^{-1}B, and precision $[(1,-1)P^{-1}(\theta|s)(l,-1)']^{-1} = P(\gamma|s)$. These results can be verified by referring to DeGroot, Chapter 5, or to the Appendix of this book.

Remember the hyperparameters μ, p, α and β must be known in order to do the posterior analysis and they can be determined by using the methods of the previous section of this chapter. On the other hand, if very little is known about the parameters, before the experiment, one perhaps would use the Jeffreys' prior (1.13) which results in a simplified posterior analysis. For example the posterior distribution of θ is a bivariate t with $n-2$ degrees of freedom, location vector $E(\theta|s) = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}$ and precision matrix

$$P(\theta|s) = \frac{(n_1+n_2-2)\begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}}{(n_1-1)s_1^2+(n_2-1)s_2^2},$$

where

$$s_i^2 = \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2(n_i-1)^{-1}, \quad i=1,2$$

and $\bar{x}_1$ and $\bar{x}_2$ are the sample means and s_1^2 and s_2^2 the sample variances.

The posterior analysis for θ using Jeffreys' prior density is based on Theorem 1.1 of Chapter 1. One may obtain the same results by letting $\alpha \to -1$, $\beta \to 0$, $p \to 0(2\times 2)$ in the joint posterior distribution of θ with density (3.12).

Regions of highest posterior density, HPD, may be found for θ, θ_i, i = 1, 2, and γ. First for θ, from (3.12) one may show (see Chapter 1, 1.28)

$$F(\theta) = \frac{2^{-1}(\theta-A^{-1}B)'A(\theta-A^{-1}B)(n+2\alpha)}{C-B'A^{-1}B} \tag{3.15}$$

has an F-distribution with 2 and $n + 2\alpha$ degrees of freedom and one may plot 99%, 95% regions for θ. If θ_1 or θ_2 or γ is the parameter of interest, one may construct an HPD interval using the Student t tables. How would one do this?

Since and θ_1 θ_2 are the primary parameters of interest, Bayesian inferences for τ will not be discussed.

Two Normal Populations with a Common Mean

Consider two normal populations $n(\theta, \tau^{-1}{}_1)$ and $n(\theta, \tau^{-1}{}_2)$ with a common mean θ but distinct precisions τ_1 and τ_2, then on the basis of random samples $s_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})$, i = 1, 2, how does one make inferences about the common mean? Suppose given θ, τ_1 and τ_2 that S_1 and S_2 are independent, then the likelihood function for θ and $\tau = (\tau_1, \tau_2)$ is

$$L(\theta,\tau|s) \propto \prod_{i=1}^{2} \tau_i^{n_i/2}\exp - \frac{\tau_i}{2}\sum_{j=1}^{n_i}(x_{ji}-\theta)^2, \quad \theta \in R, \quad \tau_i > 0, \tag{3.16}$$

where s = (S_1, S_2). The conjugate prior density is

$$\xi(\theta,\tau) \propto \prod_{i=1}^{2} \tau_i^{\alpha_i-1}e^{-\tau_i\beta_i}\tau_i^{1/2}e^{-\tau_i P_i(\theta-\mu_i)^2/2},$$

$$\theta \in R, \quad \tau_i > 0, \tag{3.17}$$

where $\alpha_i > 0$, $\beta_i > 0$, $P_i > 0$, and $\mu_i \in R$, and is the product of two normal-gamma densities, one for each population. This implies the marginal prior density of θ is

$$\xi(\theta,\tau) \propto \prod_{i=1}^{2}[2\beta_i + P_i(\theta-\mu_i)^2]^{-(2\alpha_i+1)/2}, \quad \theta \in R,$$

which is the product of two t densities on θ and is called a 2/0 univariate t density with parameters α_i, β_i, P_i, and μ_i, i = 1, 2. This is a very complicated density because its normalizing constant and moments are unknown, thus it is very difficult to assign values to the hyperparameters. How would one assign values to the hyperparameters? One solution is to consider each population separately, so for the $n(\theta, \tau^{-1}{}_1)$ population, one may set the values of μ_1, P_1, α_1, and β_1 and then repeat the process for the second population. Is this a legitimate procedure?

The marginal prior density of τ_1 and τ_2 is

$$\xi(\tau) \propto \prod_{i=1}^{2} \tau_i^{\alpha_i-1}\, e^{-\tau_i\beta_i}\frac{\tau_1^{1/2}\tau_2^{1/2}}{(\tau_1P_1+\tau_2P_2)^{1/2}} \tag{3.18}$$
$$\times\; e^{-\tau_1\tau_2P_1P_2(\mu_1-\mu_2)^2/(\tau_1P_1+\tau_2P_2)}, \qquad \tau_1 > 0, \qquad \tau_2 > 0,$$

thus τ_1 and τ_2 are not independent, a priori, even if $\mu_1 = \mu_2$. Since the populations have a common mean, it is probably wise to let μ_1 and μ_2 have a common value, in which case (3.18) simplifies but is not the density of independent precision components.

Proceeding with the posterior analysis, we have

$$\xi(\theta,\tau|s) \propto \prod_{i=1}^{2} \tau_i^{(2\alpha_i+n_i+1)/2-1} \exp - \tfrac{\tau_i}{2}\{2\beta_i + \sum_{j=1}^{n_i}(x_{ij} - \theta)^2 + P_i(\theta-\mu_i)^2\}, \tag{3.19}$$

where $\theta \in R$, $\tau_i > 0$, for the joint posterior density of the parameters.

Completing the square on θ and integrating with respect to and τ_2 gives

$$\xi(\theta,\tau|s) \propto \prod_{i=1}^{2}\{2\beta_i + \sum_{j=1}^{n_i}(x_{ij} - \overline{x_i})^2 + n_iP_i(\overline{x}i - \mu_i)^2/(n_i + P_i) + [\theta - (n_i + P_i)^{-1}(n_i\overline{x_i} + P_i\mu_i)]^2(n_i + P_i)\}^{-(n_i+2\alpha_i+1)/2}, \quad \theta \in R \tag{3.20}$$

for the marginal posterior density of θ and is a 2/0 scalar poly-t density, which, as we have seen, is the same form as the prior density of θ. For the properties of this distribution, see Box and Tiao (1973), Zellner (1971), Dreze (1977), Sedory (1980), and Yusoff (1982). Since θ is scalar, the density (3.20) is easily plotted and its moments calculated by numerical integration; however if θ is a vector, it may be necessary to use approximations to the density and this has been done by Box and Tiao (1973), Zellner (1971), Dreze (1977), and Yusoff (1982).

What is the marginal posterior density of τ_1 and τ_2 If θ is eliminated from the joint density,

$$\xi(\tau|s) \propto \prod_{i=1}^{2} \tau_i^{(n_i+2\alpha_i)/2-1} \exp - \tfrac{\tau_i}{2}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_i)^2 \frac{\exp -A/2}{[\Sigma\tau_i(n_i+P_i)]^{1/2}}, \quad \tau_i > 0 \tag{3.21}$$

is the density of the precision components, where

$$\sum_1^2 \tau_i(n_i+P_i)A = \sum_1^2 \tau_i^2 n_iP_i(\overline{x}_i-\mu_i)^2 + \tau_1\tau_2\Big\{n_1n_2(\overline{x}_1-\overline{x}_2)^2 + n_1P_2(\overline{x}_1-\mu_2)^2 + n_2P_1(\overline{x}_2-\mu_1)^2 + P_1P_2(\mu_1-\mu_2)^2\Big\},$$

which is some type of bivariate-gamma distribution and it is necessary to employ numerical integration to calculate the joint and marginal properties of the distribution.

Thus with a common mean for the two populations, it is very difficult to do the posterior analysis because the marginal posterior distributions for θ, then for τ_1 and τ_2 are not standard densities.

For example, if the precisions are equal and the means are distinct, it has been shown the posterior analysis results in standard well-known distributions for the parameters.

How should inferences be made for this problem? On the one hand, for θ, one must use the poly-t density, and on the other, to make inferences for τ_1 and τ_2, one must use the bivariate type gamma density (3.21), and both densities must be numerically determined.

Is there some other way by which posterior inferences for θ and τ are possible?

Consider the conditional posterior densities for the parameters, for example, for θ given τ_1 and τ_2, and then for τ given θ. The former distribution is normal with mean

$$E(\theta|\tau,s) = [\sum_1^2 \tau_i(n_i+P_i)]^{-1}[\sum_1^2 \tau_i(n_i\overline{x}_i + P_i\mu_i)]$$

and precision

$$P(\theta|\tau,s) = \sum_1^2 \tau_i(n_i+P_i)$$

and the conditional distribution of τ_1 and τ_2 given θ, is that of independent gamma random variables, such that $\tau_i|\theta$ is gamma with parameters $(2\alpha_i + n_i + 1)/2$ and

$$\frac{2\beta_i + \sum_{j=1}^{n_i}(x_{ij}-\overline{x}_i) + n_i(\theta-\overline{x}_i)^2 + P_i(\theta-\mu_i)^2}{2}.$$

Although the marginal posterior distributions of the parameters are intractable, the corresponding conditional distributions are well-known. How can these be used for inferences about the parameters?

Perhaps the mode of all the parameters can be found from the modes of the conditional densities, which are

Perhaps the mode of all the parameters can be found from the modes of the conditional densities, which are

$$M(\theta|\tau,s) = \frac{\sum_{1}^{2} \tau_i (n_i \bar{x}_i + P_i \mu_i)}{\sum_{1}^{2} \tau_i (n_i + P_i)} \tag{3.22}$$

and

$$M(\tau_i|\tau_j, i \neq j, \theta) = \frac{n_i + 2\alpha_i - 1}{2\beta_i + \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + n_i(\theta - \bar{x}_i)^2 + P_i(\theta - \mu_i)^2} \tag{3.23}$$

for i = 1, 2.

Using an approach of Lindley and Smith (1972), the mode of the joint density of θ, τ_1, and τ_2 perhaps, can be found by solving simultaneously the above three modal equations for the parameters.

An explicit solution cannot be found and an iterative technique can be tried by starting the process with some trial values for τ_1 and τ_2 (say the sample precisions) and finding the conditional mode of θ from (3.22) which is then substituted into the equations, (3.23), for the conditional modes of τ_1 and τ_2. The process is repeated until the estimates stabilize and then the final estimates are a solution to the modal equations, and the solution, perhaps, is the joint mode of the joint posterior density.

Since the joint density may have multiple modes, the iterative technique may not converge to the absolute maximum but instead to a local maximum, depending on the starting value of the iterative procedure.

Of course, the joint mode of the parameters, if it can be found, is a useful characteristic of the posterior distribution since it gives one a point estimate of the parameters, but it does not tell us anything about the precision of the posterior distribution.

Suppose θ is the primary parameter of interest and τ_1 and τ_2 are nuisance parameters, then how should one estimate θ? There are several possibilities. Suppose we obtain the marginal posterior mean and mode of θ from the 2/0 poly-t density (3.20) via numerical integration and the θ component of the joint posterior mode, which is found by an iterative solution to the modal equations (3.22) and (3.23) where the sample precisions are used as starting values. This will give three point estimates of θ.

Suppose the sample values are $n_1 = n_2 = 10$ and $\sum_1^{10} (x_{1j} - \bar{x}_1)^2 = 31.7882$, $\sum_1^{10} (x_{2j} - \bar{x}_2)^2 = 9.9999$, $\bar{x}_1 = -.1499$ and $\bar{x}_2 = .3611$, thus there are two normal populations with a common mean and distinct variances and the two sample means are −.1499 and.3611. The sample precisions are. 28312 and .9 respectively and are used as the starting values of the iterative solution to the modal equations.

One must choose the hyperparameters α_1, β_1, α_2 β_2 P_1 P_2, μ_1, and μ_2 of the prior distribution, and for fifty-two combinations of these parameters, Table 3.1 lists the mode of the marginal density of θ, the θ component of the solution to the modal equations, and the mean of the marginal density of θ.

We see, first of all, that for each combination of the eight hyperparameters, the three estimates of θ are very close. The marginal mean and mode of θ are not the same because the marginal density of θ, (3.20), is usually not symmetric and the θ component of the joint modal estimate is not the same as that calculated from the marginal density of θ, nevertheless, the three estimates are close to one another.

Table 3.1 reveals the sensitivity of the three posterior estimates to the values of the parameters of the prior distribution. For example, the range of the largest to the smallest values of each estimate is 5.4087, 5.3870, and 5.3865 for the marginal mode, θ component of iterative solution, and posterior marginal mean, respectively, and the last has the smallest range.

The three estimates appear to be sensitive to μ_1 and μ_2, the two means of the prior distribution of θ. As μ_1 and μ_2 increase, so do the posterior mean of θ and the other two estimates and this was not unexpected.

In order to arrive at definite conclusions about this way of estimation, more numerical studies need to be conducted.

Another way to do the posterior analysis is with an improper prior density

$$\xi(\theta, \tau_1, \tau_2) \propto \tau_1^{-1} \tau_2^{-1}, \quad \theta \in R, \quad \tau_1 > 0, \quad \tau_2 > 0 \tag{3.24}$$

for the parameter. If this is done, one will see that the marginal posterior density of θ is a 2/0 poly-t, of the form (3.20), and the marginal posterior density of τ is a bivariate type gamma density, (3.21).

Also, the analysis of this section is easily extended to $k(\geqslant 2)$ normal populations with the same mean but with distinct variances.

Table 3.1
Comparison of the Mode (from Marginal Distribution), Mode (by Iteration), and the Mean

α_1	α_2	β_1	β_2	ξ_1	ξ_2	μ_1	μ_2	Mode (marginal)	Mode (iterative)	Mean
0	0	0	0	0	0	0	0	.2458	.2418	.2261
0	0	0	0	0	0	2	2	.2458	.2418	.2261
0	0	0	0	0	0	4	4	.2458	.2418	.2261
0	0	0	0	0	0	6	6	.2458	.2418	.2261
0	0	0	0	0	0	8	8	.2458	.2418	.2261

0	0	0	0	0	0	8	8	.2458	.2418	.2261
0	0	0	0	0	0	10	10	.2458	.2418	.2261
0	0	0	0	2	2	0	0	.2060	.1995	.1867
0	0	0	0	4	4	0	0	.1661	.1698	.1589
0	0	0	0	6	6	0	0	.1528	.1478	.1384
0	0	0	0	8	8	0	0	.1262	.1308	.1225
0	0	0	0	10	10	0	0	.1130	.1173	.1099
0	0	0	0	12	12	0	0	.1130	.1064	.0997
0	0	0	0	14	14	0	0	.0997	.0973	.0911
0	0	0	0	16	16	0	0	.0864	.0896	.0840
0	0	0	0	18	18	0	0	.0864	.0831	.0779
0	0	1	3	0	0	0	0	.1894	.1989	.1867
0	0	2	6	0	0	0	0	.1528	.1679	.1592
0	0	3	9	0	0	0	0	.1528	.1443	.1389
0	0	4	12	0	0	0	0	.1162	.1260	.1231
0	0	5	15	0	0	0	0	.1162	.1113	.1105
0	0	6	18	0	0	0	0	.1162	.0992	.1001
0	0	7	21	0	0	0	0	.0797	.0891	.0915
0	0	8	24	0	0	0	0	.0797	.0806	.0841
0	0	9	27	0	0	0	0	.0797	.0733	.0778
2	2	0	0	0	0	0	0	.2458	.2418	.2305
4	4	0	0	0	0	0	0	.2458	.2418	.2330
6	6	0	0	0	0	0	0	.2458	.2418	.2346
8	8	0	0	0	0	0	0	.2458	.2418	.2357
10	10	0	0	0	0	0	0	.2458	.2418	.2365
12	12	0	0	0	0	0	0	.2458	.2418	.2371
14	14	0	0	0	0	0	0	.2458	.2418	.2376
16	16	0	0	0	0	0	0	.2458	.2418	.2380
2	2	1	1	3	3	−8	−8	−1.7741	−1.7634	−1.7636
2	2	1	1	3	3	−6	−6	−1.2957	−1.2967	−1.2974
2	2	1	1	3	3	−4	−4	−.8173	−.8225	−.8245
2	2	1	1	3	3	−2	−2	−.3389	−.3299	−.3349
2	2	1	1	3	3	0	0	.1794	.1745	.1664
2	2	1	1	3	3	2	2	.6179	.6262	.6187

2	2	1	1	3	3	4	4	1.0565	1.0572	1.0520
2	2	1	1	3	3	6	6	1.4950	1.4999	1.4966
2	2	1	1	3	3	8	8	1.9336	1.9512	1.9489
2	2	1	1	0	0	1	1	.2359	.2299	.2191
2	2	1	1	3	3	1	1	.4186	.4078	.3996
2	2	1	1	6	6	1	1	.5382	.5189	.5124
2	2	1	1	9	9	1	1	.5781	.5949	.5895
2	2	1	1	12	12	1	1	.6578	.6502	.6455
2	2	1	1	15	15	1	1	.6977	.6922	.6881
2	2	1	1	18	18	1	1	.7375	.7252	.7215
2	2	1	3	2	2	2	2	.4983	.4947	.4877
4	4	2	6	4	4	4	4	1.2425	1.2501	1.2477
6	6	3	9	6	6	6	6	2.3322	2.3318	2.3301
8	8	4	12	8	8	8	8	3.6346	3.6236	3.6229

Two Normal Populations with Distinct Means and Precisions

The Behrens-Fisher problem is the problem of comparing the means θ_1 and θ_2 of two normal populations which have distinct variances or precisions τ_1 and τ_2. If the precisions are equal, the two-sample t-test is the way to test the hypothesis the two means are equal, and the Bayesian analysis of this problem was discussed in the section on two normal populations with distinct means (see pp. 70