1, Y)$ respectively. Then (3.10) can be rewritten as

$$\begin{pmatrix} B(\sigma) & 0 \\ 0 & 0 \end{pmatrix} \leqq \begin{pmatrix} A(\sigma)_{11} & A(\sigma)_{12} \\ A(\sigma)_{21} & A(\sigma)_{22} \end{pmatrix} . \tag{3.11}$$

It follows from (3.11) that $B(\sigma) \leqq A(\sigma)_{11}$ and hence that $B^{-1}(\sigma) \geqq A^{-1}(\sigma)_{11}$. Letting $\sigma \to 0$ we see that $B^{-1} \geqq (A^{-})_{11}$ as desired.

THEOREM 3.4. *If*

$$\liminf_{n\to\infty} \mathscr{I}(\theta + X_n + \sigma W, Y_n) \geqq \mathscr{I}(\theta + X + \sigma W, Y)$$

for all $\sigma > 0$, *then*

$$\limsup_{n\to\infty} (e, \mathscr{I}^-(\theta + X_n, Y_n)e) \leqq (e, \mathscr{I}^-(\theta + X, Y)e), \qquad e \in R^d.$$

PROOF. Let $\sigma > 0$ be fixed. Choose $\lambda > 0$ such that $\mathscr{I}(\theta + X + \sigma W, Y)(e) \geqq 2\lambda|e|^2$ for all e.

For n sufficiently large

$$\mathscr{I}(\theta + X_n + \sigma W, Y_n)(e) \geqq \mathscr{I}(\theta + X + \sigma W, Y)(e) - \lambda|e|^2, \qquad e \in R^d. \tag{3.12}$$

(Here Lemma 2.1 is used to reduce (3.12) to an inequality involving only finite many e's.) It follows from (3.12) that for n sufficiently large

$$(e, \mathscr{I}^{-1}(\theta + X_n + \sigma W, Y_n)e) \leqq (e, (A(\sigma) - \lambda I)^{-1}e), \qquad e \in R^d.$$

Thus

$$\limsup_{n\to\infty} (e, \mathscr{I}^{-1}(\theta + X_n + \sigma W, Y_n)e) \leqq (e, (A(\sigma) - \lambda I)^{-1}e), \qquad e \in R^d. \tag{3.13}$$

Since

$$\lim_{\lambda\to 0} (A(\sigma) - \lambda I)^{-1} = A^{-1}(\sigma) = \mathscr{I}^{-1}(\theta + X + \sigma W, Y),$$

we can let $\lambda \to 0$ in (3.13) to conclude that

$$\limsup_{n\to\infty} (e, \mathscr{I}^{-1}(\theta + X_n + \sigma W, Y_n)e) \leqq (e, \mathscr{I}^{-1}(\theta + X + \sigma W, Y)e).$$

Since

$$\mathscr{I}^-(\theta + X_n, Y_n) \leqq \mathscr{I}^{-1}(\theta + X_n + \sigma W, Y_n)$$

it follows that

$$\limsup_{n\to\infty} (e, \mathscr{I}^-(\theta + X_n, Y_n)e) \leqq (e, \mathscr{I}^{-1}(\theta + X + \sigma W, Y)e).$$

The desired conclusion now follows from Theorem 3.1.

COROLLARY 3.2. $\mathscr{I}^-(\theta + X, Y_1, \cdots, Y_n) \downarrow \mathscr{I}^-(\theta + X, Y_1, Y_2, \cdots)$ *as* $n \to \infty$.

It is not true in general that if

$$\liminf_{n\to\infty} \mathscr{I}(\theta + X_n, Y_n) \geqq \mathscr{I}(\theta + X, Y),$$

then

$$\limsup_{n\to\infty} (e, \mathscr{I}^-(\theta + X_n, Y_n)e) \leqq (e, \mathscr{I}^-(\theta + X, Y)e), \qquad e \in R^d.$$

A counterexample can easily be constructed by letting $\mathscr{I}(\theta + X_n, Y_n)$ correspond to the 2×2 matrix

$$\begin{pmatrix} n^2 + 1 & n \\ n & 1 \end{pmatrix}.$$

One can show, however, by the argument used in proving Theorem 3.1 that if $\mathscr{I}(\theta + X_n, Y_n) \uparrow \mathscr{I}(\theta + X, Y)$ then $\mathscr{I}^-(\theta + X_n, Y_n) \downarrow \mathscr{I}^-(\theta + X, Y)$.

For $e \in R^d$ set $\mu = (e, \theta)$. An estimator T_n of μ based on the sample

$$(\theta + X_1, Y_1), \cdots, (\theta + X_n, Y_n)$$

of size n from $(\theta + X, Y)$ is said to be invariant if $T_n - \mu$ is independent of θ.

If so, we can write

$$U_n = T_n - \mu = (e, X_1) + \varphi(Y_1, X_2 - X_1, Y_2, \cdots, X_n - X_1, Y_n) .$$

THEOREM 3.5. *Let T_n be an invariant estimator of $\mu = (e, \theta)$ based on a sample of size n from $(\theta + X, Y)$ and set $U_n = T_n - \mu$. Then*

$$n\mathscr{I}^-(\mu + U_n) \geqq (e, \mathscr{I}^-(\theta + X, Y)e) .$$

PROOF. By using Theorem 2.9 we can reduce the general case to the case $n = 1$. By using Corollary 3.1 we can assume that $e = (1, 0, \cdots, 0)$. Then, letting θ_1 and X_1 denote the first coordinates of θ and X respectively,

$$\mathscr{I}^-(\theta_1 + X_1 + \varphi(Y)) \geqq \mathscr{I}^-(\theta_1 + X_1, Y) \geqq (e, \mathscr{I}^-(\theta + X, Y)e) ,$$

the last inequality following from Theorem 3.3.

4. Approximate maximum likelihood estimators. In this section we study estimators of θ based on samples from $(\theta + X + \sigma W, Y)$ for fixed $\sigma > 0$. These results are mainly tools for the proofs in Section 5 and in [8]. By definition

$$f_\sigma(x \mid y) = \int \varphi_\sigma(x - z)F(dz \mid y) ,$$

where

$$\varphi_\sigma(x) = \frac{1}{\sigma^d(2\pi)^{d/2}} \exp[-|x|^2/2\sigma^2] .$$

It follows easily that $f_\sigma(x \mid y) > 0$ for all x and y and that, for each y, $f_\sigma(x \mid y)$ is infinitely differentiable in x. Each such derivative is uniformly bounded in x and y and, for fixed y, vanishes as $|x| \to \infty$.

THEOREM 4.1. *Suppose that $E|X|^2 < \infty$ and let $\hat{\theta}(\sigma)$ be an unbiased estimator of θ based on $(\theta + X + \sigma W, Y)$. Then* $\operatorname{Cov} \hat{\theta}(\sigma) \geqq \mathscr{I}^-(\theta + X + \sigma W, Y)$.

PROOF. By the Cramér–Rao inequality (see page 265 of Rao [7]) it suffices to verify that

$$\iint \left(\frac{\max_{|\theta| \leq 1} |f_\sigma'(x - \theta \mid y)|}{f_\sigma(x \mid y)}\right)^2 f_\sigma(x \mid y)\, dx \mu_Y(dy) < \infty . \tag{4.1}$$

Since $E|X|^2 < \infty$, it is immediate that

$$E|X + \sigma W|^2 < \infty . \tag{4.2}$$

Note also that

$$\int \frac{dx}{(1 + |x|)^{d+1}} < \infty . \tag{4.3}$$

By using (4.2) and (4.3) respectively for the cases

$$f_\sigma(x \mid y) \geqq (1 + |x|)^{-(d+3)} \qquad \text{and} \qquad f_\sigma(x \mid y) < (1 + |x|)^{-(d+3)}$$

we conclude that

$$\iint (f_\sigma(x \mid y))^{-2/(d+3)} f_\sigma(x \mid y)\, dx \mu_Y(dy) < \infty . \tag{4.4}$$

Now

$$|f_\sigma'(x-\theta\,|\,y)| = |\int \varphi_\sigma'(x-z-\theta)F(dz\,|\,y)|$$
$$\leqq \int \frac{|x-z-\theta|}{\sigma^2}\varphi_\sigma(x-z-\theta)F(dz\,|\,y)$$
$$\leqq \int \frac{|x-z-\theta|}{\sigma^2}e^{(x-z,\theta)/\sigma^2}\varphi_\sigma(x-z)F(dz\,|\,y)\,.$$

By straightforward computations we can find an $L_0 \geqq 1$ such that

$$\max_{|\theta|\leqq 1}|f_\sigma'(x-\theta\,|\,y)| \leqq Lf_\sigma(x\,|\,y) + L^{-(d+2)}\,, \qquad L \geqq L_0\,. \tag{4.5}$$

It follows from (4.5) that

$$\max_{|\theta|\leqq 1}|f_\sigma'(x-\theta\,|\,y)| \leqq 2(f_\sigma(x\,|\,y))^{-1/(d+3)}f_\sigma(x\,|\,y) \quad \text{if} \quad (f_\sigma(x\,|\,y))^{-1/(d+3)} \geqq L_0\,. \tag{4.6}$$

Since $f_\sigma'(x\,|\,y)$ is uniformly bounded in x and y, we conclude from (4.6) that for some N

$$\max_{|\theta|\leqq 1}|f_\sigma'(x-\theta\,|\,y)| \leqq N(f_\sigma(x\,|\,y))^{-1(d+3)}f_\sigma(x\,|\,y) \qquad \text{for all } x, y\,. \tag{4.7}$$

Equation (4.1) follows from (4.4) and (4.7).

Set

$$L_\sigma(x\,|\,y) = \frac{f_\sigma'(x\,|\,y)}{f_\sigma(x\,|\,y)} = \frac{d}{dx}\log f_\sigma(x\,|\,y)\,.$$

Then

$$E(L_\sigma(X+\sigma W\,|\,Y), e)^2 = \mathscr{I}(\theta + X + \sigma W, Y)(e)\,, \qquad e \in R^d\,. \tag{4.8}$$

Thus by Lemma 2.1.

$$E(L_\sigma(X+\sigma W\,|\,Y), e)^2 \leqq \sigma^{-2}|e|^2\,, \qquad e \in R^d\,,$$

and hence

$$E|L_\sigma(X+\sigma W\,|\,Y)|^2 \leqq d\sigma^{-2} < \infty\,, \tag{4.9}$$

We conclude from (4.9) that

$$E|L_\sigma(X+\sigma W\,|\,Y)| \leqq d^{\frac{1}{2}}\sigma^{-1} < \infty\,. \tag{4.10}$$

Since $f_\sigma(x\,|\,y) \to 0$ as $|x| \to \infty$, it follows from (4.10) that

$$EL_\sigma(X+\sigma W\,|\,Y) = 0\,. \tag{4.11}$$

Let $A(\sigma)$ be the positive definite symmetric linear transformation from R^d to itself defined by Lemma 2.2. Then

$$(e_1, A(\sigma)e_2) = E(L_\sigma(X+\sigma W\,|\,Y), e_1)(L_\sigma(X+\sigma W\,|\,Y), e_2)\,. \tag{4.12}$$

Let $C_c^{(2)}$ denote the real-valued functions on R^d which are twice continuously differentiable and have compact support. For $g \in C_c^{(2)}$ set $L_{\sigma,g}(x\,|\,y) = g(x)L_\sigma(x\,|\,y)$. Then for each y, $L_{\sigma,g}(\cdot\,|\,y) \in C_c^{(2)}$. Also $L_{\sigma,g}$ and its first two derivatives are uniformly bounded in x and y. Set

$$A_g(\sigma) = -EL'_{\sigma,g}(X+\sigma W\,|\,Y)\,,$$
$$B_g(\sigma) = \operatorname{Cov} L_{\sigma,g}(X+\sigma W\,|\,Y)\,,$$

and

$$C_g(\sigma) = EL_{\sigma,g}(X + \sigma W \mid Y) .$$

THEOREM 4.2. *Let* $g_\nu \in C_c^{(2)}$ *be such that* $\sup_{\nu,x} |g_\nu^{(k)}(x)| < \infty$ *for* $k = 0, 1, 2$, $\lim_{\nu\to\infty} g_\nu(x) = 1$ *for all* x, *and* $\lim_{\nu\to\infty} g_\nu^{(k)}(x) = 0$ *for* $k = 1, 2$ *and all* x. *Then*

$$\lim_{\nu\to\infty} A_{g_\nu}(\sigma) = \lim_{\nu\to\infty} B_{g_\nu}(\sigma) = A(\sigma) .$$

PROOF. It is easily seen that

$$(e_1, A_g(\sigma)e_2) = \iint (L_\sigma(x \mid y), e_1)(L_\sigma(x \mid y), e_2)g(x)f_\sigma(x \mid y)\, dx\mu_Y(dy)$$

and hence by (4.9) and (4.12) that $\lim_{\nu\to\infty} A_{g_\nu}(\sigma) = A(\sigma)$. Similarly one proves that $\lim_{\nu\to\infty} B_{g_\nu}(\sigma) = A(\sigma)$.

Consider a random sample

$$(\theta + X_1 + \sigma W_1, Y_1), \cdots, (\theta + X_n + \sigma W_n, Y_n)$$

of size n from $(\theta + X + \sigma W, Y)$. Let $\bar\theta_n(\sigma)$ be the estimator of θ defined as follows: the ith coordinate of $\bar\theta_n(\sigma)$ is the sample median of the ith coordinate of

$$\theta + X_1 + \sigma W_1, \cdots, \theta + X_n + \sigma W_n$$

minus the median of the distribution of the ith coordinate of $X + \sigma W$.

THEOREM 4.3. (i) *Let* $0 < c < \infty$. *Then*

$$\limsup_{n\to\infty} n^{r/2} E \min(c, |\bar\theta_n(\sigma) - \theta|^r) < \infty , \qquad r > 0 ,$$

and for some $\varepsilon > 0$,

$$\lim_{n\to\infty} e^{\varepsilon n} P(|\bar\theta_n(\sigma) - \theta| \geqq c) = 0 .$$

(ii) *If* (1.1) *holds, then*

$$\limsup_{n\to\infty} n^{r/2} E|\bar\theta_n(\sigma) - \theta|^r < \infty , \qquad r > 0 .$$

PROOF. This result follows easily from the formula for the density of the sample median.

Let g be such that $A_g(\sigma)$ is invertible (such g's exist by Theorem 4.2). Let $\tilde\theta_n(\sigma, g)$ be the "approximate maximum likelihood estimator" of θ defined by

$$\tilde\theta_n(\sigma, g) = \bar\theta_n(\sigma) - A_g^{-1}(\sigma) \frac{1}{n} \sum_{i=1}^n [L_{\sigma,g}(\theta + X_i + \sigma W_i - \bar\theta_n(\sigma) \mid Y_i) - C_g(\sigma)] .$$

We will determine the asymptotic behavior $\tilde\theta_n(\sigma, g) - \theta$ by first comparing it to

$$-A_g^{-1}(\sigma)\frac{1}{n} \sum_{i=1}^n [L_{\sigma,g}(X_i + \sigma W_i \mid Y_i) - C_g(\sigma)] ,$$

which is asymptotically normal by the central limit theorem.

THEOREM 4.4. (i) *Let* $0 < c < \infty$. *Then*

$$\lim_{n\to\infty} nE \min\left(c, \left|\tilde\theta_n(\sigma, g) - \theta + A_g^{-1}(\sigma)\frac{1}{n} \sum_{i=1}^n [L_{\sigma,g}(X_i + \sigma W_i \mid Y_i) - C_g(\sigma)]\right|^2\right) = 0$$

and, for some $\varepsilon > 0$,

$$\lim_{n\to\infty} e^{\varepsilon n} P(|\tilde\theta_n(\sigma, g) - \theta| \geqq c) = 0 .$$

(ii) *If* (1.1) *holds, then*

$$\lim_{n\to\infty} nE\left|\tilde{\theta}_n(\sigma, g) - \theta + A_g^{-1}(\sigma)\frac{1}{n}\sum_{i=1}^n [L_{\sigma,g}(X_i + \sigma W_i \mid Y_i) - C_g(\sigma)]\right|^2 = 0 .$$

PROOF. We start with the Taylor expansion

$$\begin{aligned}\tilde{\theta}_n(\sigma, g) - \theta &= \bar{\theta}_n(\sigma) - \theta - A_g^{-1}(\sigma)\frac{1}{n}\sum_{i=1}^n [L_{\sigma,g}(X_i + \sigma W_i \mid Y_i) - C_g(\sigma)] \\ &\quad + A_g^{-1}(\sigma)\frac{1}{n}\sum_{i=1}^n L'_{\sigma,g}(X_i + \sigma W_i \mid Y_i)(\bar{\theta}_n(\sigma) - \theta) \\ &\quad + O(|\bar{\theta}_n(\sigma) - \theta|^2) \\ &= -A_g^{-1}(\sigma)\frac{1}{n}\sum_{i=1}^n [L_{\sigma,g}(X_i + \sigma W_i \mid Y_i) - C_g(\sigma)] \\ &\quad + A_g^{-1}(\sigma)\frac{1}{n}\sum_{i=1}^n [L'_{\sigma,g}(X_i + \sigma W_i \mid Y_i) + A_g(\sigma)](\bar{\theta}_n(\sigma) - \theta) \\ &\quad + O(|\bar{\theta}_n(\sigma) - \theta|^2) .\end{aligned}$$

Observe that $[L_{\sigma,g}(X_i + \sigma W_i \mid Y_i) - C_g(\sigma)]$ are i.i.d. and that each is bounded and has mean zero. The same is true of $[L'_{\sigma,g}(X_i + \sigma W_i \mid Y_i) + A_g(\sigma)]$. Recall that if Z_i have these properties, then for every $c > 0$ there is an $\varepsilon > 0$ such that

$$\lim_{n\to\infty} e^{\varepsilon n} P\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i\right| \geqq c\right) = 0 ;$$

also for $r > 0$

$$\lim_{n\to\infty} n^{r/2} E\left|\frac{1}{n}\sum_{i=i}^n Z_i\right|^r < \infty .$$

Theorem 4.4 follows from these observations together with Theorem 4.3 and Schwarz's inequality.

COROLLARY 4.1. (i) *As* $n \to \infty$

$$\mathscr{L}(n^{\frac{1}{2}}(\tilde{\theta}_n(\sigma, g) - \theta) \to N(0, A_g^{-1}(\sigma)B_g(\sigma)(A_g^{-1}(\sigma))^{\text{tr}}) ;$$

also for every $c > 0$

$$\lim_{n\to\infty} nE \min(c, |\tilde{\theta}_n(\sigma, g) - \theta|^2) = \text{trace } A_g^{-1}(\sigma)B_g(\sigma)(A_g^{-1}(\sigma))^{\text{tr}} .$$

(ii) *If* (1.1) *holds, then* $\lim_{n\to\infty} n|E\tilde{\theta}_n(\sigma, g) - \theta|^2 = 0$ *and*

$$\lim_{n\to\infty} n \operatorname{Cov} \tilde{\theta}_n(\sigma, g) = A_g^{-1}(\sigma)B_g(\sigma)(A_g^{-1}(\sigma))^{\text{tr}} .$$

5. Asymptotic behavior of the Pitman estimator. In this section we study the Pitman estimators of θ and $\mu = (e, \theta)$ and their respective minimax risks M_n and $M_n(e)$, all defined in the introduction.

THEOREM 5.1. $nM_n(e) \geqq (e, \mathscr{I}^-(\theta + X, Y)e)$.

PROOF. We can suppose that $M_n(e) < \infty$. Let P_n be the Pitman estimator of $\mu = (e, \theta)$ and set $U_n = P_n - \mu$. Then $EU_n = 0$ and $\operatorname{Var} U_n = M_n(e)$. Let W be independent of U_n and have the standard normal distribution on R. We conclude from Theorem 4.1 that

$$M_n(e) + \sigma^2 = \text{Var}\,(U_n + \sigma W) \geqq \mathscr{I}^-(\mu + U_n + \sigma W)\,.$$

By letting $\sigma \to 0$, and using Theorem 3.1, we conclude that $M_n(e) \geqq \mathscr{I}^-(\mu + U_n)$. The desired result now follows from Theorem 3.5.

COROLLARY 5.1. $nM_n \geqq \text{trace}\,\mathscr{I}^-(\theta + X, Y)$.

COROLLARY 5.2. *Suppose* $M_n < \infty$. *Then the Pitman estimator* $\hat{\theta}_n$ *satisfies*

$$n\,\text{Cov}\,\hat{\theta}_n \geqq \mathscr{I}^-(\theta + X, Y)\,. \tag{5.1}$$

THEOREM 5.2. *Suppose that* (1.1) *holds. Then the Pitman etimator* $\hat{\theta}_n$ *exists for n sufficiently large,*

$$\lim_{n\to\infty} \mathscr{L}(n^{\frac{1}{2}}(\hat{\theta}_n - \theta)) = N(0, \mathscr{I}^-(\theta + X, Y))\,, \tag{5.2}$$

and

$$\lim_{n\to\infty} n\,\text{Cov}\,\hat{\theta}_n = \mathscr{I}^-(\theta + X, Y)\,. \tag{5.3}$$

PROOF. It follows from (1.1) that for some n_0 the median of the first n_0 samples of $\theta + X$ is an invariant estimator of θ having finite total risk. Thus $M_n < \infty$ for $n \geqq n_0$. For $\sigma > 0$ let $(\theta + X_i + \sigma W_i, Y_i)$, $1 \leqq i \leqq n$, be a sample of size n from $(\theta + X + \sigma W, Y)$. Set $X_i(\sigma) = X_i + \sigma W_i$. The estimator $\tilde{\theta}_n(\sigma, g)$ from Theorem 4.4 is invariant and has finite covariance for n sufficiently large. Now $\hat{\theta}_n$ is the Pitman estimator of θ based on observations of $(\theta + X_i(\sigma), W_i, Y_i)$, $1 \leqq i \leqq n$. Thus

$$\hat{\theta}_n = \tilde{\theta}_n(\sigma, g) - E_0[\tilde{\theta}_n(\sigma, g)\,|\,X_i(\sigma) - X_1(\sigma), W_i, Y_i, 1 \leqq i \leqq n]\,.$$

Consequently

$$\text{Cov}\,(\tilde{\theta}_n(\sigma, g) - \hat{\theta}_n) = \text{Cov}\,\tilde{\theta}_n(\sigma, g) - \text{Cov}\,\hat{\theta}_n\,. \tag{5.4}$$

According to Corollary 4.1

$$\begin{aligned}\limsup_{n\to\infty} n\,\text{Var}\,(e, \hat{\theta}_n) &\leqq \lim_{n\to\infty} n\,\text{Var}\,(e, \tilde{\theta}_n(\sigma, g)) \\ &= (e, A_g^{-1}(\sigma)B_g(\sigma)(A_g^{-1}(\sigma))^{\text{tr}}e)\,.\end{aligned}$$

Thus by Theorem 4.2

$$\limsup_{n\to\infty} n\,\text{Var}\,(e, \hat{\theta}_n) \leqq (e, \mathscr{I}^{-1}(\theta + X + \sigma W, Y)e)\,.$$

By letting $\sigma \to 0$ we conclude that

$$\limsup_{n\to\infty} n\,\text{Var}\,(e, \hat{\theta}_n) \leqq (e, \mathscr{I}^-(\theta + X, Y)e)\,.$$

It now follows from Theorem 5.1 that

$$\lim_{n\to\infty} n\,\text{Var}\,(e, \hat{\theta}_n) = (e, \mathscr{I}^-(\theta + X, Y)e)\,,$$

which is equivalent to (5.3). From (5.3), (5.4) and Corollary 4.1 we see that

$$\lim_{n\to\infty} n\,\text{Cov}\,(\tilde{\theta}_n(\sigma, g) - \hat{\theta}_n) = A_g^{-1}(\sigma)B_g(\sigma)(A_g^{-1}(\sigma))^{\text{tr}} - \mathscr{I}^-(\theta + X, Y)\,.$$

Let g_ν be as in Theorem 4.2. Then by that theorem

$$\lim_{\sigma\to 0}\lim_{\nu\to\infty}\lim_{n\to\infty} n\,\mathrm{Cov}\,(\tilde{\theta}_n(\sigma, g_\nu) - \hat{\theta}_n) = 0\,.$$

Equation (5.2) now follows easily from Theorem 4.2 and Corollary 4.1.

REFERENCES

[1] BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

[2] GIRSHICK, M. A. and SAVAGE, L. J. (1951). Bayes and minimax estimates for quadratic loss functions. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 53-73.

[3] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73-101.

[4] KIEFER, J. (1957). Invariance, minimax sequential estimation and continuous time processes. *Ann. Math. Statist.* **28** 573-601.

[5] KUDÔ, H. (1955). On minimax invariant estimates of the transformation parameter. *Natur. Sci. Rep. Ochanomizu Univ.* **6** 31-73.

[6] PITMAN, E. J. G. (1938). The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika* **30** 391-421.

[7] RAO, C. R. (1965). *Linear Statistical Inference and its Applications.* Wiley, New York.

[8] STONE, C. J. (1973). Asymptotic properties of estimators of a location parameter (preprint).

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90024