ON PITMAN EMPIRICAL DISTRIBUTION AND STATISTICAL ESTIMATION

PRANAB KUMAR SEN

University of North Carolina at Chapel Hill

The posterior distribution of a parameter (with respect to a uniform weight function), termed the Pitman empirical distribution, provides the Pitman estimator as well as a posterior Pitman closest estimator. A systematic account of various properties of this empirical d.f. is provided with due emphasis on the related asymptotics.

1. Introduction. Let $X_1, \ldots, X_n$ be independent and identically distributed random vectors (i.i.d.r.v.), each with density $f(x, \theta)$ with respect to a sigma-finite measure μ, where $\theta \in \Theta \subset R^d$, for some $d \geq 1$. Then the joint density (i.e., the likelihood) function of $X_1, \ldots, X_n$ is given by

$$\ell_n(\theta) = \prod_{i=1}^{n} f(x_i, \theta), \quad \theta \in \Theta. \tag{1.1}$$

The classical Pitman-estimator (P.E.) of θ is defined as

AMS Classifications: 62C10, 62B99, 62F10.

Keywords and Phrases: Ancillarity; asymptotic Gaussian; equivariance; MLE; Pitman closeness; Pitman estimator; posterior Pitman closest estimator; robustness; shrinkage estimation.

Short Title: Pitman Empirical distribution.

$$(1.2) \qquad \hat{\theta}_{P,n} = \left\{\int \underset{\Theta}{\cdots} \int \theta\, \ell_n(\theta)d\theta\right\} \Big/ \left\{\int \underset{\Theta}{\cdots} \int \ell_n(\theta)d\theta\right\}$$

[see, Pitman (1939)]. In this context, we define the posterior density of θ (with respect to the uniform weight function) by

$$(1.3) \qquad g_{P,n}(\theta) = \ell_n(\theta) \Big/ \left\{\int \underset{\Theta}{\cdots} \int \ell_n(y)dy\right\}, \ \theta \in \Theta,$$

and the corresponding distribution function (d.f.)

$$(1.4) \qquad G_{P,n}(\theta) = \int \underset{y \le \theta}{\cdots} \int g_{P,n}(y)dy, \ \theta \in \Theta,$$

is termed the Pitman empirical d.f. (PEDF) of θ. Note that by (1.2)-(1.4),

$$(1.5) \qquad \hat{\theta}_{P,n} = \int \underset{\Theta}{\cdots} \int \theta\, dG_{P,n}(\theta) = E\{\theta | X_1, \ldots, X_n, U\},$$

where U stands for the (improper) prior distribution of θ generated by the uniform weight function on Θ. Thus, $\hat{\theta}_{P,n}$ is a Bayes estimator, and various properties of $\hat{\theta}_{P,n}$ for specific models (such as location-scale and exponential families) have been studied in detail in the literature; Lehmann (1983) and Berger (1985) are excellent sources. In the case of a real parameter θ, if we let

$$(1.6) \qquad \tilde{\theta}_{P,n} = G_{P,n}^{-1}(\tfrac{1}{2}) = \inf\{\theta : G_{P,n}(\theta) \ge \tfrac{1}{2}\}$$

$$= \text{median of the PEDF } G_{P,n},$$

then $\tilde{\theta}_{P,n}$ is a posterior Pitman closest (PPC) estimator of θ [viz., Ghosh and Sen (1991)]; the result holds in the multi-parameter case under additional regularity conditions [viz., Bose (1991)]. $\tilde{\theta}_{P,n}$ is unique whenever $G_{P,n}$ has a unique median. Further, whereas the PE $\hat{\theta}_{P,n}$ is adapted to squared error loss,

$\tilde{\theta}_{P,n}$ is adapted to absolute error loss, and under the Pitman closeness criterion (PCC), $\tilde{\theta}_{P,n}$ may dominate $\hat{\theta}_{P,n}$ [viz., Sen and Saleh (1991) and Kubokawa (1991)]. It is quite clear that both $\hat{\theta}_{P,n}$ and $\tilde{\theta}_{P,n}$ are functionals of the PEDF $G_{P,n}$, and there is ample room for other functionals of $G_{P,n}$ as competing estimators of θ (under other loss criteria). For this reason, it is of interest to study the basic properties of the PEDF $G_{P,n}$ itself and to incorporate these in optimal (Pitman-type) estimation of θ. This task constitutes the primary objective of the current study.

In a (multi-variate) location model, we set $f(x;\theta) = f(x-\theta)$, $x \in R^p$, $\theta \in R^p$, for some $p \geq 1$, where $f(\cdot)$ does not depend on θ. In this setup, $\hat{\theta}_{P,n}$ is translation-equivariant and unbiased for θ. For $p = 1$, $\tilde{\theta}_{P,n}$ is also translation-equivariant, but may not be universally unbiased. Also, within the class of translation-equivariant estimators, with respect to a quadratic loss, $\hat{\theta}_{P,n}$ is minimax for the location family, and it is admissible under additional regularity conditions. On the other hand, in a general estimation problem, with respect to an absolute error loss, $\tilde{\theta}_{P,n}$ may dominate $\hat{\theta}_{P,n}$; $\tilde{\theta}_{P,n}$ may also be "posterior Pitman closer" then $\hat{\theta}_{P,n}$. The parameter in a Poisson or gamma distribution are noteworthy examples in this context. Thus, the relative picture depends on the loss function or other criteria adapted, and most of the issues arising in this context can be better resolved by reference to the basic properties of the PEDF $G_{P,n}$. This provides further incentives for our contemplated study, so that the role of a conventional quadratic risk may be properly examined.

Pitman estimators are known to possess affinity to maximum likelihood estimators (MLE), especially in the asymptotic case, and to sufficient statistics whenever they exist. A similar picture holds for $\tilde{\theta}_{P,n}$ as well.

This affinity remains in tact for the PEDF $G_{P,n}$ which provides both the estimators $\hat{\theta}_{P,n}$ and $\tilde{\theta}_{P,n}$. These results are presented in Sections 2 and 3. The MLE are known to be dominated by their shrinkage or Stein-rule versions in the light of quadratic risk and the Pitman closeness measure [viz., Sen, Kubokawa and Saleh (1989)]. As such, the dominance of Pitman-type estimators by their shrinkage versions is also briefly considered (in the concluding section). Often, the MLE are not so robust (for small departures from the assumed model), and this drawback is shared by the PE $\hat{\theta}_{P,n}$ as well. For this reason, some functional Pitman estimators are considered in the concluding section (with due emphasis on the robustness aspects). The results in the concluding section are mostly asymptotic in nature.

2. PEDF and the exponential family. Suppose that the parameter $\theta \in \Theta \subset R^d$, for some $d \geq 1$ and the joint density $\ell_n(\theta)$ (of $X_1,\ldots,X_n$) admits a sufficient statistic T_n, so that $\ell_n(\theta)$ can be factorized as

(2.1) $$\ell_n(\theta) = h_n(T_n,\theta)\ \ell_n^*(X_1,\ldots,X_n;\ T_n),\ \theta \in \Theta,$$

where $\ell_n^*(\cdot)$ does not depend on θ. As for the sufficient statistic, we (assume a full rank model and) choose a suitable version, such that T_n is itself an estimator of θ in a meaningful sense (viz., sample mean vs. sum in a normal or exponential model). In this setup, $h_n(T_n,q)$ can be taken as the pdf of T_n, so that

(2.2) $$\int h_n(T_n,\theta)\ d\,\mu_n(T_n) = 1 \quad \text{a.a.}\theta.$$

From (1.3) and (2.1), we have

(2.3) $$g_{P,n}(\theta) = h_n(T_n,\theta)/\left\{\int_{\Theta}\cdots\int h_n(T_n,y)dy\right\},\ \theta \in \Theta,$$

so that

$$(2.4) \qquad G_{P,n}(\theta) = \left\{\int \underset{y \le \theta}{\cdots} \int h_n(T_n, y)dy\right\} \Big/ \left\{\int \underset{\Theta}{\cdots} \int h_n(T_n, y)dy\right\}, \ \theta \in \Theta.$$

Thus, whenever a sufficient statistic (estimator) T_n exists, the PEDF $G_{P,n}(\cdot)$ depends solely on T_n through its density $h_n(T_n;\cdot)$. In fact, $G_{P,n}(\cdot)$ is a random d.f. (defined on Θ) and is itself a sufficient statistic (process) whenever a sufficient statistic exists.

Let us consider a general exponential family of densities (in a minimal canonical form), where we set

$$(2.5) \qquad f(x,\theta) = a(\theta)b(x) \exp\{\alpha(\theta) \cdot t(x)\}, \ \theta \in \Theta,$$

where $a(\theta)$ is > 0, $b(x) \ge 0$, $\alpha(\theta) \in R^d$ and $t(x) \in R^d$, for some $d \ge 1$; without any loss of generality, we assume that the elements of $t(x)$ are affinely independent [viz., Barndorff-Nielsen (1978)]. Then, we have

$$(2.6) \qquad \ell_n(\theta) = ([a(\theta)]^n \prod_{i=1}^{n} b(X_i)) \exp\{n\ \alpha(\theta) \cdot T_n\},$$

where $T_n = n^{-1} \Sigma_{i=1}^n t(X_i) \in R^d$. Therefore,

$$(2.7) \qquad g_{P,n}(\theta) = \frac{(a(\theta))^n \exp\{n\ \alpha(\theta) \cdot T_n\}}{\int \underset{\Theta}{\cdots} \int (a(y))^n \exp\{n\ \alpha(y) \cdot T_n\}dy}, \ \theta \in \Theta,$$

which belongs to the conjugate family (of densities on Θ) of the form

$$(2.8) \qquad d_n(\alpha, T_n)[a(\theta)]^n \exp\{n\ \alpha(\theta) \cdot T_n\}, \ \theta \in \Theta$$

where $d_n(\alpha, T_n)$ depends on T_n (given) and $\alpha(\cdot)$, but not on θ. Thus, the PEDF $G_{P,n}$ can be characterized by the conjugate family in (2.8) wherein the sufficient statistic T_n is regarded as fixed while θ varies over Θ. This conjugate density may often suggest an appropriate loss function (or other criteria) and provide the desired estimators.

Suppose that $\{gX,\ g \in \mathcal{G}\}$ be a group of transformations (of the sample space onto itself) which leave the model invariant. The transformation gX induces a transformation (on Θ) $\theta \to \bar{g}\theta = \bar{\theta}$ where the $\bar{g}$ also form a group, denoted by $\bar{\mathcal{G}}$. Further, if T_n is equivariant (with respect to $g \in \mathcal{G}$), then

$$T_n(gx_1,\ldots,gx_n) = g^* \, T_n(x_1,\ldots,x_n), \quad g \in \mathcal{G}, \tag{2.9}$$

where the transformations g^* form a group, denoted by $\mathcal{G}^*$. Recall that the Jacobian of the transformation $T_n \to T_n^* = g^* \, T_n$ does not involve θ, and hence, the inherent ($\mathcal{G}$-) equivariance of the model implies that $h_n(T_n,\theta)$ expressed in its new coordinate system $(T_n^*,\ \bar{\theta})$ as $h_n^*(T_n^*,\ \bar{\theta})$ satisfies the following:

$$h_n^*(T_n^*,\bar{\theta}) = d(T_n,g^*) \; h_n(T_n,\theta), \tag{2.10}$$

for every $T_n^* = g^* \, T_n$ and $\bar{\theta} = \bar{g}\theta$, $g^* \in \mathcal{G}^*$ and $\bar{g} \in \bar{\mathcal{G}}$. We denote the PEDF for the transformed model by $G^*_{p,n}$, so that by (2.3), (2.4) and (2.10), we obtain that

$$G^*_{p,n}(\bar{g}(\theta+y)) = G_{p,n}(\theta+y), \; g \in \mathcal{G}, \; (\theta + y) \in \Theta, \tag{2.11}$$

where the transitiveness of $\bar{\mathcal{G}}$ is tacitly assumed to justify (2.10).

Thus, whenever a group of transformations $\mathcal{G}$ leave the model invariant the PEDF $G_{p,n}$ is also $\mathcal{G}$-equivariant. The picture simplifies a bit more for the classical location model where $f(x,\theta) = f(x-\theta)$, $x \in E^p$, $\theta \in R^p$, so that $h_n(T_n,\theta) = h_n(T_n-\theta)$ and $h_n(u)$ is independent of θ (but the form may depend on n). If we let

$$\bar{H}_n(u) = \int \underset{y \geq u}{\cdots} \int h_n(u), \; u \in R^p, \tag{2.12}$$

then, we have by (2.4) and (2.12),

$$G_{P,n}(x) = \bar{H}_n(T_n - x), \quad x \in R^P, \tag{2.13}$$

so that for every $x \in R^P$,

$$\begin{aligned} G_{P,n}(\theta + x) &= \bar{H}_n(T_n - \theta - x) \\ &= P_\theta\{(T_n - \theta) \geq x\} \\ &= P_0\{T_n \geq x\} \\ &= G^{(0)}_{P,n}(x), \text{ say,} \end{aligned} \tag{2.14}$$

where $G^{(0)}_{P,n}(x)$, the PEDF of T_n, under $\theta = 0$, is independent of θ. Thus, if we consider the group $\mathcal{G}$ of affine transformations

$$X \to X^* = a + BX, \text{ a real p-vector, B nonsingular,} \tag{2.15}$$

then on letting $\bar{\theta} = a + B\theta$ and $T^*_n = a + BT_n$, we have

$$G^*_{P,n}(\bar{\theta} + y^*) = G^{(0)}_{P,n}(y) = G_{P,n}(\theta + y), \tag{2.16}$$

whenever $y^* = a + By$, $y \in R^P$. Note that if $H_n(u)$, $u \in R^P$ is (diagonally) symmetric about 0, then $\bar{H}_n(u) = H_n((-1)u)$, $\forall\, u \in R^P$, and hence, for every $x \in R^P$,

$$\begin{aligned} \bar{H}_n(T_n - x) &= \bar{H}_n((T_n - \theta) - (x - \theta)) \\ &= H_n(x - T_n), \ \forall\, T_n \in R^P, \end{aligned} \tag{2.17}$$

so that in (2.13), we may as well replace $\bar{H}_n(T_n - x)$ by $\bar{H}_n(x - T_n)$. In any case, (2.16) insures that the PEDF $G_{P,n}$ is $\mathcal{G}$-equivariant for the location model. Conclusions about the uniqueness of the median of $G_{P,n}$ (for p=1) and the multivariate-median (for $p \geq 2$) can be drawn as in Ghosh and Sen (1991) and Bose (1991), and the posterior Pitman closest property of T_n can be established thereof.

It may be quite appropriate to consider another example. Consider a multi-normal population with null mean vector and dispersion matrix Σ (positive definite (p.d.) but arbitrary). Then T_n is the sample covariance matrix, so that (2.1) holds with $h_n(\cdot)$ related to the classical Wishart density with n degrees of freedom (DF) and the PEDF $G_{P,n}$ reduces to the conjugate Wishart distribution. However, $G_{P,n}$ fails to satisfy the requirement of multivariate median unbiasedness (and it is not diagonally symmetric). The group $\mathcal{G}$ of all nonsingular matrices B (i.e., $X \to Y = BX$) leaves the model invariant and T_n is $\mathcal{G}$-equivariant:

$$T_n^* = BT_nB' \text{ and } \Sigma^* = B\Sigma B'. \tag{2.18}$$

In this context, the following two loss functions are generally used:

$$L_q(\Sigma, T_n) = \mathrm{Tr}(\Sigma^{-1} T_n - I)^2, \tag{2.19}$$

$$L_\ell(\Sigma, T_n) = \mathrm{tr}(\Sigma^{-1} T_n) - \ell og|\Sigma^{-1} T_n| - p. \tag{2.20}$$

Recall that $A_n = n\, T_n$ has the Wishart distribution $W_p(\Sigma, n)$. The quadratic risk of an estimator $T_n(a) = a\, A_n$, $a > 0$ is minimized at $a = (n+p+1)^{-1}$, and the likelihood risk of $T_n(a)$ is minimized at $a = n^{-1}$. Although the minimum risk is constant (for its respective loss), the estimator $T_n(a)$ is not minimax. In the (generalized) Pitman closeness (GPC) sense, within the class

$$\mathscr{C}_1 = \{T_n : T_n = a\, A_n,\ a > 0\}, \tag{2.21}$$

the closest estimator of Σ is

$$T_n^* = A_n\{p/\mathrm{med}(\chi^2_{np})\}. \tag{2.22}$$

Note that an unbiased estimator (sufficient statistic) of Σ is

(2.23) $$T_n = n^{-1} A_n,$$

and that $\mathscr{C}_1$ is the class of all equivariant estimators of Σ under the group of nonsingular transformations (too big to achieve a best estimator w.r.t. L_q or L_ℓ). For this purpose, we may note that there exists a lower triangular U_n, such that $A_n = U_n U_n'$. We consider the group of lower triangular transformations, i.e., $A_n \to H A_n H'$ and $\Sigma \to H \Sigma H'$, for lower triangular H. The corresponding class of equivariant estimators is

(2.24) $$\mathscr{C}_2 = \{T_n : T_n = U_n D U_n', \text{ D diagonal and p.d.}\}.$$

The minimum risk (w.r.t. L_ℓ) estimator in $\mathscr{C}_2$ is

(2.25) $$T_{nJS} = U_n B_n U_n';\ B_n = \text{diag}\left[\frac{1}{n+p+1-2j},\ 1 \le j \le p\right]$$

[viz., James and Stein (1961)]. However, in the GPC sense (with respect to $L_\ell(\cdot)$), a best (closest) equivariant estimator in $\mathscr{C}_2$ does not exist (for $p \ge 2$) [viz., Sen, Nayak and Khattree (1991)]. A similar picture holds for $L_q(\cdot)$. These results show that the classical Pitman estimator of Σ (based on the conjugate Wishart distribution) does not have the "bestness", and there is a need to probe further into the $G_{P,n}$ to characterize alternative estimators which perform better.

The density in (2.6) involving (T_n, θ) may not have naturally independent (or uncorrelated) coordinate variables. Suppose that $\theta = (\theta^{(1)}, \theta^{(2)})$ and consider a transformation $\theta \to \tau = (\tau^{(1)}, \tau^{(2)})$ where $\tau^{(2)} = \theta^{(2)}$ and $\tau^{(1)}$ is a function of $\theta^{(1)}$ and $\theta^{(2)}$, such that $\tau^{(1)}$ and $\tau^{(2)}$ (or $\theta^{(2)}$) are L-independent. Then, by Theorem 9.12 of Barndorff-Nielsen (1978), we conclude that $\tau^{(1)}$ and $\theta^{(2)}$, considered as r.v.'s on Θ are stochastically independent under the conjugate family, so that under the transformation $\theta \to \tau$, the PEDF $G_{P,n}(\tau)$ can be expressed as the product of $G^{(1)}_{P,n}(\tau^{(1)})$ and $G^{(2)}_{P,n}(\tau^{(2)})$ where the component PEDF's can be obtained from the respective reduced models in (2.8). Example 9.7 (on p. 149) in Barndorff-Nielsen (1978) provides a nice application of this orthogonalization of PEDF's in the context of estimation of a minor of the covariance matrix Σ. Motivated by this, we may consider this conditional approach in a quasi-independence case as follows.

We consider a partition of θ and T_n as $(\theta^{(1)}, \theta^{(2)})$ and $(T_n^{(1)}, T_n^{(2)})$ respectively, such that in (2.1),

(2.26) $$h_n(T_n, \theta) = h_{n1}(T_n^{(1)}, \theta)\, h_{n2}(T_n^{(2)}, \theta^{(2)}), \quad \theta \in \Theta,$$

so that $T_n^{(1)}$, $T_n^{(2)}$ are independent, the density of $T_n^{(2)}$ depends only on $\theta^{(2)}$, but the density of $T_n^{(1)}$ may depend on both $\theta^{(1)}$, $\theta^{(2)}$. Suppose further that

(2.27) $$\int h_{n1}(T_n^{(1)}, \theta)d\theta^{(1)} \equiv 1(\text{wlog}),\ \forall\ \theta^{(2)} \in \Theta^{(2)}.$$

Then, under (2.26) and (2.27), (2.3) reduces to

(2.28) $$g_{P,n}(\theta^{(1)}, \theta^{(2)}) = h_{n1}(T_n^{(1)}, \theta)\, \frac{h_{n2}(T_n^{(2)}, \theta^{(2)})}{\int \cdots_{\Theta} \int h_{n2}(T_n^{(2)}, y^{(2)})dy^{(2)}}$$

$$= g^{(1)}_{P,n}(\theta^{(1)} | \theta^{(2)}) \cdot g^{(2)}_{P,n}(\theta^{(2)}), \quad \theta \in \Theta,$$

where $g^{(2)}_{P,n}(\theta^{(2)})$ is the marginal Pitman density of $\theta^{(2)}$ and $g^{(1)}_{P,n}(y^{(1)} | y^{(2)})$ is the conditional density of $\theta^{(1)}$, given $\theta^{(2)} = y^{(2)}$. The PEDF corresponding to $g^{(1)}_{P,n}$, denoted by $G^{(1)}_{P,n}(\cdot | y^{(2)})$, is termed the conditional PEDF of $\theta^{(1)}$ given $\theta^{(2)} = y^{(2)}$. In an estimation problem if $\theta^{(1)}$ is the parameter of interest and

$\theta^{(2)}$ is a nuisance parameter, (2.28) provides a convenient way of deriving (conditional) Pitman estimators and other related ones. For example, if $G_{P,n}^{(1)}(\cdot|\theta^{(2)})$ has location parameter (mean/median) $T_n^{(1)}$ for every $\theta^{(2)}$, then $T_n^{(1)}$ is a convenient (Pitman type) estimator of $\theta^{(1)}$. As an illustrative example, we consider the case of a multinormal density with mean vector $\theta^{(1)}$ and dispersion matrix $\theta^{(2)}$. Then $G_{P,n}^{(1)}(\theta^{(1)}|\theta^{(2)})$ is $\mathcal{N}_p(\bar{X}_n, \frac{1}{n}\theta^{(2)})$, which has the natural location parameter $\bar{X}_n$ (the PE of $\theta^{(1)}$ when $\theta^{(2)}$ is known), which can as well be adopted in an unconditional setup. In the same vein, let $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)})$ and $\theta^{(2)} = ((\theta_{ij}^{(2)})_{i,j=1,2}$. Then $g_{P,n}^{(2)}(\theta^{(2)})$ has the conjugate Wishart form, $g_{P,n}^{(12)}(\theta_2^{(1)}|\theta^{(2)})$ is normal with mean vector $\bar{X}_n^{(2)}$ and dispersion matrix $n^{-1}\theta_{22}^{(2)}$, while $g_{P,n}^{(11)}(\theta_1^{(1)}|\theta_2^{(1)},\theta^{(2)})$ is normal with mean vector $\bar{X}_n^{(1)} + (\theta_{22}^{(2)})^{-1}(\theta_2^{(1)} - \bar{X}_n^{(2)})$ and dispersion matrix $n^{-1}(\theta_{11}^{(2)} - \theta_{12}^{(2)}(\theta_{22}^{(2)})^{-1}(\theta_{21}^{(2)})) = n^{-1}\theta_{11.2}^{(2)}$. The PE of $\theta_2^{(1)}$ is $\bar{X}_n^{(2)}$, so that $G_{P,n}^{(11)}(\theta^{(1)}|\bar{X}_n^{(2)}, \theta^{(2)})$ provides the usual Pitman estimator $\bar{X}_n^{(1)}$ of $\theta_1^{(1)}$. Such conclusions may also be derived for some non-regular models. For example, uniform $(\theta_1 - \frac{1}{2}\theta_2, \theta_1 + \frac{1}{2}\theta_2)$, θ_1 real $\theta_2 > 0$, density. The sample extreme values $X_{n:1}$ and $X_{n:n}$ are jqintly sufficient for (θ_1,θ_2), $T_n^{(1)} = \frac{1}{2}(X_{n:1} + X_{n:n})$ and $T_n^{(2)}$

$= (X_{n:n} - X_{n:1})$. The d.f. of $T_n^{(2)}$ does not depend on θ_1, while $T_n^{(1)}$ has a density depending on both θ_1 and θ_2. However, the conditional density of $T_n^{(1)}$, given $T_n^{(2)}$, is symmetric about θ_1, so that a similar characterization of

$G^{(1)}_{P,n}(\theta_1|\theta_2)$ works out well.

The results presented here are all of exact nature. As the sample size n becomes large, under fairly general regularity conditions $T_n \to \theta$, in probability, and hence, asymptotically, $G_{P,n}(y)$ becomes degenerate at the point θ. This calls for suitable (Pitman) neighborhoods of θ on which the PEDF behave property, and this will be considered in the next section in a more general setup where sufficient statistics may not exist.

3. <u>PEDF and MLE</u>. Pitman estimators are known to have affinity to MLE's. This affinity extends directly to the PEDF through a local asymptotic normality (LAN) pattern, and this will be considered here. In this context, exponential family, sufficiency and equivariance do not play any basic role.

We assume that the density $f(x,\theta)$, $x \in R^p$, $\theta \in \Theta \subset R^d$, for some $p \geq 1$, $d \geq 1$, satisfy the following (essentially, Cramér-type) regularity conditions:

(i) $(\partial/\partial\theta)f(x,\theta)$ and $(\partial^2/\partial\theta\partial\theta')f(x,\theta)$ exist almost everywhere and are dominated by some integrable functions (w.r.t. x).

(ii) $(\partial/\partial\theta)\log f(x,\theta)$ and $(\partial^2/\partial\theta\partial\theta')\log f(x,\theta)$ exist almost everywhere, and are such that

(a) The Fisher information (matrix) $\mathcal{I}_\theta$, defined by

$$\mathcal{I}_\theta = E_\theta\{[(\partial/\partial\theta)\log f(x,\theta)][(\partial/\partial\theta')\log f(x,\theta)]\} \tag{3.1}$$

is finite and p.d., and

(b) for every $\delta > 0$, there exists a $\psi_\delta (> 0)$, such that

$$E_\theta\left\{\sup_{u:\|u\|\leq\delta} \left\| \frac{\partial^2}{\partial\theta\partial\theta'}\log f(x;\theta+u) - \frac{\partial^2}{\partial\theta\partial\theta'}\log f(x;\theta)\right\|\right\} = \psi_\delta \quad \text{exists} \tag{3.2}$$

and $\psi_\delta \downarrow 0$ as $\delta \downarrow 0$.

(iii) If we define $Z(u) = -\log\{f(x,\theta)/f(x;\theta+u)\}$, $u \in R^d$ and define the Kullback-Leibler information by

(3.3) $$I(u;\theta) = E_\theta Z(u) = \int_{R^p} \{-\log\{f(x;\theta)/f(x,\theta+u)\}\}f(x,\theta)d\mu,$$

for $u \in R^d$, then

(3.4) $$I(u;\theta) > 0, \ \forall u \neq 0,$$

and, either Θ is a compact subset of R^d, or, letting

(3.5) $$Z_n(u) = \int Z(u)dF_n = \frac{1}{n} \Sigma_{i=1}^n\{-\log\{f(X_i,\theta)/f(X_i,\theta+u)\}\},$$

(and noting that $E_\theta Z_n(u) = \int Z(u)dF = I(u;\theta)$), we have:

(a) There exists a positive number k, such that

(3.6) $$\int_{u \in R^d} \exp\{-k \ I(u;\theta): du < \infty,$$

and (b) there exists a compact Θ^* (containing $u = 0$ as an inner point), such that on letting $\bar{\Theta}^* = \Theta \backslash \Theta^*$,

(3.7) $$\sup_{\theta \in \bar{\Theta}^*} |Z_n(u)/I(u;\theta) - 1| \to 0, \text{ in probability,}$$

as $n \to \infty$.

Note that (3.6) insures that as $n \to \infty$,

(3.7) $$\int_{R^d} \exp\{-n \ I(u;\theta)\}du \to 0$$

(at an exponential rate), and moreover,

(3.8) $$\int_{\bar{\Theta}^*} \exp\{-n \ Z_n(u)\} \ du \overset{P}{\to} 0, \text{ as } n \to \infty.$$

Further, if for every $\epsilon > 0$, we denote by $\Theta_\epsilon^* = \{\theta' \in \Theta^* : \|\theta' - \theta\| > \epsilon\}$, then by (3.4), (3.5) and some standard steps, as $n \to \infty$,

(3.9) $$\int_{\Theta_\epsilon^*} \exp\{-n \ Z_n(\theta')\} \ d\theta' \overset{P}{\to} 0, \quad \text{as } n \to \infty,$$

where again the rate of convergence is exponential in n. For the time being let us assume that $\Theta = R^d$, for some $d \geq 1$, and write $\Theta = \bar{\Theta}^* \cup \Theta^*$ $= \bar{\Theta}^* \cup \Theta^*_\epsilon \cup \{\theta' : \|\theta'-\theta\| < \epsilon\}$. If Θ is conpact, we may drop $\bar{\Theta}^*$ in this decomposition. Let then

(3.10)
$$U_n = \frac{1}{\sqrt{n}} (\partial/\partial\theta)\log \ell n(\theta)$$
$$= n^{-½} \Sigma_{i=1}^n (\partial/\partial\theta)\log f(X_i,\theta)|_\theta,$$

(3.11)
$$V_n = -n^{-1}(\partial^2/\partial\theta\partial\theta')\log \ell n(\theta)$$
$$= n^{-1} \Sigma_{i=1}^n\{-(\partial^2/\partial\theta\partial\theta') \log f(X_i,\theta)|_\theta.$$

Then, under the stated regularity conditions, as $n \to \infty$

(3.12)
$$U_n \sim \mathcal{N}_d(0,\mathcal{I}_\theta),$$

(3.13)
$$V_n \to \mathcal{I}_\theta \text{ a.s., as } n \to \infty.$$

We also write

(3.14)
$$W_n = (V_n^{-1} U_n(\sim \mathcal{N}_d(0,\mathcal{I}_\theta^{-1}), \text{ as } n \to \infty).$$

[Recall that W_n is a form of Studentized scores at θ.] Next, we write for every $u \in R^d$,

(3.15)
$$\prod_{i=1}^n f(x_i, \theta+u)/\int_\Theta\cdots\int \prod_{i=1}^n f(x_i,y)dy$$
$$= \prod_{i=1}^n \frac{f(x_i,\theta+u)}{f(x_i,\theta)} / \int_\Theta\cdots\int \prod_{i=1}^n \frac{f(x_i,y)}{f(x_i,\theta)} dy$$
$$= \exp\{-n Z_n(u)\}/\int_\Theta\cdots\int \exp\{-n Z_n(y-\theta)\}dy.$$

First, consider the domain

(3.16)
$$D_n = \{u : u + n^{-½}t, \|t\| < K\},$$

where $K(< \infty)$ is arbitrarily large (but fixed). Then, under Assumptions (i) and (iia), (iib),

(3.17) $$\sup_{\|t\| \le K} |n Z_n(n^{-1/2} t) + t' U_n - \frac{1}{2} t' V_n t| \xrightarrow{P} 0, \text{ as } n \to \infty.$$

Next, over the domain $\Theta_\epsilon^* \setminus D_n$, we note that for ϵ small,

(3.18) $$\sup_{u \in \Theta_\epsilon^* \setminus D_n} |n Z_n(u) + n^{1/2} u' U_n - \frac{1}{2} n u' V_n u + o(n)\|u\|^2|$$

converges, in probability, to zero as $n \to \infty$. Over the complementary part, we make use of (3.8) and (3.9). Hence, multiplying both the numerator and denominator of (3.15) by $\exp\{-\frac{1}{2} U_n'(V_n)^{-1} U_n\}$ (and noting that by (3.14), $U_n'(V_n)^{-1} U_n = O_p(1)$), we obtain from (3.15) through (3.18) that for every $K(< \infty)$, $\|t\| < K$,

(3.19) $$\exp\{-Z_n(n^{-1/2} t)\} / \int \dots \int_{D_n} \exp\{-Z_n(u)\} du$$

$$= (2\pi)^{-\frac{1}{2}d} |V_n|^{-1/2} \exp\{-\frac{1}{2}(U_n - V_n t)' V_n^{-1} (U_n - V_n t)\} + o(1),$$

where in the denominator, the domain may as well be replaced by Θ. We rewrite the exponent in (3.19) as

(3.20) $$\exp\{-\frac{1}{2}(t - W_n)' V_n (t - W_n)\},$$

so that from (3.19) and (3.20), we conclude that as $n \to \infty$,

(3.21) $$|G_{P,n}(\theta + n^{-1/2} t) - \Phi_d(t; W_n, V_n^{-1})| \xrightarrow{P} 0,$$

uniformly in t $(\in R^d)$, where $\Phi(t; \mu, \Sigma)$ stands for a d-variate normal d.f. with mean vector μ and dispersion matrix Σ. Let us now denote the MLE of θ by $\hat{\theta}_n$; actually, we may take a BAN estimator of θ and denote by $\hat{\theta}_n$. Then, by virtue of (3.17), we have

$$n^{1/2}(\hat{\theta}_n - \theta) = V_n^{-1} U_n + o_p(1) \tag{3.22}$$
$$= W_n + o_p(1).$$

Consequently, from (3.21) and (3.22), we conclude that

$$G_{P,n}(\hat{\theta}_n + n^{-1/2}\xi) = G_{P,n}(\theta + (\hat{\theta}_n - \theta) + n^{-1/2}\xi) \tag{3.23}$$
$$= \Phi_d(\xi;\ 0,\ V_n^{-1}) + o_p(1),$$

uniformly in ξ. Thus, asymptotically, in a Pitman (i.e., $O(n^{-1/2})$) neighborhood of the MLE (BAN) $\hat{\theta}_n$ (and hence, around θ), the PEDF $G_{P,n}$ is Gaussian with mean $\hat{\theta}_n$ and dispersion matrix $n^{-1} V_n^{-1}$.

This representation is of prime importance in studying the asymptotic properties of Pitman-type estimators of θ (which are based on the $G_{P,n}$). In this characterization, because of the diagonal symmetry of Φ_d (around 0), we conclude that, asymptotically, $G_{P,n}$ attains diagonal symmetry in a Pitman neighborhood of θ (or $\hat{\theta}_n$). Hence if we use the Posterior Pitman closest (PPC) characterization of Ghosh and Sen (1991) [and Bose (1991)], we can claim that the classical Pitman estimator $\hat{\theta}_{P,n}$ is asymptotically equivalent to the MLE (BAN) $\hat{\theta}_n$ and is a PPC estimator of θ. This representation also provides a natural justification for using a quadratic loss in an asymptotic setup. At the same time, it raises some other issues which will be discussed in the next section.

<u>4. Some general remarks</u>. We have observed that (3.23) provides a justification for the adaptability of the PPC criterion (in an asymptotic setup) on $G_{P,n}$, and this entails that the classical Pitman estimator $\hat{\theta}_{P,n}$ is asymptotically a PPC estimator of θ [in the sense of Ghosh and Sen (1991)].

The underlying (multivariate) median unbiasedness property of the (posterior) Pitman empirical d.f. (in an asymptotic setup) insures that this asymptotic PPC property holds for a general class of (location-) measures of the PEDF $G_{P,n}$ which include the PC $\hat{\theta}_{P,n}$ as a special member. For example, for the multivariate normal dispersion matrix ($\theta = \Sigma$) estimation problem, treated in detail in Section 2 [(2.18) through (2.25)], the $G_{P,n}$ is not strictly Gaussian (for finite n), and (3.23) holds only in an asymptotic setup. In such a case, the influence of the tails of $G_{P,n}$ in a measure of its location may be quite perceptible for Pitman-type estimators but less for some alternative ones. As for example, we may consider the alternative estimator in (2.25) or even modify it by taking B as a diagonal matrix with the elements $\text{med}^{-1}(\chi^2_{p(n+1+p-2j)})$, $1 \le j \le p$. [Incidentally, in this context, we have used the entropy loss function instead of the conventional quadratic one.] Keeping this in mind, we may, for example, consider the marginal PEDFs $G^{(j)}_{P,n}(\theta_j)$, $1 \le j \le d$, denote the respective medians by $\tilde{\theta}^{(j)}_{P,n}$, $1 \le j \le d$, and let $\tilde{\theta}_{P,n} = (\tilde{\theta}^{(1)}_{P,n},\ldots,\tilde{\theta}^{(d)}_{P,n})'$. Then, $\tilde{\theta}_{P,n}$ is also a PPC estimator of θ. The main advantage of prescribing $\tilde{\theta}_{P,n}$ as an alternative to $\hat{\theta}_{P,n}$ is that $\hat{\theta}_{P,n}$ is much more sensitive to the tail-behavior of $G_{P,n}$ then $\tilde{\theta}_{P,n}$, so that $\tilde{\theta}_{P,n}$ is likely to be more robust than $\hat{\theta}_{P,n}$. Of course, it should be kept in mind that the PEDF $G_{P,n}$ is a conditional d.f., given $X_1,\ldots,X_n$, and it is highly influenced by the MLE $\hat{\theta}_n$ (or any other BAN estimator of θ based on $X_1,\ldots,X_n$). Quite often, the MLE (or BAN) $\hat{\theta}_n$ is attacked on the ground of plausible lack of robustness. Thus, if $\hat{\theta}_n$ is not so robust, (3.23), in turn, would imply that the Pitman-type estimators (based on the $G_{P,n}$) may share the same drawback to a certain extent. On the other hand, (3.23) ensures that the Gaussian approximation holds well in an $O(n^{-1/2})$-neighborhood of $\hat{\theta}_n$ (and hence θ), and the 'closeness' of this approximation may not be that fine in the tails to ensure that $\hat{\theta}_{P,n}$ is so robust. This may suggest that either $\tilde{\theta}_{P,n}$ or some other measure (of location of $G_{P,n}$) which is quite insensitive to the tail-behavior of $G_{P,n}$ should be preferable. In the case where $G_{P,n}$ is itself strictly Gaussian (viz., $f(x,\theta) = f(x-\theta)$, f (multi-) normal density), this point of distinction may not be very pertinent. However, in the negation of the exact Gaussian form for $G_{P,n}$, for finite sample sizes, such as a robustness consideration merits attention. Basically, one then prescribe a measure of location of the PEDF $G_{P,n}$ which is less sensitive to the tails. Trimmed mean, Winsorized mean, median or, in general, an L-functional with practically no weight attached to the tails would be a desirable solution. This leads us to the following robustification of PE.

Let $\tau_n = \tau(G_{P,n})$ be a functional of the PEDF $G_{P,n}$, such that τ_n is robust and translation-equivariant. Then $\tau_n = \hat{\theta}_{P,n}(\tau)$ is termed a functional Pitman estimator (FPE) of θ. The main point in this construction is that by virtue of (3.23), $G_{P,n}$ is non-degenerate only in an $O(n^{-½})$-neighborhood of $\hat{\theta}_n$ (and hence, θ), so that the tails of $G_{P,n}$ are all adapted to this shrinking balls around $\hat{\theta}_n$. As long as τ_n admits a first order representation (where the leading term is a linear functional), we may use (3.23) to claim that $n^{½}\|\hat{\theta}_n-\hat{\theta}_{P,n}(\tau)\| \to 0$, in probability, as $n \to \infty$, while depending on the particular form of $\tau(\cdot)$, we may be able to achieve more robustness. In addition to this, in (1.1) (and elsewhere), the likelihood $\ell_n(\theta)$ may be modified in a way to induce more robustness, and with that modification one may consider suitable Pitman-type estimators. While this works out well for the location model, there are some technicalities for a general model, and we shall not enter into these problems here.

Another important feature of multiparameter estimation is that in the case of $d \geq 3$ (with respect to a quadratic loss), the MLEs are generally dominated (at least asymptotically) by their shrinkage or Stein-rule versions. For the multinormal location model, this was the Stein-phenomenan introduced more than 35 years ago by Stein (1956), and since then a considerable amount of work has been done in this area. In the light of a generalized Pitman closeness measure, a similar dominance result has been established by Sen, Kubokawa and Saleh (1989), where d may be as low as 2. As such, the Gaussian approximation in (3.23) suggests that similar dominance results (under quadratic loss or Pitman closeness measure) should hold for the Pitman-type estimators (based on $G_{P,n}$). Indeed this is the case (in an asymptotic setup): The PE $\hat{\theta}_{P,n}$ or, in general, an FPE $\hat{\theta}_{P,n}(\tau)$ is dominated by suitable shrinkage or Stein-rule versions. For example, with a pivot θ_o (= 0, WOLOG), we may set

$$\hat{\theta}^S_{P,n} = \{1 - (d-2)\, \mathcal{L}_n^{-1}\}\hat{\theta}_{P,n}, \tag{4.1}$$

where $\mathcal{L}_n$ is the likelihood ratio test statistic for testing $H_0 : \underset{\sim}{\theta} = \underset{\sim}{0}$ vs. $H_1 : \theta \neq 0$, and d is ≥ 3; $\hat{\theta}^S_{P,n}$ is a James-Stein version of $\hat{\theta}_{P,n}$. More complex versions (including the positive-rule ones) can also be considered in the same vein. Moreover, with respect to PCC, the shrinkage factor $(d-2)$ may also be replaced by $b : 0 < b < (d-1)(3d+1)/2d$, $d \geq 2$. As such, the results in Sen (1986) and Sen, Kubokawa and Saleh (1989) can directly be adapted along with (3.23) to conclude that Stein-rule versions of the Pitman type estimators are asymptotically better than the original ones (in either mode).

REFERENCES

BARNDORFF-NIELSEN, O. (1978). Information and Exponential Families in Statistical Estimation. John Wiley, New York.

BERGER, J.O. (1985). Statistical Decision Theory and Bayesian Analysis. Springer-Verlag, New York.

BOSE, S. (1991). Some properties of posterior Pitman closeness. *Commun. Statist. Theory Meth.* 20, in press.

GHOSH, M. and SEN, P.K. (1989). Median unbiasedness and Pitman closeness. *J. Amer. Statist. Assoc.* 84, 1089-1091.

GHOSH, M. and SEN, P.K. (1991). Bayesian Pitman closeness. *Commun. Statist. Theory Meth.* 20, in press.

KUBOKAWA, T. (1991). Equivariant estimation under the Pitman closeness criterion. *Commun. Statist. Theory Meth.* 20, in press.

LEHMANN, E.L. (1983). Theory of Point Estimation. John Wiley, New York.

PITMAN, E.J.G. (1939) The estimation of location and scale parameters of a continuous population of any given form. *Biometrika* 30, 391-421.

SEN, P.K. (1986). On the asymptotic distributional risk of shrinkage and preliminary test versions of maximum likelihood estimators. *Sankhya, Ser. A.* 48, 354-372.

SEN, P.K., KUBOKAWA, T. and SALEH, A.K.M.E. (1989). The Stein paradox in the sense of the Pitman measure of closeness. *Ann. Statist.* 17, 1375-1386.

SEN, P.K. and SALEH, A.K.M.E. (1991). On Pitman closeness of Pitman estimators. *Gujarat Statist. Rev.*, in press.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* 1, 197-206.