

**Approximate studentization with  
marginal and conditional inference**

D.A.S. Fraser

Department of Mathematics and Statistics  
York University  
North York, Ontario, Canada. M3J 1P3

A.C.M. Wong <sup>1</sup>

Department of Statistics and Actuarial Science  
University of Waterloo  
Waterloo, Ontario, Canada. N2L 3G1

**Summary**

Many inference problems lead naturally to a marginal or conditional measure of departure that depends on a nuisance parameter. As a device for first order elimination of the nuisance parameter, we suggest averaging with respect to an exact or approximate confidence distribution function. It is shown that for many standard problems where an exact answer is available by other methods, the averaging method reproduces the exact answer. Moreover, for the gamma mean problem, where the exact answer is not explicitly available, the averaging method gives results that agree closely with those obtained from higher order asymptotic methods. Examples are discussed; detailed asymptotic calculations will be examined elsewhere.

**Some keywords:** Approximate inference; Averaging; Conditional inference; Confidence distribution function; Gamma mean; Marginal inference; Studentization.

---

<sup>1</sup>Now at Department of Statistics and Applied Probability, University of Alberta, Edmonton, Alberta, Canada. T6G 2G1

## 1. Introduction

Studentization can be defined generally as the process of eliminating nuisance parameters from a pivotal quantity and has its origins in the Student (1908) investigation of the mean of the  $N(0, 1)$  distribution: the measure of departure

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}}$$

is standard normal but involves the nuisance parameter  $\sigma$ ; the adjusted pivotal

$$\frac{\bar{y} - \mu}{s_y/\sqrt{n}}$$

with  $\sigma$  removed has the Student  $(n - 1)$  distribution.

Consider a sample  $(y_1, \dots, y_n)$  from the gamma distribution with mean  $\mu$  and shape  $\tau$ , where  $\mu$  is the parameter of interest and  $\tau$  is the nuisance parameter. Let  $\bar{y} = \sum y_i/n$ ,  $\tilde{y} = (\prod y_i)^{1/n}$ ,  $d_i = \log y_i - \log \tilde{y}$ , and  $e^w = \sum e^{d_i}/n$ . Then  $(\bar{y}, \mathbf{d})$  is a sufficient statistic for  $\mathbf{y}$ . Moreover, we have the factorization

$$f(\mathbf{y}; \mu, \tau) = f_1(\bar{y}|\mathbf{d}; \mu, \tau)f_2(\mathbf{d}; \tau)J(\bar{y}, \mathbf{d})$$

where  $J(\bar{y}, \mathbf{d})$  is the appropriate Jacobian. For inference concerning  $\mu$ , the conditional model  $f_1(\bar{y}|\mathbf{d}; \mu, \tau)$  is a reasonable first choice; however, it still depends on the nuisance parameter  $\tau$ . Through some non-trivial iterative calculations, Jensen (1986) constructs a conditional pivotal and through some non-trivial iterative calculation obtains tables for forming confidence interval for  $\mu$  at levels 95% and 98% for sample sizes  $n = 10, 20, 40, \infty$ . The averaging method presented in this paper will give a simple approximate solution for this gamma mean problem and a detailed discussion is given in Section 3.

In general, many problems lead easily to a factorization

$$f(\mathbf{x}; \psi, \lambda) = f_1(t|s; \psi, \lambda)f_2(s; \lambda) \tag{1}$$

of the density for a sufficient statistic  $\mathbf{x}$ , where  $\psi$  is the parameter of interest,  $\lambda$  is the nuisance parameter, and  $(s, t)$  is a equivalent statistic to  $\mathbf{x}$ . Since  $\psi$  is completely contained in the conditional density of  $t$  given  $s$ , that density will typically be used for inference concerning  $\psi$ .

However, this conditional density still depends on the nuisance parameter  $\lambda$ . The averaging method is developed as a device to eliminate  $\lambda$  from  $f_1(t|s; \psi, \lambda)$ . In this paper, the case of scalar  $\psi$  and  $\lambda$  will be considered in detail. Generalization of the proposed method beyond a two-parameter problem, though theoretically possible, would be difficult because it will involve high dimensional integrals.

Section 2 outlines the averaging method. Section 3 then applies the averaging method to the gamma mean problem where a simple exact solution is unavailable. Section 4 records examples for location-scale models and show that the averaging method gives the exact pivotal distribution as indicated in Fisher (1934); see also Fraser (1968, 1976, 1979). Moreover, the averaging method is shown to give the same result as the marginal analysis indicated for the ratio of means problems. Some concluding remarks are recorded in Section 5.

## 2. The method of averaging

Consider the factorization (1) and define  $p(\lambda)$  to be the **confidence distribution function** for the parameter  $\lambda$  at a data point  $s$  by

$$p(\lambda) = \int_{-\infty}^s f_2(u; \lambda) du \quad \lambda \in \Lambda. \quad (2)$$

For testing the hypothesis  $H_0 : \lambda = \lambda_0$ , the observed level of significance is

$$p(\lambda_0) \quad \text{or} \quad 1 - p(\lambda_0)$$

for one-sided assessment and

$$2 \min\{p(\lambda_0), 1 - p(\lambda_0)\}$$

for two-sided assessment. Similarly, a  $(1 - \alpha) \times 100\%$  confidence interval for  $\lambda$  is

$$(\min\{p^{-1}(\alpha/2), p^{-1}(1 - \alpha/2)\}, \max\{p^{-1}(\alpha/2), p^{-1}(1 - \alpha/2)\}).$$

Thus  $p(\lambda)$  gives appropriate summary about  $\lambda$  based on the data.

For inference concerning  $\psi$ , we would normally use the conditional distribution,  $f_1(t|s; \psi, \lambda)$ , but it depends on the nuisance parameter  $\lambda$ . The averaging method is to use the confidence distribution function for  $\lambda$  as defined in (2) to give a first order elimination of  $\lambda$  from

$f_1(t|s; \psi, \lambda)$ . Let

$$f_A(t|s; \psi) = \int_{\Lambda} f_1(t|s; \psi, \lambda) |dp(\lambda)| \quad (3)$$

be the resulting averaged conditional density for  $t$  given  $s$ . Then  $f_A(t|s; \psi)$  can be used for inference concerning  $\psi$ . See Fraser & Gebotys (1987) for an example in the literature.

### 3. Gamma mean problem

Consider a sample  $(y_1, \dots, y_n)$  from the gamma distribution with mean  $\mu$  and shape  $\tau$ . Jensen (1986) gives tables for forming confidence interval for  $\mu$  at levels 95% and 98% for sample sizes  $n = 10, 20, 40, \infty$ . The tables were obtained from a conditional pivotal density using an indirect saddlepoint approximation and has third order accuracy. A related procedure (Fraser, Reid & Wong, 1990) involves the direct implementation of a numerical saddlepoint procedure developed in Fraser, Reid & Wong (1991). A second order procedure which does not involve iteration is discussed in Fraser (1991).

Let  $\bar{y} = \sum y_i/n$ ,  $\tilde{y} = (\prod y_i)^{1/n}$ ,  $d_i = \log y_i - \log \tilde{y}$ , and  $e^w = \sum e^{d_i}/n$ . The density for  $\mathbf{y}$  can be factored into

$$f(\mathbf{y}; \mu, \tau) = f_1(\bar{y}|\mathbf{d}; \mu, \tau) f_2(\mathbf{d}; \tau) J(\bar{y}, \mathbf{d})$$

where  $J(\bar{y}, \mathbf{d})$  is the appropriate Jacobian, and

$$\begin{aligned} f_1(\bar{y}|\mathbf{d}; \mu, \tau) &= \frac{(n\tau)^{n\tau}}{\Gamma(n\tau)} e^{n\tau[\log \bar{y} - \log \mu - e^{\log \bar{y} - \log \mu}]} \frac{1}{\bar{y}} \\ f_2(\mathbf{d}; \tau) &= \frac{\Gamma(n\tau)}{\Gamma^n(\tau) n^{n\tau - \frac{1}{2}}} e^{-n\tau w} \end{aligned}$$

with  $d_i > 0$ ,  $w > 0$ ,  $\bar{y} > 0$ , and  $\bar{y}$  and  $\mathbf{d}$  are independent.

Since  $f_2(\mathbf{d}; \tau)$  has exponential family form with minimal sufficient statistic  $w$ , the saddlepoint approximation can be used to approximate the density for  $w$  and with the Stirling's formula, we then obtain

$$f_2(w; \tau) \approx \Gamma^{-1} \left( \frac{n-1}{2} \right) n\tau (n\tau w)^{\frac{n-1}{2}-1} e^{-n\tau w}$$

which gives the confidence distribution function for  $\tau$

$$p(\tau) \approx \int_0^w \Gamma^{-1} \left( \frac{n-1}{2} \right) n\tau (n\tau u)^{\frac{n-1}{2}-1} e^{-n\tau u} du \quad \tau > 0.$$

Then from (3), averaging  $f_1(\bar{y}|\mathbf{d}; \mu, \tau)$  or equivalently  $f_1(\bar{y}|w; \mu, \tau)$  with respect to  $p(\tau)$ , we obtain

$$f_A(\bar{y}|w; \mu) \approx \int_0^\infty f_1(\bar{y}|w; \mu, \tau) |dp(\tau)| \quad \bar{y} > 0.$$

By applying the Stirling's formula once again and using the transformation  $t = \log \bar{y} - \log \mu$ , we obtain

$$f_A(t|w) \approx \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{\sqrt{2\pi}} w^{\frac{n-1}{2}} (w + e^t - t - 1)^{-\frac{n}{2}} \quad -\infty < t < \infty.$$

Thus inference concerning  $\mu$  can be obtained directly from  $f_A(t|w)$ .

For illustration, consider the sample discussed in Jensen (1986) where  $n = 20$ ,  $\bar{y} = 113.45$ , and  $w = 0.0579$ . Table 1 records the 95% confidence interval for  $\mu$  obtained by four different methods: the maximum likelihood departure standardized by the observed information (MLE), the second order method reported in Fraser (1991), the third order method in Jensen (1986) or Fraser, Reid & Wong (1990), and the present averaging method. Note that the averaging method gives a result which is very close to that of the third order method.

**Table 1:** 95% confidence interval for  $\mu$   
( $n = 20$ ,  $\bar{y} = 113.45$  and  $w = 0.0579$ )

Method	95% CI
MLE	(96.7, 130.8)
2nd Order	(96.3, 135.4)
3rd Order	(96.8, 133.5)
Averaging	(97.0, 134.7)

As a second numerical example, consider  $n = 2$ ,  $\bar{y} = 2.5$ , and  $w = 0.2231$ . Table 2 records the left tail probability which is  $P(\bar{Y} \leq \bar{y}|w; \mu)$  for  $\mu = 1, 3, 5, 7$  and 9. Because of the small sample size  $n = 2$ , we are able to obtain the exact left tail probability by direct numerical integration. The averaging method provides a pretty accurate approximation. Even though it is not as accurate as the third order method, it is much easier and does give very good approximations.

**Table 2:** Left tail probability for  $mu = 1, 3, 5, 7, 9$   
 $(n = 2, \bar{y} = 2.5 \text{ and } w = 0.2231)$

$\mu$	1	3	5	7	9
MLE	0.905	0.331	0.014	$4.09 \times 10^{-5}$	$6.37 \times 10^{-9}$
2nd Order	0.916	0.330	0.223	0.181	0.160
3rd Order	0.918	0.430	0.293	0.242	0.215
Averaging	0.892	0.512	0.330	0.280	0.257
Exact	0.901	0.484	0.318	0.256	0.225

#### 4. Location-scale problems and ratio of means problems

First consider a sample  $(y_1, \dots, y_n)$  from the  $N(\mu, \sigma^2)$ . We have the factorization

$$f(\mathbf{y}; \mu, \sigma^2) = f_1(\bar{y}|s_y^2; \mu, \sigma^2) f_2(s_y^2; \sigma^2) J(\bar{y}, s_y^2)$$

where  $J(\bar{y}, s_y^2)$  is the appropriate Jacobian,  $\mu$  is the parameter of interest and  $\sigma^2$  is the nuisance parameter. Notice that  $\bar{y} \sim N(\mu, \sigma^2/n)$ ,  $(n-1)s_y^2 \sim \sigma^2 \chi_{(n-1)}^2$ , and  $\bar{y}$  and  $s_y^2$  are independent. The confidence distribution function for  $\sigma^2$  with  $s_y^2$  at its observed value is

$$p(\sigma^2) = \int_0^{s_y^2} \Gamma^{-1} \left( \frac{n-1}{2} \right) e^{-\frac{(n-1)s^2}{2\sigma^2}} \left[ \frac{(n-1)s^2}{2\sigma^2} \right]^{\frac{n-1}{2}-1} \frac{n-1}{2\sigma^2} ds^2.$$

Then

$$|dp(\sigma^2)| = \Gamma^{-1} \left( \frac{n-1}{2} \right) e^{-\frac{(n-1)s_y^2}{2\sigma^2}} \left[ \frac{(n-1)s_y^2}{2\sigma^2} \right]^{\frac{n-1}{2}-1} \frac{(n-1)s_y^2}{2\sigma^4} d\sigma^2$$

and from (3)

$$f_A(\bar{y}|s_y^2; \mu) = \frac{\sqrt{n}}{\sqrt{\pi}} \Gamma^{-1} \left( \frac{n-1}{2} \right) \left[ \frac{1}{(n-1)s_y^2} \right]^{\frac{1}{2}} \Gamma \left( \frac{n}{2} \right) \left[ 1 + \frac{n(\bar{y} - \mu)^2}{(n-1)s_y^2} \right]^{-\frac{n}{2}}.$$

The transformation  $t = \sqrt{n}(\bar{y} - \mu)/s_y$ , with  $s_y$  fixed, then gives

$$f_A(t) = \frac{\Gamma \left( \frac{n}{2} \right)}{\Gamma \left( \frac{1}{2} \right) \Gamma \left( \frac{n-1}{2} \right) \sqrt{n-1}} \left[ 1 + \frac{t^2}{n-1} \right]^{-\frac{n}{2}} \quad -\infty < t < \infty$$

which is the appropriate Student  $(n-1)$  density for the problem.

A similar result is obtained for the location-scale model  $y_i = \mu + \sigma u_i$  where  $(u_1, \dots, u_n)$  is a sample from the Uniform  $(0, 1)$  distribution. Let  $(b_y, s_y)$  be the transformation variable as defined in Fraser (1968, P. 25). Without loss of generality, for this problem, let  $(b_y, s_y) = (y_{(1)}, y_{(n)} - y_{(1)})$ . Moreover, let  $d_i = (y_i - b_y)/s_y$ . Then the conditional distribution of  $(b_y, s_y)$  given  $\mathbf{d}$  is

$$f(b_y, s_y | \mathbf{d}; \mu, \sigma) = n(n-1) \frac{s_y^{n-2}}{\sigma^n}$$

with  $0 < s_y < \sigma$ , and  $\mu < b_y < \mu + \sigma - s_y$ ; see Fraser (1968, 1976, 1979). Note that this conditional density is independent of  $\mathbf{d}$ , and can be factored into

$$f(b_y, s_y; \mu, \sigma) = f_1(b_y | s_y; \mu, \sigma) f_2(s_y; \sigma)$$

where

$$f_1(b_y | s_y; \mu, \sigma) = \frac{1}{\sigma - s_y} \quad s_y < b_y - \mu + s_y < \sigma,$$

$$f_2(s_y; \sigma) = \frac{n(n-1)s_y^{n-2}}{\sigma^n} (\sigma - s_y) \quad 0 < s_y < \sigma.$$

Then

$$|dp(\sigma)| = \frac{n(n-1)s_y^{n-2}}{\sigma^{n-1}} (s_y - \sigma)$$

and from (3), the averaging method gives the averaged density for  $b_y$  with  $s_y$  fixed:

$$f_A(b_y | s_y; \mu) = \frac{n-1}{\left(1 + \frac{b_y - \mu}{s_y}\right)^n} \frac{1}{s_y} \quad 0 < b_y - \mu < \infty.$$

Let  $t = (b_y - \mu)/s_y$  with  $s_y$  fixed, we have

$$f_A(t) = \frac{n-1}{(1+t)^n} \quad t > 0.$$

This agrees with the direct Student-type analysis; see Fisher (1934) and also Fraser (1968, 1976, 1979).

In fact, the averaging method works in general for the location-scale model

$$y_i = \mu + \sigma e_i$$

where  $(e_1, \dots, e_n)$  is a sample from a known distribution with density  $f(\cdot)$ . Again, let  $(b_y, s_y)$  be a transformation variable and let  $d_i = (y_i - b_y)/s_y$ . Then by change of variables, the joint

conditional density for  $(b_y, s_y)$  given  $\mathbf{d}$  is

$$\begin{aligned} f(b_y, s_y | \mathbf{d}; \mu, \sigma) &= k(\mathbf{d}) \prod f\left(\frac{b_y - \mu + s_y d_i}{\sigma}\right) \frac{s_y^{n-2}}{\sigma^n} \\ &= f_1(b_y | s_y, \mathbf{d}; \mu, \sigma) f_2(s_y | \mathbf{d}; \sigma) \end{aligned}$$

with  $-\infty < b_y < \infty$  and  $s_y > 0$ , where

$$\begin{aligned} f_1(b_y | s_y, \mathbf{d}; \mu, \sigma) &= c(s_y, \mathbf{d}) \prod f\left(\frac{b_y - \mu + s_y d_i}{\sigma}\right), \\ f_2(s_y | \mathbf{d}; \sigma) &= \int_{-\infty}^{\infty} f(b_y, s_y | \mathbf{d}; \mu, \sigma) db_y, \end{aligned}$$

and  $c(s_y, \mathbf{d})$  is the normalizing constant. By applying (3) and let  $t = (b_y - \mu)/s_y$ , the averaging method gives the averaged conditional density for  $t$  given  $\mathbf{d}$  with  $s_y$  fixed:

$$f_A(t | \mathbf{d}) = \int_0^{\infty} k(\mathbf{d}) \frac{s_y^n}{\sigma^{n+1}} \prod f\left(\frac{s_y}{\sigma}(t + d_i)\right) d\sigma.$$

This agrees with the general Student-type analysis derived from Fisher's (1934) conditional approach; see also Fraser (1968, 1976, 1979).

Consider the two sample problem where  $(y_{11}, \dots, y_{1n})$  and  $(y_{21}, \dots, y_{2n})$  are samples from independent exponential distributions with means  $\lambda^{-1}$  and  $(\psi\lambda)^{-1}$  respectively. For inference concerning  $\psi$ , the indicated exact analysis is to use the marginal distribution of  $y_1/(\psi y_2)$  which has an  $F(2n, 2n)$  distribution where  $y_1 = \sum y_{1i}$  and  $y_2 = \sum y_{2i}$ .

The joint density of  $(y_1, y_2)$  has the factorization

$$\begin{aligned} f(y_1, y_2; \psi, \lambda) &= f_1(y_2 | y_1; \psi, \lambda) f_2(y_1; \lambda) \\ &= \left\{ \frac{1}{\Gamma(n)} (\psi\lambda)^n e^{-\psi\lambda y_2} y_2^{n-1} \right\} \times \left\{ \frac{1}{\Gamma(n)} \lambda^n e^{-\lambda y_1} y_1^{n-1} \right\}. \end{aligned}$$

The confidence distribution function for  $\lambda$  obtained from  $f_2(y_1; \lambda)$  is

$$p(\lambda) = \int_0^{y_1} \frac{1}{\Gamma(n)} \lambda^n e^{-\lambda u} u^{n-1} du \quad \lambda > 0.$$

Then from (3), averaging  $f_2(y_2 | y_1; \psi, \lambda)$  with respect to  $p(\lambda)$  gives

$$f_A(y_2 | y_1; \psi) = \frac{\Gamma(2n)}{\Gamma(n)\Gamma(n)} \frac{\psi^n y_1^n y_2^{n-1}}{(y_1 + \psi y_2)^{2n}}.$$

With  $t = y_1/(\psi y_2)$  and  $y_1$  fixed, we have

$$f_A(t) = \frac{\Gamma(2n)}{\Gamma(n)\Gamma(n)} \frac{t^{n-1}}{(1+t)^{2n}} \quad t > 0$$

which is the density of  $F(2n, 2n)$  distribution as indicated by the exact marginal analysis.

The averaging method can also be extended to two independent samples  $(y_{11}, \dots, y_{1m})$  and  $(y_{21}, \dots, y_{2n})$ , from gamma distributions with shapes  $\alpha_1, \alpha_2$  and scales  $\lambda\psi, \lambda$  respectively. Assume  $\alpha_1$  and  $\alpha_2$  are known. Again, the averaging method gives the standard result when a direct method is available.

## 6. Discussion

A confidence distribution function for a parameter records in effect a full spectrum of confidence intervals for a parameter. In this paper, we have used such a confidence distribution function for a nuisance parameter as a tool to eliminate or average out the nuisance parameter in a pivotal quantity for a primary parameter of interest. For many standard problems where an exact answer is available by other means, the technique reproduces the exact answer; see Sections 4.

For the gamma mean problem (Section 3), the technique leads to a simple inference distribution that agrees closely with results obtained by third order asymptotics with non-trivial iterative and numerical calculations.

More generally, we suggest that a first order asymptotic confidence distribution function for a nuisance parameter, say from the standardized maximum likelihood, standardized score, or likelihood ratio, be used to average the nuisance parameter from the distribution of a pivotal quantity that focuses on a primary parameter of interest.

A Bayesian approach could use a prior distribution for  $\lambda$  and combine it say with  $f_2(s; \lambda)$  in (1) to obtain a distribution to eliminate  $\lambda$  from  $f_1(t|s; \psi, \lambda)$ ; see Cox (1975). This seems not to be a pure Bayesian procedure but could provide reasonable approximate methods. Our preference however is for the frequentist approach combined with the confidence distribution functions.

## References

- Cox, D.R. (1975). A note on partially Bayes inference and linear models. *Biometrika* **62**, 651-654.

- Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. Royal Soc. A* **144**, 285-307.
- Fraser, D.A.S. (1968). *The Structure of Inference*. New York: Wiley.
- Fraser, D.A.S. (1976). *Probability and Statistics: Theory and Applications*. Toronto: DAI Press.
- Fraser, D.A.S. (1979). *Inference and Linear Models*. New York: McGraw-Hill.
- Fraser, D.A.S. (1991). Statistical inference from likelihood to significance. *J. Amer. Statist. Assoc.* **86**, 258-265.
- Fraser, D.A.S., and Gebotys, R.J. (1987). Nonnested linear models: a conditional confidence approach. *Can. J. Statist.* **15**, 375-385.
- Fraser, D.A.S., Reid, N. and Wong, A. (1990). Inference for the gamma mean. Submitted to *Technometrics*.
- Fraser, D.A.S., Reid, N. & Wong, A. (1991). Exponential linear model: a two pass procedure for saddlepoint approximation. *J. R. Statist. Soc. B* **53**, 483-492.
- Jensen, J.L. (1986). Inference for the mean of a gamma distribution with unknown shape parameter. *Scand. J. Statist.* **13**, 135-151.
- Student (1908). The probable error of a mean. *Biometrika* **6**, 1-25.