Moments of the Product-Cauchy Density and Bayesian Applications

by

Guorui Bian and James M. Dickey*

Technical Report No. 552
August 1990

*Research supported in part by NSF Grant DMS-8911548.

Moments of the Product-Cauchy Density and Bayesian Applications

by

Guorui Bian and James M. Dickey*

Abstract

In general, the moments of a product-t or poly-t density are not available in closed form. However, as a special case, the mean and variance of the product-Cauchy density have simple mathematical forms, which are derived here and applied to Bayesian estimation for normal or Cauchy locations from very small samples.

1. Basic result

Theorem. Let X be distributed according to the product-Cauchy density,

$$f(x) = k[1+ (x-\mu_1)^2/s_1^2]^{-1} [1+ (x-\mu_2)^2/s_2^2]^{-1}, \quad -\infty< x <\infty. \tag{1}$$

Then the normalizing constant, the mean, and the variance are, respectively,

(i) $$k = \frac{s_1+s_2}{\pi s_1 s_2}[1+\frac{(\mu_1-\mu_2)^2}{(s_1+s_2)^2}], \tag{2}$$

(ii) $$E\,X = \frac{s_1\mu_2+s_2\mu_1}{s_1+s_2} = (s_1^{-1} + s_2^{-1})^{-1} [\frac{\mu_1}{s_1}+\frac{\mu_2}{s_2}], \tag{3}$$

and

(iii) $$\text{Var}\,X = s_1 s_2[1+\frac{(\mu_1-\mu_2)^2}{(s_1+s_2)^2}]. \tag{4}$$

Proof. (i) The normalizing constant k is easily obtained, since the convolution of two Cauchy densities is a Cauchy density. (ii) To calculate the mean, write

$$E\,X = \frac{s_1+s_2}{\pi s_1 s_2}[1+\frac{(\mu_1-\mu_2)^2}{(s_1+s_2)^2}] \int \frac{x}{[1+ (x-\mu_1)^2/s_1^2][1+ (x-\mu_2)^2/s_2^2]} dx$$
$$= \mu_1 + \pi(s_1+s_2)[1+\frac{\mu^2}{(s_1+s_2)^2}]\, h(\mu), \tag{5}$$

* Reseach supported in part by National Science Foundation Grant DMS-8911548.

where $\mu = \mu_2 - \mu_1$ and

$$h(\mu) = \int \frac{y}{\pi^2 s_1 s_2[1+ y^2/s_1^2]\,[1+ (y-\mu)^2/s_2^2]}\, dy.$$

In order to calculate $h(\mu)$, take the Fourier transformation and use the convolution formula, to obtain

$$g(t) = \int h(\mu)\, e^{i\mu t}\, d\mu = \psi(t)\, \phi(s_2 t), \qquad (6)$$

where $\phi(s_2 t) = \int \frac{1}{\pi s_2[1+ y^2/s_2^2]} e^{iyt}\, dy = e^{-s_2|t|}$

and $\psi(t) = \int \frac{y}{\pi s_1[1+ y^2/s_1^2]} e^{iyt}\, dy.$

$\psi(t)$ can be easily expressed as

$$\psi(t) = -i \int \frac{1}{\pi s_1[1+ y^2/s_1^2]} \frac{\partial}{\partial t} e^{iyt}\, dy$$

$$= -i \int \frac{\partial}{\partial t} \left(\frac{1}{\pi s_1[1+ y^2/s_1^2]} e^{iyt}\right) dy. \qquad (7)$$

Because the integral (7) exists, the order of integration and differentiation is exchangeable. Therefore,

$$\psi(t) = -i \frac{\partial}{\partial t} \int \frac{1}{\pi s_1[1+ y^2/s_1^2]} e^{iyt}\, dy = i s_1 e^{-s_1 t}, \text{ if } t \geq 0,$$

and $\psi(-t) = \psi(t)$.

If $\psi(t)$ is substituted into (6), then $g(t)$ can be expressed as an exponential function of t,

$$g(t) = i s_1 e^{-(s_1+s_2)t}, \text{ if } t \geq 0,$$

and $g(-t) = g(t)$.

Taking the inverse Fourier transformation, we have

$$h(\mu) = \frac{1}{2\pi} \int g(t) e^{-i\mu t} dt = \frac{s_1}{\pi} \int_0^\infty \sin(\mu t) e^{-(s_1+s_2)t}\, dt$$

$$= \frac{s_1}{\pi} \frac{\mu}{\mu^2+(s_1+s_2)^2} .$$

Then, substituting into the expression (5), obtain

$$EX = \mu_1+\pi(s_1+s_2)[1+\frac{(\mu_1-\mu_2)^2}{(s_1+s_2)^2}] \frac{s_1}{\pi} \frac{\mu_2-\mu_1}{(\mu_1-\mu_2)^2+(s_1+s_2)^2}$$

$$= \mu_1+ \frac{s_1(\mu_2-\mu_1)}{s_1+s_2} = (s_2\mu_1+s_1\mu_2)/(s_1+s_2).$$

To obtain the variance (iii), we work first with an uncentered moment,

$$E(X-\mu_1)^2 = \pi(s_1+s_2)[1+ \frac{(\mu_1-\mu_2)^2}{(s_1+s_2)^2}] \int \frac{(x-\mu_1)^2}{\pi^2 s_1 s_2[1+(x-\mu_1)^2/s_1^2][1+(x-\mu_2)^2/s_2^2]} dx$$

$$= \pi(s_1+s_2)[1+ \frac{(\mu_1-\mu_2)^2}{(s_1+s_2)^2}] \int \frac{x^2}{\pi^2 s_1 s_2[1+x^2/s_1^2][1+[x-(\mu_2-\mu_1)]^2/s_2^2]} dx .$$

Writing in the numerator of the integral $x^2 = s_1^2[(1+x^2/s_1^2) - 1]$, obtain

$$E(X-\mu_1)^2 = s_1^2\{\pi(s_1+s_2)[1+\frac{(\mu_1-\mu_2)^2}{(s_1+s_2)^2}]\int\frac{1}{\pi^2 s_1 s_2[1+[x-(\mu_2-\mu_1)]^2/s_2^2]}dx - 1\}$$

$$= s_1\{(s_1+s_2)[1+\frac{(\mu_1-\mu_2)^2}{(s_1+s_2)^2}] - s_1\} = \frac{s_1(\mu_1-\mu_2)^2}{s_1+s_2} + s_1 s_2.$$

Hence

$$\mathrm{Var}(X) = E[(X-\mu_1)^2] - [EX-\mu_1]^2 = s_1 s_2\,[1+\frac{(\mu_1-\mu_2)^2}{(s_1+s_2)^2}].$$

2. Applications

We consider applications of the theorem in Bayesian estimation of location based on very few observations. In the cases here, the posterior density of the interesting location parameter is a product-Cauchy density.

Example 1. Normal population with unknown mean and variance.

(1) One observation

Suppose we have only one observation, x, drawn from a normal population $N(\mu, \sigma^2)$, and μ and σ^2 are prior independently distributed as a translated and scaled Cauchy and a scaled reciprocal chi-square distribution on one degree of freedom, respectively, i.e.

$$\mu \sim x_o + d_o Z,$$
$$\sigma^2 \sim s_o^2/\chi_1^2, \qquad (8)$$

in which Z has the standard Cauchy density $\pi^{-1}(1+z^2)^{-1}$. This kind of independent proper prior model was discussed by Stone (1965) and Dickey (1975). Under these assumptions, the marginal posterior density of μ is the product-Cauchy,

$$f(\mu|\, x) \propto [1+(\mu-x_o)^2/d_o^2]^{-1}\,[1+(\mu-x)^2/s_o^2]^{-1}. \qquad (9)$$

For a quadratic loss, the Bayesian estimate of μ is the posterior mean, which is obtained by the theorem as the weighted average,

$$E(\mu|\, x) = (d_o x + s_o x_o)/(d_o+s_o). \qquad (10)$$

The posterior variance is

$$\mathrm{Var}(\mu|\, x) = d_o s_o[1+\frac{(x-x_o)^2}{(d_o+s_o)^2}]. \qquad (11)$$

By way of contrast, if the prior distribution is the usual (dependent) conjugate prior having the same marginal distributions as (8), then the posterior distribution of μ is a scaled and translated Student-t with 2 degrees of freedom, and the estimate is

$$E_C(\mu|\, x) = (d_o^2 x + s_o^2 x_o)/(d_o^2+s_o^2) \qquad (12)$$

Comparing $E(\mu | x)$ to $E_C(\mu | x)$ and the classical estimate $\hat{\mu} = x$, we find that the difference between any two of those three estimates can be appreciable when x is relatively large. In addition, the posterior variance of μ under the independent prior (8) is a finite value specified by (11). Whereas, under the corresponding conjugate prior, the posterior variance of μ does not exist. This suggests that the posterior tails of μ under these two priors can differ in important respects.

(2) Two observations.

Suppose we have just two observations x_1 and x_2 drawn from $N(\mu, \sigma^2)$, μ is prior distributed as the Cauchy density with center x_0 and scale d_0, and $\log(\sigma)$ has an independent locally uniform prior, $f(\sigma) \propto 1/\sigma$. Then the marginal posterior density of μ is the product-Cauchy,

$$f(\mu | x_1, x_2) \propto [1+(\mu-x_0)^2/d_0^2]^{-1} [1+2(\mu-\bar{x})^2/s^2]^{-1}, \quad (13)$$

where $\bar{x} = (x_1+x_2)/2$ and $s^2 = (x_1-x_2)^2/2$. Hence, the posterior mean would be the weighted average,

$$E(\mu | x_1, x_2) = w\,\bar{x} + (1-w)\,x_0. \quad (14)$$

where $w = 2d_0/(2d_0+|x_1-x_2|)$.

This shows that when the two observations are close together, the Bayesian estimate of μ is near the sample mean, and when $|x_1-x_2|$ is large, the estimate is near the prior mean. The weight on $\bar{x}$ is a decreasing function of the distance between these two observations. Thus the estimate is a nonlinear function of the observations and the prior center x_0, and hence an adaptive weighted average of the sample mean and prior center.

Note that when d_0 goes to infinity, the prior goes to the usual Jeffreys' (1961) improper prior, and the posterior distribution (13) goes to Jeffreys' posterior distribution, identical to Fisher's (1935) fiducial distribution.

Example 2. Behrens-Fisher problem.

(1) One observation from each population.

Suppose we have two observations, x_1 and x_2, drawn from two normal populations, $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively, and μ_1, μ_2, σ_1^2 and σ_2^2 are unknown. The difference $\eta = \mu_1 - \mu_2$ is of interest with $\zeta = (\mu_1 + \mu_2)/2$ a nuisance parameter. We assume that in the relevant prior η, ζ, σ_1^2 and σ_2^2 are approximately independent, η is Cauchy with center y_0 and scale d_0, ζ is locally uniform, and σ_1^2 and σ_2^2 are distributed according to the chi-square distributions with 1 degree of freedom and scales s_{10} and s_{20}, respectively. Namely,

$$f(\eta, \zeta, \sigma_1^{-2}, \sigma_2^{-2}) \propto [d_o^2+(\eta-y_o)^2]^{-1}\sigma_1\sigma_2\exp[-\tfrac{1}{2}(s_{1o}^2\sigma_1^{-2} + s_{2o}^2\sigma_2^{-2})]. \quad (15)$$

The marginal posterior density of η is then the product-Cauchy,

$$f(\eta| x_1,x_2) \propto [d_o^2+(\eta-y_o)^2]^{-1} \{(s_{1o}+s_{2o})^2+[\eta-(x_1-x_2)]^2\}^{-1}. \quad (16)$$

So, by the theorem, the Bayesian estimate of η is the weighted average,

$$E(\eta| x_1,x_2) = w (x_1-x_2) + (1-w)y_o, \quad (17)$$

where $w = d_o/(d_o+s_{1o}+s_{2o})$.

Note that when d_o goes to infinity, the prior of η goes to a noninformative prior, and the Bayesian estimate converges to the classical estimate x_1-x_2.

(2) Two observations from each population

Suppose x_{11}, x_{12} and x_{21}, x_{22} are independently sampled from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Again, we are interested in the difference, $\eta = \mu_1 - \mu_2$, with $\zeta = (\mu_1 + \mu_2)/2$. We assume the prior in which η, ζ, σ_1^2 and σ_2^2 are independent, η is Cauchy distributed with center y_o and scale d_o, and the other three parameters are distributed according to the usual improper priors. In notation,

$$f(\eta, \zeta, \sigma_1, \sigma_2) \propto [d_o^2+(\eta-y_o)^2]^{-1} \sigma_1^{-1} \sigma_2^{-1}. \quad (18)$$

Then the marginal posterior density of η is

$$f(\eta| x_{11}, x_{12}, x_{21}, x_{22}) \propto \{d_o^2+(\eta-y_o)^2\}^{-1}\{[(|x_{11}-x_{12}|+|x_{21}-x_{22}|)/2]^2+[\eta-(\bar{x}_1-\bar{x}_2)]^2\}^{-1}, \quad (19)$$

and by the theorem, the Bayesian estimate is

$$E(\eta| x_{11}, x_{12}, x_{21}, x_{22}) = w (\bar{x}_1-\bar{x}_2) + (1-w) y_o, \quad (20)$$

where $w = 2d_o/(2d_o+|x_{11}-x_{12}|+|x_{21}-x_{22}|)$.

Similarly to (14), the weight on $\bar{x}_1$-$\bar{x}_2$ is a decreasing function of $|x_{11}-x_{12}|+|x_{21}-x_{22}|$. Hence, this estimate is an adaptive weighted average of the prior center y_o, and the difference of sample means, $\bar{x}_1$-$\bar{x}_2$.

Letting d_o go to infinite, we obtain a special case of Jeffreys' (1940) posterior density, again identical to Fisher's (1935) fiducial distribution in the problem.

Example 3. Cauchy population with unknown center

We consider also sampling from a Cauchy distribution and treat two cases according to the state of information on the scale parameter.

(1) Scale parameter known

Of course, the case of a single observation from a Cauchy population with known scale is identical in form to the case of one normal observation with unknown scale, which produced the marginal likelihood function of μ in the posterior density (9). It will be more interesting to treat two Cauchy observations.

Suppose we have two observations x_1 and x_2 independently drawn from the Cauchy population

$$f(x) \propto [\sigma^2+(x-\mu)^2]^{-1}, \qquad (21)$$

where μ is unknown and σ is known, and μ is locally uniformly distributed. Then μ has the posterior product-Cauchy density,

$$f(\mu | x_1, x_2) \propto [\sigma^2+(x_1-\mu)^2]^{-1}[\sigma^2+(x_2-\mu)^2]^{-1}. \qquad (22)$$

By the theorem, the posterior mean is

$$E(\mu | x_1, x_2) = (x_1+x_2)/2. \qquad (23)$$

This estimate is the sample mean and median, and is independent of the scale parameter. The posterior density (22) is a symmetric function of μ about $(x_1+x_2)/2$, and it may be bimodal when the two observations are not close together.

Letting $\frac{\partial}{\partial\mu} f(\mu | x_1, x_2) = 0$, we get the cubic equation

$$(\mu - \frac{x_1+x_2}{2})[\mu^2-(x_1+ x_2)\mu +x_1x_2 +\sigma^2] = 0, \qquad (24)$$

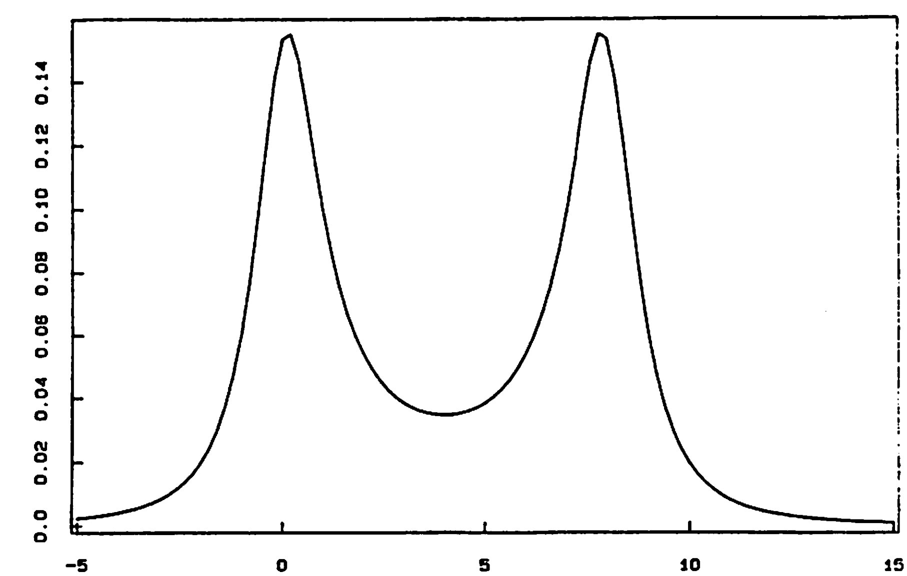
which has the three roots

$$\mu_1 = \frac{x_1+x_2}{2} - \frac{1}{2}\sqrt{(x_1-x_2)2 - 4\sigma^2},$$
$$\mu_2 = \frac{x_1+x_2}{2} + \frac{1}{2}\sqrt{(x_1-x_2)2 - 4\sigma^2}, \qquad (25)$$

and $\mu_3 = (x_1+x_2)/2.$

Figure 1. Posterior density of μ

$(x_1 = 0, x_2 = 8, \sigma = 1)$



If $|x_1 - x_2| \le 2\sigma$, (24) has only one real root μ_3, and so $f(\mu| x_1, x_2)$ is unimodal with the posterior mode equal to the posterior mean $E(\mu| x_1, x_2) = (x_1+x_2)/2$. However, if $|x_1 - x_2| \ge 2\sigma$, (24) has three real roots specified by (25), hence, $f(\mu| x_1, x_2)$ is bimodal, and has two modes, μ_1 and μ_2, both of which bear the same highest posterior density. The posterior mean $E(\mu| x_1, x_2) = (x_1+x_2)/2$ is then the intermediate local minimum of the posterior density μ_3. Figure 1 shows a bimodal curve for $f(\mu| x_1, x_2)$ with $\sigma = 1$, $x_1=0$ and $x_2 = 8$. The posterior mean is the local minimum $\mu_3 = 4$, and the posterior local modes $\mu_1 = 0.127$ and $\mu_2 = 7.873$ possess the same highest density. In this case, the posterior mean may fail to be in the highest posterior density region. Note that, since the likelihood function too is specified by (22), the maximum likelihood estimate will not be unique in this case.

(2) Scale parameter unknown

Consider again the model specified by (21) but assume σ is unknown. Suppose μ and σ are independent, μ is locally uniformly distributed and σ is distributed according to any distribution $f(\sigma)$. Similarly to the above, we obtain again

$$E(\mu| x_1, x_2, \sigma) = (x_1+x_2)/2.$$

This conditional expectation is independent of σ. Hence, the posterior mean is again

$$E(\mu| x_1, x_2) =(x_1+x_2)/2. \tag{26}$$

Example 4. Two Cauchy populations

Similarly to Example 2, we consider the difference of two population centers, but in this case, we take Cauchy populations. Suppose there are two independent observations, x_1 and x_2, one from each of the Cauchy populations with unknown centers μ_1 and μ_2 and known scales σ_1 and σ_2, respectively,

$$f(x_i| \mu_i, \sigma_i) \propto [\sigma_i^2 + (x_i - \mu_i)^2]^{-1}, \qquad i = 1, 2. \tag{27}$$

Again, we are interested in the difference, $\eta = \mu_1 - \mu_2$, with $\zeta = (\mu_1 + \mu_2)/2$. Assume the joint prior in which η and ζ are independent, η is Cauchy distributed with center y_o and scale d_o, and ζ is locally uniform,

$$f(\eta, \zeta) \propto [d_o^2+(\eta-y_o)^2]^{-1}. \tag{28}$$

Then the joint posterior density is

$$f(\eta, \zeta| x_1, x_2) \propto [d_o^2+(\eta-y_o)^2]^{-1} [\sigma_1^2+ (x_1 - \mu_1)^2]^{-1} [\sigma_2^2+ (x_2 - \mu_2)^2]^{-1}$$
$$\propto [d_o^2+(\eta-y_o)^2]^{-1} \{\sigma_1^2 + [\zeta-(x_1-\eta/2)]^2\}^{-1} \{\sigma_2^2+ [\zeta-(x_2+\eta/2)]^2\}^{-1}. \tag{29}$$

So the marginal posterior density of η is

$$f(\eta| x_1, x_2) \propto [d_o^2+(\eta-y_o)^2]^{-1} \int \{\sigma_1^2+[\zeta-(x_1-\eta/2)]^2\}^{-1}\{\sigma_2^2+[\zeta-(x_2+\eta/2)]^2\}^{-1} d\zeta$$
$$\propto [d_o^2+(\eta-y_o)^2]^{-1} \{(\sigma_1+\sigma_2)^2+[\eta-(x_1-x_2)]^2\}^{-1}, \tag{30}$$

where the second proportionality holds because the convolution of two Cauchy densities is again Cauchy. This marginal posterior density of η is in the same product-Cauchy form as in the Behrens-Fisher problem (16). By the theorem, the posterior mean is the weighted average similar to (17),

$$E(\eta \mid x_1, x_2) = w\,(x_1 - x_2) + (1-w)y_0, \qquad (31)$$

where $w = d_0/(d_0+\sigma_1+\sigma_2)$.

Conclusion

Although the results here are diminished by the special sample sizes involved, the variety of prior and sampling contexts addressed and the light shed on classical and more traditional Bayesian inferences are, perhaps, unanticipated.

References

(1) 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.

(2) Fisher, R.A. (1935), The fiducial argument in statistical inference. Ann. Eugen. 6, 391-98.

(3) Jeffreys, H. (1940), Note on the Behrens-Fisher formula. Ann. Eugen. 10, 48-51.

(4) Jeffreys, H. (1961), Theory of Probability. 3rd edition, Oxford, Clarendon Press.

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