Properties of Multivariate Cauchy and Poly-Cauchy Distributions with Bayesian g-Prior Applications

by

Guorui Bian and James M. Dickey*

Technical Report No. 567
October 1991

*The authors are grateful for helpful discussion with Kathryn Chaloner. Work of the second author was supported, in part, by NSF grant DMS-8911548.

Abstract

The sum of two independent Cauchy random variables is well-known to be distributed according to a (scaled) Cauchy distribution. This reproductive property is no longer tenable, however, in the general multivariate case; necessary and sufficient conditions are here investigated. A distribution whose density is proportional to a product of t-densities has been called a poly-t distribution, or double-t for two factors in the density. Simple mathematical forms for the normalizing constant and the lower-order moments of the multivariate double-Cauchy density are here derived when the two Cauchy-density factors satisfy the aforementioned conditions. These forms are then applied to obtain new Bayesian estimates of the multivariate normal location parameter and slope coefficients in the linear multiple regression sampling model with independent Cauchy-type g-prior distributions. The new estimators are adaptive, and differ substantially from the usual Bayesian estimates obtained using natural conjugate prior distributions.

1. Introduction

The well-known Cauchy distribution has heavy tails and possesses a reproductive property, whereby the sum of two independent scaled Cauchy random quantities is, still, a scaled Cauchy quantity. The Cauchy distribution is often employed in robustness studies. Having no mean or higher moments, it is often used to constructt counterexamples. The poly-Cauchy is a distribution with density proportional to a product of Cauchy densities. The double-Cauchy, with two Cauchy factors, is sometimes encountered in Bayesian inference for a normal population, for example, as the posterior distribution of the mean when it is prior distributed according to a Cauchy distribution with the unknown variance independently reciprocal chi-squared distributed. (For early work involving even more general multivariate-t and matrix-t factors, see Stein (1962), Lindley (1962), Anscombe (1963), Tiao and Zellner (1964), and Stone (1964), followed later by Dickey (1967, 1968, 1975) and others.) In the univariate case, Bian and Dickey (1990) found simple mathematical forms for the mean and variance of the double-Cauchy and used them to develop new Bayesian estimates of the normal location parameter under Cauchy-type prior distributions. Since the estimates are obtained from non-conjugate prior distributions, they exhibit new properties, including adaptability to the dispersion in the data, properties which do not appear in the usual Bayesian estimates based on conjugate or non-informative priors.

In this paper, we study properties of the multivariate Cauchy and the lower-order moments of the multivariate double-Cauchy. The multivariate Cauchy is a direct generalization of the univariate Cauchy, but in general, the reproductive property does not hold. Necessary and sufficient conditions for the property will be investigated here, in Section 2. The normalizing constant and lower moments of the multivariate double-Cauchy density, derived in Section 3, will have simple mathematical forms when the two Cauchy component densities satisfy the conditions. These are applied, in Sections 4 and 5, to Bayesian estimation of the multivariate normal location parameter and the coefficient parameter in multiple linear regression. The location or regression parameters are prior-distributed as multivariate Cauchy, independently of the dispersion component σ^2. In all these applications, the parameters of interest, the mean vector or regression coefficients, are posterior distributed as multivariate double-Cauchy.

2. Reproductive property of multivariate Cauchy distributions

We review briefly necessary formulae and concepts of the multivariate Cauchy distribution. Suppose the random vector $\underset{\sim}{Z} = (Z_1, \ldots, Z_p)'$ to be distributed multivariate normal $N(0, I_p)$ and Y distributed as a chi-square quantity with 1 degree of freedom, independently of $\underset{\sim}{Z}$.

Denote

$$\underset{\sim}{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} = Y^{-1/2}\,\underset{\sim}{Z} = \begin{bmatrix} Z_1 Y^{-1/2} \\ \vdots \\ Z_p Y^{-1/2} \end{bmatrix} \tag{2.1}$$

Then the distribution of $\underset{\sim}{X}$ is standard p-dimension multivariate Cauchy and its density function is

$$f_C(\underset{\sim}{x}; \underset{\sim}{0}, I_p) = c(p)\,(1 + \underset{\sim}{x}'\underset{\sim}{x})^{-(1+p)/2}, \tag{2.2}$$

where $c(p) = \Gamma[(p+1)/2]\, \pi^{-(p+1)/2}$ (Raiffa & Schlaifer 1961, p.256).

The characteristic function of $\underset{\sim}{X}$ is

$$\phi_C(\underset{\sim}{t};\ \underset{\sim}{0}, I_p) = E[\exp(\, i\underset{\sim}{t}'\underset{\sim}{X}\,)] = E(\,\exp[\,i \sum t_j Z_j\, Y^{-1/2}]\,).$$

Since $\sum t_j Z_j$ is distributed as $N(0, \|\underset{\sim}{t}\|)$, independently from Y, then $\sum(t_j Z_j) Y^{-1/2}$ is distributed as a scaled Cauchy. Thus,

$$\phi_C(\underset{\sim}{t}; \underset{\sim}{0}, I_p) = \exp(\,-\|\underset{\sim}{t}\|), \tag{2.3}$$

where $\|\underset{\sim}{t}\| = (\sum_{j=1}^{p} t_j^2)^{1/2}$. Further, if $\underset{\sim}{W}$ is the linear transformation

$$\underset{\sim}{W} = (V^{1/2})'\, \underset{\sim}{X} + \underset{\sim}{\mu}, \tag{2.4}$$

then the density of $\underset{\sim}{W}$ is

$$f_C(\underset{\sim}{w}; \underset{\sim}{\mu}, V) = c(p)[\det(V)]^{-1/2}\, [\,1 + (\underset{\sim}{w} - \underset{\sim}{\mu})'V^{-1}(\underset{\sim}{w} - \underset{\sim}{\mu})]^{-(1+p)/2}, \tag{2.5}$$

and the characteristic function of $\underset{\sim}{W}$ is

$$\phi_C(\underset{\sim}{t}; \underset{\sim}{\mu}, V) = \exp(\,i\underset{\sim}{t}'\underset{\sim}{\mu})\, \phi_C(V^{1/2}\underset{\sim}{t}; \underset{\sim}{0}, I_p) = \exp(\,i\underset{\sim}{t}'\underset{\sim}{\mu} - \|V^{1/2}\,\underset{\sim}{t}\|), \tag{2.6}$$

where $\|V^{1/2}\,\underset{\sim}{t}\| = (\underset{\sim}{t}'V\underset{\sim}{t})^{1/2}$. We call $\underset{\sim}{\mu}$ and V the center and dispersion matrix of the multivariate Cauchy random variable $\underset{\sim}{W}$, respectively. $\underset{\sim}{X}$ and $\underset{\sim}{W}$ do not have finite means and variances. The density function (2.5) and characteristic function (2.6) depend only on quadratic forms in their arguments, and the standard multivariate Cauchy distribution (2.2) is invariant under orthogonal transformations (Eaton 1983).

The reproductive property of the univariate Cauchy distribution is no longer automatic in the multivariate case, but when the dispersion matrices of the two Cauchy factors are proportional, the property holds. Before proving a theorem to this effect, we need the following lemma.

Lemma 1. (Bian 1989) Suppose both A and B are n x n semi-positive definite matrices, then there exists an n x n matrix C such that

$$(\underset{\sim}{x}'A\underset{\sim}{x})^{1/2}+(\underset{\sim}{x}'B\underset{\sim}{x})^{1/2} = (\underset{\sim}{x}'C\underset{\sim}{x})^{1/2}, \text{ for all } \underset{\sim}{x} \in R^n, \tag{2.7}$$

iff A is proportional to B.

Theorem 1. Suppose $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are independently p-dimensional standard Cauchy. Then the vector

$$Y = B_1\underset{\sim}{X}_1 + B_2\underset{\sim}{X}_2 \tag{2.8}$$

is distributed as a scaled Cauchy if and only if B_1B_1' is proportional to B_2B_2'.

Proof. Using (2.6), the characteristic functions of $B_1\underset{\sim}{X}_1$ and $B_2\underset{\sim}{X}_2$ respectively are $\phi_1(\underset{\sim}{t}) = \exp[-(\underset{\sim}{t}'B_1B_1'\underset{\sim}{t})^{1/2}]$ and $\phi_2(\underset{\sim}{t}) = \exp[-(\underset{\sim}{t}'B_2B_2'\underset{\sim}{t})^{1/2}]$. Since $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are independent, the characteristic function of Y is the product

$$\phi_y(\underset{\sim}{t}) = \phi_1(\underset{\sim}{t})\phi_2(\underset{\sim}{t}) = \exp(-[||B_1'\underset{\sim}{t}\,||+||B_2'\underset{\sim}{t}\,||]). \tag{2.9}$$

The distribution is determined by the characteristic function. Hence, Y is distributed as a scaled Cauchy if and only if there is a semi-positive definite matrix C for which

$$\phi_y(\underset{\sim}{t}) = \exp(-[||B_1'\underset{\sim}{t}\,||+||B_2'\underset{\sim}{t}\,||]) = \exp(-(\underset{\sim}{t}'C\,\underset{\sim}{t})^{1/2}). \tag{2.10}$$

This is equivalent to

$$(\underset{\sim}{t}'B_1B_1'\underset{\sim}{t})^{1/2} + (\underset{\sim}{t}'B_2B_2'\underset{\sim}{t})^{1/2} = (\underset{\sim}{t}'C\,\underset{\sim}{t})^{1/2}, \text{ for all } \underset{\sim}{t}. \tag{2.11}$$

By Lemma 1, (2.11) holds if and only if B_1B_1' is proportional to B_2B_2'.

3. Lower-order moments of poly-Cauchy distributions

The double-Cauchy density is proportional to a product of two Cauchy densities, i.e.

$$f(\underset{\sim}{x}) \propto f_C(\underset{\sim}{x}; \underset{\sim}{\mu}_1, V_1)\, f_C(\underset{\sim}{x}; \underset{\sim}{\mu}_2, V_2). \tag{3.1}$$

Simple mathematical forms for the normalizing constant, mean, and variance of a univariate poly-Cauchy density were found by Bian & Dickey (1990). As Theorem 1 has shown, the multivariate case is more complicated. In general, there is no closed form for the normalizing constant and lower-order moments. However, when the dispersion matrices of the Cauchy components are proportional, simple mathematical forms are available.

Lemma 2. (i) Suppose f is a measurable real function, so that the following integrals are meaningful. Then

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1^2+\ldots+x_n^2)dx_1\ldots dx_n = \frac{2\pi^{n/2}}{\Gamma(n/2)}\int_0^{\infty} f(y^2)y^{n-1}dy\ , \tag{3.2}$$

(ii) $$\int_0^{\infty} x^{n-1}\exp(-ax)\mathrm{Sin}(bx)dx = \Gamma(n)\ (a^2+b^2)^{-n/2}\mathrm{Sin}[n\ \mathrm{arctg}(b/a)] \tag{3.3}$$

(Gradsteyn & Ryzhik 1965, pp.620, 490).

Lemma 3.

$$\int_0^{\pi/2} \frac{\mathrm{Sin}x\ \mathrm{Cos}^{n-2}x}{(1+a^2\mathrm{Sin}^2x)^{n/2}}\ \mathrm{Sin}[n\ \mathrm{arctg}(a\mathrm{Sin}x)]\ dx$$

$$= a\ 2^{n-2}\ B[(n+1)/2,\ (n-1)/2]\ (1+a^2)^{-(n+1)/2}, \tag{3.4}$$

where B(,) is the beta function.

Proof. (i) Case of odd n.

Taking the change of variable Sin x = Sinφ/(y+bCosφ), where φ = arctgb, and using the identity

$$\int_0^{\infty} \frac{x^{k-1}}{(x^2+2bx\mathrm{Cos}\varphi+b^2)^{m/2}}\ \mathrm{Sin}[m\ \mathrm{arctg}(\frac{b\mathrm{Sin}\varphi}{(x+b\mathrm{Cos}\varphi)})]\ dx$$

$$= B(k,\ m-k)\ b^{m-k}\ \mathrm{Sin}(m-k)\varphi$$

(Grobner and Hofreiter 1950, p.184), we obtain

$$\text{LHS of (3.4)} = \mathrm{Sin}^2\varphi \int_0^{\infty} \frac{(y^2+2y\mathrm{Sin}\varphi)^{(n-3)/2}}{(y^2+2by\mathrm{Cos}\varphi+b^2)^{n/2}} \mathrm{Sin}[n\, \mathrm{arctg}(\frac{b\mathrm{Sin}\varphi}{(y+b\mathrm{Cos}\varphi)})]\, dy$$

$$= \mathrm{Sin}^2\varphi \sum_{j=0} \binom{(n-3)/2}{j} 2^j \mathrm{Sin}^j\varphi\, B(j+2, n-j-2)\, b^{j+2}\, \mathrm{Sin}(j+2)\varphi, \qquad (3.5)$$

where $\mathrm{Sin}\varphi = b\mathrm{Cos}\varphi = b(1+b^2)^{-1/2}$.

(ii) Case of even n.

Let $\mathrm{tg}y = b\mathrm{Sin}x$, then $\mathrm{Sin}y = b\mathrm{Sin}x/(1+b^2\mathrm{Sin}^2x)^{1/2}$ and $\mathrm{Cos}y = (1+b^2\mathrm{Sin}^2x)^{-1/2}$.

Since $\mathrm{Sin}\, ny = n\, \mathrm{Cos}^{n-1}y\, \mathrm{Sin}y - \binom{n}{3} \mathrm{Cos}^{n-3}y\, \mathrm{Sin}^3y + \binom{n}{5} \mathrm{Cos}^{n-5}y\, \mathrm{Sin}^5y - \ldots,$ (3.6)

and $\displaystyle\int_0^1 \frac{x^{k/2}(1-x)^{(n-3)/2}}{(1+b^2x)^n} dx$

$$= (1+b^2)^{-(k/2+1)} \sum_{j=0}^{(n-1-k)/2} (1+b^2)^{-j} \binom{(n-1-k)/2}{j} B[k/2+1+j, n-(k/2+1+j)],$$

if $n - k + 3$ is even (3.7)

(Gradsteyn and Ryzhik 1965 p.27 and Grobner and Hofreiter 1950 p.175), we obtain by substituting (3.6) into (3.4) and then taking the change of variable $z = \mathrm{Sin}^2x$,

$$\text{LHS of (3.4)} = \frac{1}{2} \int_0^1 \frac{(1-z)^{(n-3)/2}}{(1+b^2z)^n} [nbz^{1/2} - \binom{n}{3} b^3z^{3/2} + \binom{n}{5} b^5z^{5/2} - \ldots]\, dz$$

$$= (1/2) \sum_{m=0}^{[(n-1)/2]} (-1)^m \binom{n}{2m+1} b^{2m+1}(1+b^2)^{-(m+3/2)}$$

$$\cdot \sum_{j=0}^{n/2-m-1} (1+b^2)^{-j} \binom{n/2-m-1}{j} B[m+3/2+j, n-(m+3/2+j)]. \qquad (3.8)$$

The relation (3.4) can be verified by using the forms (3.5) and (3.8).

Theorem 2. Let $\underset{\sim}{X}$ be distributed according to the multivariate poly-Cauchy density,

$$f(\underset{\sim}{x}) = k[1+(\underset{\sim}{x} - \underset{\sim}{\mu}_1)'(\sigma_1{}^2V)^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu}_1)]^{-(1+p)/2}[1+(\underset{\sim}{x} - \underset{\sim}{\mu}_1)'(\sigma_2{}^2V)^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu}_1)]^{-(1+p)/2}. \tag{3.9}$$

Then

(i) the normalizing constant is

$$k = c(p)[\det(V)]^{-1/2}[(\sigma_1+\sigma_2)/\sigma_1\sigma_2]^p[1+ \| V^{-1/2}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2) \|^2/(\sigma_1+\sigma_2)^2)]^{(p+1)/2}, \tag{3.10}$$

(ii) the mean is

$$E(\underset{\sim}{X}) = (\sigma_1\underset{\sim}{\mu}_2+\sigma_2\underset{\sim}{\mu}_1)/(\sigma_1+\sigma_2) = (\underset{\sim}{\mu}_1/\sigma_1 + \underset{\sim}{\mu}_2/\sigma_2)/(\sigma_1{}^{-1} + \sigma_2{}^{-1}), \tag{3.11}$$

(iii) and if $\Sigma = E\{[\underset{\sim}{X} - E(\underset{\sim}{X})][\underset{\sim}{X} - E(\underset{\sim}{X})]'\}$ is the covariace matrix of $\underset{\sim}{X}$, then

$$\mathrm{Trace}(V^{-1}\Sigma) = \sigma_1\sigma_2[\ 1+ \| V^{-1/2}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2) \|^2/(\sigma_1+\sigma_2)^2)]. \tag{3.13}$$

Proof. (i) The normalizing constant k is found by Theorem 1, since the convolution of two Cauchy densities is a Cauchy density in this case.

The proofs of (ii) and (iii) are given in an Appendix.

4. Estimating the multivariate normal mean.

Statistical inference concerning the mean of a multinormal process has been studied deeply from the classical point of view (Stein 1962) and from the Bayesian point of view with the usual conjugate or non-informative priors (Raiffa & Schlaifer 1961, Press 1982). Further, Stone (1964) and Dickey (1968, 1971, 1975) considered Bayesian inference regarding multinormal sampling, using independent priors for mean and variance. But the problem of estimating the mean vector with these independent priors has not, itself, received much treatment, in part, because of the complication of the posterior distributions, as pointed out by Jeffreys (1948 p.123-124). Here, we consider Bayesian estimation of the multinormal mean, with an independent Cauchy-type prior distribution.

Theorem 3. Suppose that $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n$ are independently drawn from a p-dimension normal population $N(\underset{\sim}{\mu}, \sigma^2V)$ with unknown mean vector $\underset{\sim}{\mu}$ and variance component σ^2, and that $\underset{\sim}{\mu}$ and σ^2 are prior independently distributed according to a scaled Cauchy distribution and a reciprocal chi-square type distribution, respectively. In notation,

$$\underset{\sim}{\mu} = f_C(\underset{\sim}{\mu}; \underset{\sim}{m}_o, d_o^2 V),$$

$$f(\sigma^{-2}) \propto (\sigma^{-2})^{k/2-1} \exp(-\sigma^{-2} s_o^2 /2). \tag{4.1}$$

Then

(i) the joint posterior density is

$$f(\underset{\sim}{\mu}, \sigma^{-2} | \bar{x}, s_n) \propto (\sigma^{-2})^{(np+k)/2-1} f_C(\underset{\sim}{\mu}; \underset{\sim}{m}_o, d_o^2 V)$$

$$\cdot \exp\{-(\sigma^{-2}/2)[n||V^{-1/2}(\bar{x} - \underset{\sim}{\mu})|| + ns_n^2 + s_o^2\}, \tag{4.2}$$

(ii) the marginal posterior density of $\underset{\sim}{\mu}$ is a poly-t with density

$$f(\underset{\sim}{\mu} | \bar{x}, s_n) \propto f_C(\underset{\sim}{\mu}; \underset{\sim}{m}_o, d_o^2 V)[1 + n||V^{-1/2}(\bar{x} - \underset{\sim}{\mu})||/(ns_n^2 + s_o^2)]^{-(np+k)/2}, \tag{4.3}$$

(iii) furthermore, if $k = (1-n)p + 1$, the marginal posterior density of $\underset{\sim}{\mu}$ is a poly-Cauchy with the density

$$f(\underset{\sim}{\mu} | \bar{x}, s_n) \propto f_C(\underset{\sim}{\mu}; \underset{\sim}{m}_o, d_o^2 V)[1 + n||V^{-1/2}(\bar{x} - \underset{\sim}{\mu})||/(ns_n^2 + s_o^2)]^{-(p+1)/2}, \tag{4.4}$$

and the conditional predictive distribution of $\bar{x}$ for given s_n is Cauchy with the density

$$f(\bar{x} | s_n) \propto \left[1 + \frac{||V^{-1/2}(\bar{x} - \underset{\sim}{m}_o)||^2}{[(s_n^2 + s_o^2/n)^{1/2} + d_o]^2}\right]^{-(p+1)/2}, \tag{4.5}$$

where $\bar{x} = (1/n) \sum \underset{\sim}{x}_i$, and $s_n^2 = (1/n) \sum ||V^{-1/2}(\underset{\sim}{x}_i - \bar{x})||^2$.

<u>Proof</u>. (i) The density of the sufficient statistic $(\bar{x}, s_n^2)$ is

$$f(\bar{x}, s_n^2 | \underset{\sim}{\mu}, \sigma) \propto (\sigma^{-2})^{np/2} (s_n^2)^{(n-1)p/2-1} \exp\{-(n\sigma^{-2}/2)[\, ||V^{-1/2}(\bar{x} - \underset{\sim}{\mu})|| + s_n^2\,]\}.$$

The joint density of $\bar{x}$, s_n^2 and $\underset{\sim}{\mu}$, σ^{-2} is proportional to the product of the likelihood function and the prior density,

$$f(\bar{x}, s_n^2, \underset{\sim}{\mu}, \sigma^{-2}) \propto (s_n^2)^{(n-1)p/2-1} (\sigma^{-2})^{(np+k-2)/2}$$

$$\cdot f_C(\underset{\sim}{\mu}; \underset{\sim}{m}_o, d_o^2 V) \exp\{-(n\sigma^{-2}/2)[\, ||V^{-1/2}(\bar{x} - \underset{\sim}{\mu})|| + s_n^2\,]\}. \tag{4.6}$$

The joint posterior density of $\underset{\sim}{\mu}$ and σ^{-2} is proportional to the joint density (4.6), so that

$$f(\underset{\sim}{\mu}, \sigma^{-2} \mid \bar{x}, s_n) \propto (\sigma^{-2})^{(np+k-2)/2} f_C(\underset{\sim}{\mu}; \underset{\sim}{m}_o, d_o^2 V)$$

$$\cdot \exp\{-(\sigma^{-2}/2)[n\|V^{-1/2}(\bar{x} - \underset{\sim}{\mu})\| + ns_n^2 + s_o^2]\}.$$

(ii) The marginal density (4.3) is obtained immediately by integrating out σ^{-2} from the joint density (4.2).

(iii) The density (4.4) is a special case of (4.2) with $k = (1-n)p+1$. To derive (4.5), integrate out σ^{-2} from $f(\bar{x}, s_n^2, \underset{\sim}{\mu}, \sigma^{-2})$ specified by (4.6) with $k = (1-n)p+1$, to obtain

$$f(\bar{x}, s_n^2, \underset{\sim}{\mu}) \propto (s_n^2)^{(n-1)p/2-1} f_C(\underset{\sim}{\mu}; \underset{\sim}{m}_o, d_o^2 V)[n\|V^{-1/2}(\bar{x} - \underset{\sim}{\mu})\| + ns_n^2 + s_o^2]^{-(p+1)/2}$$

$$\propto (s_n^2)^{[(n-1)p-3]/2} f_C(\underset{\sim}{\mu}; \underset{\sim}{m}_o, d_o^2 V)\, s_n^{-p}[1 + n\|V^{-1/2}(\bar{x} - \underset{\sim}{\mu})\|/(ns_n^2 + s_o^2)]^{-(p+1)/2}.$$

Then integrating out $\underset{\sim}{\mu}$, since the convolution of two Cauchy densities is a Cauchy density, we obtain the predictive density,

$$f(\bar{x}, s_n^2) \propto (s_n^2)^{[(n-1)p-3]/2}[(s_n^2 + s_o^2/n)^{1/2} + d_o]^{-p}$$

$$\cdot \{1 + \|V^{-1/2}(\bar{x} - \underset{\sim}{m}_o)\|^2 / [(s_n^2 + s_o^2/n)^{1/2} + d_o]^2\}^{-(p+1)/2}. \qquad (4.7)$$

Density (4.5) is obtained by fixing s_n^2 in (4.7), and the theorem is proved.

Note that the posterior density of $\underset{\sim}{\mu}$, (4.3), is a poly-t. When the prior location $\underset{\sim}{m}_o$ is near the sample mean $\bar{x}$, the posterior density is unimodal, but when the prior location differs substantially from the sample mean, this means that the prior is somewhat inconsistent with the sample, and the posterior density of $\underset{\sim}{\mu}$ is a bimodal function. The modes and valley are located on the so-called "contract curve", the line segment connecting the prior location and the sample mean. Bimodality of the posterior density may suggest incoherency of the prior opinion with the sample. However, the bimodality would disappear under the more usual Bayesian inference even if the prior distribution is substantially inconsistent with the observed sample. This is because the conjugate prior stipulates a joint prior density having the same mathematical form as the likelihood function. This implies that the prior information can be parameterized by imaginary data, called "prior data", drawn from the same underlying population. In other words, the conjugate prior assumes that the prior distribution should be summarized by a statistic combining it in a standard way with the real observations.

Theorem 4. Suppose that the assumptions of Theorem 3 hold and that the prior parameter k is the function of sample size and dimension, $k = (n-1)p - 2$. Then the optimal estimate of μ, under quadratic loss, is the posterior mean

$$E(\mu \mid \bar{x}, s_n) = \frac{1}{d_o+(s_n^2+s_o^2/n)^{1/2}} \left[(s_n^2+s_o^2/n)^{1/2}\underset{\sim}{m}_o + d_o\bar{x}\right]$$

$$= \frac{1}{d_o^{-1}+(s_n^2+s_o^2/n)^{-1/2}} \left[d_o^{-1}\underset{\sim}{m}_o + (s_n^2+s_o^2/n)^{-1/2}\bar{x}\right]. \quad (4.8)$$

Proof. This expression is a direct conclusion of Theorem 2 and Theorem 3.

The new Bayesian estimate (4.7) is seen to be a weighted average of $\underset{\sim}{m}_o$, the prior location, and $\bar{x}$, the sample mean. The weights depend on the sample dispersion s_n^2 and the prior dispersion d_o^2. When the sample dispersion goes up, the uncertainty from the sample mean becomes larger. Thus the estimate moves closer to the prior location. The new estimate adjusts automatically to the sample dispersion. In this sense, it is adaptive. However, if the independent Cauchy-type prior (4.1) is replaced by the natural conjugate prior, the usual Bayesian estimate of μ obtains,

$$E_C(\mu \mid \bar{x}, s_n) = \frac{1}{(n_o+n)} \left[n_o\underset{\sim}{m}_o + n\bar{x}\right], \quad (4.9)$$

which is a weighted average too, but the weights are determined only by the sample size n and prior sample size n_o. (See Raiffa & Schlaifer 1961, p. 316-325). Comparing (4.8) with (4.9), we see that the new estimate (4.8) may differ substantially from the usual Bayesian estimate when the prior location is inconsistent with the sample mean. It is important to consider the effects of reasonable priors in Bayesian inference, especially in small sample cases.

5. Estimation in linear regression

We consider, in this section, estimates of the coefficient parameters in the normal multiple linear regression model (NLR) with the standard form

$$y = X\beta + e, \quad (5.1)$$

where y is an n x 1 vector of observations on the dependent variable, X is an n x p design matrix with rank p, β is a p x 1 vector of regression parameters with unknown value, and e is the n x 1 vector of disturbance or error items. It is assumed that the elements of e are independently drawn from a normal distribution with mean 0 and finite variance σ^2. The Bayesian analysis of the NLR with the natural conjugate or the usual non-informative prior distributions has appeared in Jeffreys (1948), Raiffa and Schlaifer (1961), Tiao and Zellner (1964), Box and Tiao (1973), and elsewhere. Zellner (1986) considered Bayesian inference regarding NLR with the normal g-prior specified by the following forms,

$$p(\beta, \sigma) \propto h(\sigma)\, f(\beta \mid \sigma, g), \qquad (5.2)$$

in which $f(\beta \mid \sigma, g) \propto \sigma^{-p} \exp\{-g(\beta - \beta_o)X'X(\beta - \beta_o)/2\sigma^2\}$

and $h(\sigma) \propto 1/\sigma$.

This is a special case of the natural conjugate prior. Based on this normal g-prior, the posterior density is

$$f(\beta, \sigma \mid y) \propto \sigma^{-(n+p+1)} \exp\{-[(y - X\beta)'(y - X\beta) + g(\beta - \beta_o)X'X(\beta - \beta_o)]/2\sigma^2\}. \qquad (5.3)$$

Hence, the Bayesian estimate of ß, under quadratic loss, is the posterior mean

$$E_C(\beta \mid y, g) = (\hat{\beta} + g\beta_o)/(1 + g), \qquad (5.4)$$

where $\hat{\beta}$ is the OLS estimate of ß,

$$\hat{\beta} = (X'X)^{-1} X'y. \qquad (5.5)$$

The Bayesian estimate (5.4) is a weighted average of $\hat{\beta}$, the OLS estimate, and β_o, the prior center of ß. The weights depend on the prior parameter g only. The Zellner normal g-prior distribution (5.2) assumes that there is little prior knowledge regarding the variance and that the prior knowledge regarding ß is dependent on the variance. In some practical cases, the prior knowledge regarding ß and σ may come from different sources and the information regarding ß is then a bit looser. In that case, it may be appropriate to describe as independent the prior knowledge regarding ß and σ^2 by assuming the Cauchy-type g-prior and a reciprocal chi-square distribution, respectively. In notation,

$$p(\beta, \sigma) \propto h(\sigma)\, f(\beta \mid g), \qquad (5.6)$$

where $f(\beta \mid g) \propto [1 + g(\beta - \beta_o)'(X'X)(\beta - \beta_o)]^{-(p+1)/2}$,

and $f(\sigma^{-2}) \propto (\sigma^{-2})^{k/2-1} \exp(-\sigma^{-2} s_o^2/2)$.

Theorem 5. Suppose the NLR model has the standard form (5.1) with the Cauchy-type g-prior specified by (5.6). Then

(i) the posterior density of β and σ^{-2} is

$$f(\beta, \sigma^{-2}| y) \propto (\sigma^{-2})^{(n+k)/2-1} [1+ g^{1/2} || (X'X)^{1/2}(\beta - \beta_o)||^2]^{-(p+1)/2}$$

$$\cdot \exp\{- [|| y -X\hat{\beta} ||^2 + || (X'X)^{1/2}(\beta - \hat{\beta})||^2+ s_o^2]/2\sigma^2\}, \quad (5.7)$$

(ii) the marginal posterior density of β is a poly-t with the density

$$f(\beta| y) \propto [1+ g^{1/2} || (X'X)^{1/2}(\beta - \beta_o)||^2]^{-(p+1)/2}$$

$$\cdot[|| (X'X)^{1/2}(\beta - \hat{\beta})||^2 + || y -X\hat{\beta} ||^2+ s_o^2]^{-(n+k)/2}. \quad (5.8)$$

The omitted proof is similar to that of Theorem 3.

The posterior density of β, (5.8), obtained using the independent Cauchy-type g-prior, is a poly-t, which may have two modes when the prior center of β is far away from the OLS estimate.

Theorem 6. Suppose that NLR has the standard form (5.1) with the Cauchy-type g-prior specified by (5.6) with the prior parameter k related to sample size and dimension by $k = p+1-n$. Then

(i) the marginal posterior density of β is a poly-Cauchy with density

$$f(\beta| y) \propto [1+ g^{1/2} || (X'X)^{1/2}(\beta - \beta_o)||^2]^{-(p+1)/2}$$

$$\cdot[|| (X'X)^{1/2}(\beta - \hat{\beta})||^2 + || y -X\hat{\beta} ||^2+ s_o^2]^{-(p+1)/2}, \quad (5.9)$$

(ii) the Bayesian estimate of β, under quadratic loss, is the posterior mean,

$$E(\beta| y, g) = w\hat{\beta} + (1-w)\beta_o, \quad (5.10)$$

where $w = (1 + g^{1/2}[|| y -X\hat{\beta} || + s_o^2]^{1/2})^{-1}$, and $\hat{\beta}$ is the OLS estimate of β, (5.5).

Proof. This is the direct conclusion of Theorem 2 and Theorem 5.

Both the usual Bayesian estimate (5.4) and the new Bayesian estimate (5.10) are weighted averages of the prior location and the OLS estimate. The weights in (5.4) depend only on the prior parameter g. However, the weights in (5.10) depend on the residual $\| y - X\hat{\beta} \|$ and the prior parameter g. The weight on $\hat{\beta}$ is a monotone decreasing function of the residual. Hence, the Bayesian estimate (5.10) is adjusted automatically by the residual $\| y - X\hat{\beta} \|$. The smaller the residual, the closer to the OLS estimate $\hat{\beta}$. In this sense, this is again an adaptive estimate. The two Bayesian estimates (5.4) and (5.10), which are obtained using different priors, may differ considerably, and both may differ considerably from the OLS estimate when the OLS estimate is far from the prior location.

6. Conclusions

Use of an independent Cauchy prior can lead to a posterior mean for the multinormal mean vector or the regression coefficient vector that is an adaptive weighted average of the prior location and the classical estimate. The weights depend on the sample dispersion, so the estimates are adjusted automatically by the dispersion of observations. In this sense, they are adaptive. These estimates may differ substantially from the usual Bayesian estimates obtained using the natural conjugate prior distributions. From these results, it appears that investigating Bayesian inferences with more realistic prior distributions and considering the effects of realistic prior distributions on the statistical inference can be important in theory and practice.

References

Anscombe, F. J. (1963), Bayesian inference concerning many parameters with reference to supersaturated designs. Bull. Int. Statist. Inst., 40, 721-33.

Bian, G. (1989), Bayesian statistical analysis with independent bivariate priors for the normal location and scale parameters. Unpublished Ph.D. thesis, University of Minnesota.

Bian, G. and J. M. Dickey (1990), Moments of the product-Cauchy density and Bayesian applications. Tech. Report No.552, School of Statistics, University of Minnesota.

Box, G. E. P. and G. C. Tiao (1973), Bayesian Inference in Statistical Analysis. Addision-Wesley.

Dickey, J. M. (1967), Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. Ann. Math. Statist. 38, pp. 511-18.

Dickey, J. M. (1968), Three multidimensional-integral identities with Bayesian applications. Ann. Math. Statist. 39, pp. 1615-27.

Dickey, J. M. (1971), The weighted likelihood ratio, linear hypotheses on normal location parameter. Ann. of Math. Statist., 42, pp. 204-23

Dickey, J. M. (1975), Bayesian alternatives to the F-test and least-squares estimate in the normal linear model. In "Studies in Bayesian Econometrics and Statistics" edited by S.E.Fienberg and A.Zellner, North-Holland, Amsterdam. pp. 515-54.

Eaton, M. L. (1983), Multivariate Statistics: A Vector Space Approach. Wiley & Sons, New York.

Gradshteyn, I. S. and I. M. Ryzhik (1980), Table of Integrals, Series, and Products. Academic Press, New York.

Grobner, W. and N. Hofreiter (1950), Integraltafel, Zweiter Teil, Bestimmte Integrale. Springer Verlag, Wien and Innsbruck.

Jeffereys, H. (1948), Theory of Probability. 2nd edition, Oxford, Clarendon Press.

Lindley, D. V. (1962), Discussion to C. M. Stein's, Confidence sets for the mean of a multivariate normal distribution. J. Roy. Statist. Soc.B 24, pp. 285-87.

Press, S. J. (1972), Applied Multivariate Analysis. Holt, Rinehart & Winston, New York.

Raiffa, H. and R. Schlaifer (1961), Applied Statistical decision Theory. Harvard University, Boston.

Stein, C. M. (1962), Confidence sets for the mean of a multivariate normal distribution. J. Roy. Statist. Soc.B 24, pp. 265-96.

Stone, M. (1964), Comments on a posterior distribution of Geisser and Cornfield. J. Roy. Statist. Soc.B 26, pp. 274-76.

Tiao, G. C., and A. Zellner (1964), Bayes's theorem and the use of prior knowledge in regression analysis. Biometrika, 51 pp. 233-43.

Zellner, A. (1986), On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In "Bayesian Inference and Decision Techniques" edited by P. Goel and A. Zellner, North-Holland, Amsterdam, pp. 233-43.

Appendix - The proof of Parts (ii) and (iii) of Theorem 2

(ii) For convenience, take the transformation

$$\underset{\sim}{Y} = \sigma_1^{-1} V^{-1/2} (\underset{\sim}{X} - \mu_1), \qquad (3.14)$$

and denote $\sigma = \sigma_2/\sigma_1$, and $\underset{\sim}{\mu} = \sigma_1^{-1}V^{-1/2}(\underset{\sim}{\mu}_2 - \underset{\sim}{\mu}_1)$. Then the density function of $\underset{\sim}{Y}$ is reduced to

$$f(\underset{\sim}{y}) = k_1[1 + \| \underset{\sim}{y} \|^2]^{-(1+p)/2}[1 + \sigma^{-2}(\underset{\sim}{y} - \underset{\sim}{\mu})'(\underset{\sim}{y} - \underset{\sim}{\mu})]^{-(1+p)/2}, \qquad (3.15)$$

where $k_1 = c(p)[(1+\sigma)/\sigma]^p[1 + \| \underset{\sim}{\mu} \|^2/(1+\sigma)^2)]^{(p+1)/2}$.

Because the Cauchy distribution is invariant under orthogonal transformations, without losing the generality, assume $\underset{\sim}{\mu} = \underset{\sim}{\mu}_o = (\mu, 0,\ldots,0)'$. Thus, for the last p-1 means, obtain

$$E(Y_j) = 0, \quad \text{for } j = 2, \ldots, p. \qquad (3.16)$$

The mean of Y_1 is

$$E(Y_1) = k_1 \int y_1[1+\| \underset{\sim}{y} \|^2]^{-(1+p)/2}[1+ \sigma^{-2}(\underset{\sim}{y} - \underset{\sim}{\mu}_o)'(\underset{\sim}{y} - \underset{\sim}{\mu}_o)]^{-(1+p)/2}d\underset{\sim}{y}$$

$$= c^{-1}(p)(1+\sigma)^p[1+ \| \underset{\sim}{\mu} \|^2/(1+\sigma)^2)]^{(p+1)/2}h(\underset{\sim}{\mu}_o), \qquad (3.17)$$

where

$$h(\underset{\sim}{\mu}) = c^2(p)\sigma^{-p} \int y_1[1+\| \underset{\sim}{y} \|^2]^{-(1+p)/2}[1+ \sigma^{-2}(\underset{\sim}{y} - \underset{\sim}{\mu})'(\underset{\sim}{y} - \underset{\sim}{\mu})]^{-(1+p)/2}d\underset{\sim}{y}.$$

To calculate $h(\underset{\sim}{\mu})$, take the Fourier transform and use the convolution formula, to obtain

$$g(\underset{\sim}{t}) = \int h(\underset{\sim}{\mu})\exp(i\underset{\sim}{\mu}'\underset{\sim}{t})d\underset{\sim}{\mu} = \psi(\underset{\sim}{t})\,\phi(\sigma\underset{\sim}{t}), \qquad (3.18)$$

where

$$\phi(\sigma\underset{\sim}{t}) = \int c(p)\sigma^{-p}[1+\sigma^{-2}\| \underset{\sim}{y} \|^2]^{-(1+p)/2}\exp(i\underset{\sim}{y}'\underset{\sim}{t})d\underset{\sim}{y} = \exp(-\sigma\| \underset{\sim}{t} \|), \qquad (3.19)$$

and $$\psi(\underset{\sim}{t}) = \int c(p)y_1[1+\| \underset{\sim}{y} \|^2]^{-(1+p)/2}\exp(i\underset{\sim}{y}'\underset{\sim}{t})d\underset{\sim}{y}.$$

We express $\psi(\underset{\sim}{t})$ as

$$\psi(\underset{\sim}{t}) = -i \int c(p)[1+\| \underset{\sim}{y} \|^2]^{-(1+p)/2}\frac{\partial}{\partial t_1}\exp(i\underset{\sim}{y}'\underset{\sim}{t})\,d\underset{\sim}{y}$$

$$= -i \int \frac{\partial}{\partial t_1}\{c(p)[1+ \| \underset{\sim}{y} \|^2]^{-(1+p)/2}\exp(i\underset{\sim}{y}'\underset{\sim}{t})\}\,d\underset{\sim}{y}. \qquad (3.20)$$

Because the integral (3.20) exists, the order of integration and differentiation is interchangeable. Therefore,

$$\psi(\underset{\sim}{t}) = -i\frac{\partial}{\partial t_1}\int c(p)[1+\| \underset{\sim}{y} \|^2]^{-(1+p)/2}\exp(i\underset{\sim}{y}'\underset{\sim}{t})\, d\underset{\sim}{y}$$

$$= -i\frac{\partial}{\partial t_1}\exp(-\| \underset{\sim}{t} \|) = i\frac{| t_1 |}{\| \underset{\sim}{t} \|}\exp(-\| \underset{\sim}{t} \|). \qquad (3.21)$$

Substituting (3.19) and (3.21) into (3.18), we have

$$g(\underset{\sim}{t}) = i\frac{| t_1 |}{\| \underset{\sim}{t} \|}\exp(-(\sigma+1)\| \underset{\sim}{t} \|).$$

Taking the inverse Fourier transformation, we have

$$h(\underset{\sim}{\mu}) = (2\pi)^{-p}\int g(\underset{\sim}{t})\exp(-i\underset{\sim}{\mu}'\underset{\sim}{t}))\, d\underset{\sim}{t}$$

$$= i(2\pi)^{-p}\int \frac{| t_1 |}{\| \underset{\sim}{t} \|}\exp[-(\sigma+1)\| \underset{\sim}{t} \|]\exp(-i\underset{\sim}{\mu}'\underset{\sim}{t})\, d\underset{\sim}{t}.$$

When $\underset{\sim}{\mu}_o = (\mu, 0,\ldots,0)'$,

$$h(\underset{\sim}{\mu}_o) = i(2\pi)^{-p}\int \frac{| t_1 |}{\| \underset{\sim}{t} \|}\exp[-(\sigma+1)\| \underset{\sim}{t} \|]\exp(-i\mu t_1))\, d\underset{\sim}{t}.$$

Use (3.2) to reduce the above integral to

$$h(\underset{\sim}{\mu}_o) = i\, 2^{-p+1}\pi^{-(p+1)/2}\Gamma^{-1}[(p-1)/2]$$

$$\cdot\int_0^\infty\int_{-\infty}^\infty t\exp(-i\mu t)\, y^{p-2}(t^2+y^2)^{-1/2}\exp[-(\sigma+1)(t^2+y^2)^{1/2}]\, dt\, dy$$

$$= 2^{-(p-2)}\pi^{-(p+1)/2}\Gamma^{-1}[(p-1)/2]\, I, \qquad (3.22)$$

where $$I = \int_0^\infty\int_0^\infty t\, y^{p-2}(t^2+y^2)^{-1/2}\exp[-(\sigma+1)(t^2+y^2)^{1/2}]\mathrm{Sin}(\mu t)\, dt\, dy.$$

To calculate I, change variables, t = z Sinx, and y = z Cosx, and then use (3.3) and (3.4),

$$I = \int_0^{\pi/2} \int_0^{\infty} z^{p-1} \exp[-(1+\sigma)z] \text{Sinx Cos}^{p-2}x \text{ Sin}(\mu z \text{Sinx})\, dz\, dx$$

$$= \Gamma(p) \int_0^{\pi/2} \text{Sinx Cos}^{p-2}x\, [(1+\sigma)^2+\mu^2\text{Sin}^2x]^{-p/2} \text{ Sin}[p \text{ arctg}(\frac{\mu \text{Sinx}}{1+\sigma})]\, dx$$

$$= \Gamma(p)\mu(1+\sigma)^{-(p+1)} 2^{p-2} B[(p-1)/2,(p+1)/2]\, [1+\mu^2/(1+\sigma)^2]^{-(p+1)/2}. \qquad (3.23)$$

Substituting (3.23) and (3.22) into (3.17), obtain

$$h(\underset{\sim}{\mu}_o) = \pi^{-(p+1)/2}\, \Gamma[(p+1)/2]\, \mu(1+\sigma)^{-(p+1)}\, [1+\mu^2/(1+\sigma)^2]^{-(p+1)/2},$$

and $\quad E(Y_1) = \mu/(1+\sigma). \qquad (3.24)$

By combining (3.24) with (3.16), the mean of $\underset{\sim}{Y}$ is

$$E(\underset{\sim}{Y}) = (1+\sigma)^{-1}\, \underset{\sim}{\mu}. \qquad (3.25)$$

Using the transform (3.14), the expectation of $\underset{\sim}{X}$ is seen to be

$$E(\underset{\sim}{X}) = \underset{\sim}{\mu}_1 + \sigma_1 V^{1/2} E(\underset{\sim}{Y})$$

$$= \underset{\sim}{\mu}_1 + \sigma_1 V^{1/2} (1+\sigma)^{-1}\, \underset{\sim}{\mu} = (\sigma_1\underset{\sim}{\mu}_2+\sigma_2\underset{\sim}{\mu}_1)/(\sigma_1+\sigma_2).$$

(iii) Similarly to the proof of Part (ii), consider

$$E(\| \underset{\sim}{Y} \|) = k_1 \int \| \underset{\sim}{y} \|^2\, [1+\| \underset{\sim}{y} \|^2]^{-(1+p)/2} [1 + \sigma^{-2} (\underset{\sim}{y} - \underset{\sim}{\mu}_o)'(\underset{\sim}{y} - \underset{\sim}{\mu}_o)]^{-(1+p)/2} d\underset{\sim}{y}$$

$$= c^{-1}(p)(1+\sigma)^p[1+ \| \underset{\sim}{\mu} \|^2/(1+\sigma)^2)]^{(p+1)/2} h_1(\underset{\sim}{\mu}_o), \qquad (3.26)$$

where

$$h_1(\underset{\sim}{\mu}) = c^2(p)\sigma^{-p} \int \| \underset{\sim}{y} \|^2\, [1+\| \underset{\sim}{y} \|^2]^{-(1+p)/2} [1+\sigma^{-2} (\underset{\sim}{y} - \underset{\sim}{\mu})'(\underset{\sim}{y} - \underset{\sim}{\mu})]^{-(1+p)/2} d\underset{\sim}{y}.$$

To calculate $h_1(\underset{\sim}{\mu})$, take the Fourier transform and use the convolution formula, to obtain

$$g_1(\underset{\sim}{t}) = \int h_1(\underset{\sim}{\mu}) \exp(i\underset{\sim}{\mu}'\underset{\sim}{t})\, d\underset{\sim}{\mu} = \psi_1(\underset{\sim}{t})\, \phi(\sigma\underset{\sim}{t}), \tag{3.27}$$

where

$$\phi(\sigma\underset{\sim}{t}) = \exp(-\sigma\| \underset{\sim}{t} \|) \tag{3.28}$$

and

$$\psi_1(\underset{\sim}{t}) = \int c(p)\| \underset{\sim}{y} \|^2 [1+\| \underset{\sim}{y} \|^2]^{-(p+1)/2} \exp(i\underset{\sim}{y}'\underset{\sim}{t})d\underset{\sim}{y}$$

$$= \int c(p)[1+\| \underset{\sim}{y} \|^2]^{-(p-1)/2} \exp(i\underset{\sim}{y}'\underset{\sim}{t})d\underset{\sim}{y} - \exp(-\sigma\| \underset{\sim}{t} \|). \tag{3.29}$$

Using (3.2), the first term of the right hand side of $\psi_1(\underset{\sim}{t})$ is expressed as

$$I = \frac{4c(p)}{c(p-2)} \int_0^\infty \int_0^\infty z^{p-2}(1+z^2+x^2)^{-(p-1)/2} \exp(ix\| \underset{\sim}{t} \|)\, dxdz$$

$$= \frac{4c(p)}{c(p-2)} \int_0^\infty \int_0^{\pi/2} u^{p-1}(1+u^2)^{-(p-1)/2} \mathrm{Sin}^{p-2}w\, \mathrm{Cos}(u\| \underset{\sim}{t} \|\mathrm{Cos}w)\, dudw.$$

Since $\displaystyle\int_0^{\pi/2} \mathrm{Sin}^{p-2}w\, \mathrm{Cos}(u\| \underset{\sim}{t} \|\mathrm{Cos}w)\, dw = \frac{\pi^{1/2}2^{p/2-2}}{(\| \underset{\sim}{t} \|u)^{p/2-1}}\, \Gamma(\frac{p-1}{2})\, J_{p/2-1}(u\| \underset{\sim}{t} \|)$,

and $\displaystyle\int_0^\infty u^{p/2}(1+u^2)^{-(p-1)/2} J_{p/2-1}(u\| \underset{\sim}{t} \|)du = \frac{\pi^{1/2}\| \underset{\sim}{t} \|^{p/2-2}}{2^{p/2-1}\exp(\| \underset{\sim}{t} \|)\Gamma((p-1)/2)}$,

where $J_k(.)$ is the Bessel function (Gradsteyn & Ryzhik 1965, p.403, p.686), we have

$$I = (p-1)\, \| \underset{\sim}{t} \|^{-1} \exp(-\| \underset{\sim}{t} \|),$$

and

$$\psi_1(\underset{\sim}{t}) = [(p-1)\, \| \underset{\sim}{t} \|^{-1} - 1]\exp(-\| \underset{\sim}{t} \|). \tag{3.30}$$

Substituting (3.30) and (3.28) into (3.27), obtain

$$g_1(\underset{\sim}{t}) = [(p-1)\, \| \underset{\sim}{t} \|^{-1} - 1]\exp(-(1+\sigma)\| \underset{\sim}{t} \|). \tag{3.31}$$

Now, taking the inverse Fourier transformation for $g_1(\underset{\sim}{t})$,

$$h_1(\underset{\sim}{\mu}) = (2\pi)^{-p}\int g_1(\underset{\sim}{t}) \exp(-i\underset{\sim}{\mu}'\underset{\sim}{t})\, d\underset{\sim}{t}$$

$$= (2\pi)^{-p}\int [(p-1)\| \underset{\sim}{t} \|^{-1}-1] \exp(-(\sigma+1)\| \underset{\sim}{t} \|) \exp(-i\underset{\sim}{\mu}'\underset{\sim}{t})\, d\underset{\sim}{t}$$

$$= (2\pi)^{-p}\int (p-1)\| \underset{\sim}{t} \|^{-1} \exp(-(\sigma+1)\| \underset{\sim}{t} \|) \exp(-i\underset{\sim}{\mu}'\underset{\sim}{t})\, d\underset{\sim}{t}$$

$$- c(p)(1+\sigma)^{-p}[1+\| \underset{\sim}{\mu} \|^2/(1+\sigma)^2)]^{-(p+1)/2}. \tag{3.32}$$

Cauculating the first term of the right hand side of (2.32),

$$I_1 = (p-1)(2\pi)^{-p}\int \| \underset{\sim}{t} \|^{-1}\exp(-(\sigma+1)\| \underset{\sim}{t} \|)\exp(-i\underset{\sim}{\mu}'\underset{\sim}{t})\, d\underset{\sim}{t}$$

$$= c^{-1}(p-2)(p-1)(2\pi)^{-p}\iint z^{p-2}(z^2+x^2)^{-1/2}\exp[-(\sigma+1)(z^2+x^2)]\exp(iy\| \underset{\sim}{\mu} \|)\, dxdz$$

$$= c^{-1}(p-2)(p-1)(2\pi)^{-p}\int_0^{\infty}\int_0^{\pi/2} u^{p-2}\exp[-(1+\sigma)u]\mathrm{Sin}^{p-2}w\ \mathrm{Cos}(u\| \underset{\sim}{\mu} \|\mathrm{Cos}w)\, dudw$$

$$= \frac{p-1}{(2\pi)^{p/2}\| \underset{\sim}{\mu} \|^{p/2-1}}\int_0^{\infty} u^{p/2-1}\exp[-(1+\sigma)u]J_{p/2-1}(u\| \underset{\sim}{\mu} \|)du\ .$$

Since $\int_0^{\infty} u^{p/2-1}\exp[-(1+\sigma)u]J_{p/2-1}(u\| \underset{\sim}{\mu} \|)du$

$$= \frac{2^{p/2-1}\| \underset{\sim}{\mu} \|^{p/2-1}}{\pi^{1/2}[1+\| \underset{\sim}{\mu} \|^2/(1+\sigma)^2]^{(p-1)/2}}\Gamma(\frac{p-1}{2})$$

(Gradsteyn & Ryzhik 1965, p.1145), we have

$$I_1 = c(p)\ [1+\| \underset{\sim}{\mu} \|^2/(1+\sigma)^2]^{-(p-1)/2}. \tag{3.33}$$

Substituting (3.23), (2.32) into (2.26), obtain

$$E(\| \underset{\sim}{Y} \|) = \| \underset{\sim}{\mu} \|^2/(1+\sigma) + \sigma. \tag{3.34}$$

Using the transform (3.14),

$$E(\| V^{-1/2}(\underset{\sim}{X}-\underset{\sim}{\mu}_1) \|^2) = E[(\underset{\sim}{X}-\underset{\sim}{\mu}_1)'V^{-1}(\underset{\sim}{X}-\underset{\sim}{\mu}_1)]$$

$$= \sigma_1\sigma_2 + \frac{\sigma_1 \| V^{-1/2}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2) \|^2}{\sigma_1+\sigma_2}. \qquad (3.35)$$

Thus $\quad \mathrm{Trace}(V^{-1}\Sigma) = E(\| V^{-1/2}(\underset{\sim}{X}-\underset{\sim}{\mu}_1) \|^2) - \| V^{-1/2}(E(\underset{\sim}{X}) - \underset{\sim}{\mu}_1) \|^2)$

$$= \sigma_1\sigma_2[\, 1 + \| V^{-1/2}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2) \|^2/(\sigma_1+\sigma_2)^2)].$$