

Chapter 3

t-tests and normal variance tests

We deal first with the Bayesian analogues of the one- and two-sample *t*-tests, the bread-and-butter of simple normal-based frequentist analysis. These sections of this chapter are from unpublished papers by Aitkin (2006) and Aitkin and Liu (2007). A remarkable feature of Bayesian analysis has been the absence, until very recently, of formal Bayes procedures paralleling the two-sample *t*-test – reliance has been placed instead on the credible interval for the mean difference. Bayes factors have not been used, for reasons we have already seen. We comment below on the most recent Bayesian alternatives to the two-sample *t*-test.

3.1 One sample *t*-test

3.1.1 Credible interval

In the one-sample case, we observe data $y_i, i = 1, \dots, n$ from the model $Y \sim N(\mu, \sigma^2)$. The likelihood function is

$$\begin{aligned} L(\mu, \sigma) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2} \frac{(y_i - \mu)^2}{\sigma^2}\right\} \\ &= c \cdot \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} [RSS + n(\bar{y} - \mu)^2]\right\} \\ &= c \cdot \frac{\sqrt{n}}{\sigma} \exp\left\{-\frac{1}{2} \frac{n(\bar{y} - \mu)^2}{\sigma^2}\right\} \cdot \frac{1}{\sigma^{n-1}} \exp\left\{-\frac{1}{2} \frac{RSS}{\sigma^2}\right\} \end{aligned}$$

where $RSS = \sum_i (y_i - \bar{y})^2$. A conjugate prior for μ, σ based on a (hypothetical) *prior sample of size m*, has the same form as the likelihood:

$$\pi(\mu, \sigma) = c' \cdot \frac{\sqrt{m}}{\sigma} \exp\left\{-\frac{1}{2} \frac{m(\mu_P - \mu)^2}{\sigma^2}\right\} \cdot \frac{1}{\sigma^{m-1}} \exp\left\{-\frac{1}{2} \frac{PSS}{\sigma^2}\right\},$$

where μ_P is the prior mean of μ and PSS is the “prior sum of squares” defining the prior precision of σ . Thus μ has a normal prior distribution $N(\mu_P, \sigma^2/m)$ conditional on σ , and PSS/σ^2 has a marginal χ_{m-1}^2 prior distribution. The posterior distribution of μ given σ is $N((n\bar{y} + m\mu_P)/(n+m), \sigma^2/(n+m))$, while the marginal posterior distribution of $(RSS + PSS)/\sigma^2$ is $\chi_{n-1+m-1}^2$. Letting $m \rightarrow 0$ (with $PSS = 0$), the diffuse limit of the conjugate prior is the improper prior $1/\sigma$, and the resulting posterior distribution of μ and σ^2 can be expressed as

$$\mu \mid \sigma, \mathbf{y} \sim N(\bar{y}, \sigma^2/n), \quad RSS/\sigma^2 \mid \mathbf{y} \sim \chi_{n-1}^2.$$

Integrating over σ , standard calculations show that the marginal posterior distribution of $t = \sqrt{n}(\mu - \bar{y})/s$ is t_{n-1} , where $s^2 = RSS/(n-1)$. The $100(1-\alpha)\%$ credible intervals for μ correspond exactly, for the diffuse prior, to the usual $100(1-\alpha)\%$ frequentist t confidence intervals. Here the posterior distribution is analytic, but we note for later use that the distribution of μ can be simulated in two stages:

- make M random draws of RSS/σ^2 from its marginal χ_{n-1}^2 distribution, and hence M random draws $\sigma^{[m]}$ of σ ;
- for each $\sigma^{[m]}$, make a random draw $\mu^{[m]}$ of μ from its conditional normal $N(\bar{y}, \sigma^{[m]2}/n)$ distribution.

The M draws $\mu^{[m]}$ ($m = 1, \dots, M$) provide a random sample from the marginal posterior distribution of μ . An approximate 95% credible interval for μ is defined by the 2.5% and 97.5% points of the empirical distribution of the $\mu^{[m]}$. This approximation can be made as accurate as required by setting M sufficiently large.

3.1.2 Model comparisons

To test the null hypothesis $H_1 : \mu = \mu_1$ against the alternative $H_2 : \mu$ unspecified, we compute the likelihood ratio and deviance difference:

$$\begin{aligned} LR_{12} &= \frac{c \cdot \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}[RSS + n(\bar{y} - \mu_1)^2]\right\}}{c \cdot \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}[RSS + n(\bar{y} - \mu)^2]\right\}} \\ &= \exp\left\{-\frac{1}{2\sigma^2}[n(\bar{y} - \mu_1)^2 - n(\bar{y} - \mu)^2]\right\} \\ DD_{12} &= \frac{1}{\sigma^2}[n(\bar{y} - \mu_1)^2 - n(\bar{y} - \mu)^2] \\ &= \frac{n(\bar{y} - \mu_1)^2}{s^2} \cdot \frac{s^2}{\sigma^2} - \frac{n(\bar{y} - \mu)^2}{\sigma^2} \\ &= t^2 \cdot W/(n-1) - Z^2 \end{aligned}$$

where t is the frequentist one-sample t -statistic, $W = RSS/\sigma^2$ has a χ_{n-1}^2 distribution, and Z has an independent standard normal distribution.

Thus the posterior distribution of DD is that of the difference between a $\chi_{n-1}^2/(n-1)$ variate scaled by t^2 , and an independent χ_1^2 variate. This distribution has no closed form density, but can be directly simulated by making M independent draws from the two χ^2 variates and forming the M draws from the scaled difference DD .

3.1.3 Example

We modify slightly an example from Gönen, Johnson, Lu and Westfall (2005). In a sample of $n = 10$, we observe a sample mean $\bar{y} = 5$ with a sample standard deviation $s = 8.74$. The null hypothesis is $\mu = 0$, the alternative is $\mu \neq 0$. The t -statistic is 1.809, with a p -value of 0.104, well above the conventional criterion for rejection. The 95% credible interval for μ is $\bar{y} \pm 2.262s/\sqrt{n}$, here $(-1.252, 11.252)$. This is identical to the 95% confidence interval for μ (with the diffuse prior used).

We make 10,000 draws from the posterior distribution of DD as described in Section 2. The cdf of these values is shown in Figure 3.1, and a smooth approximation to the density in Figure 3.2.

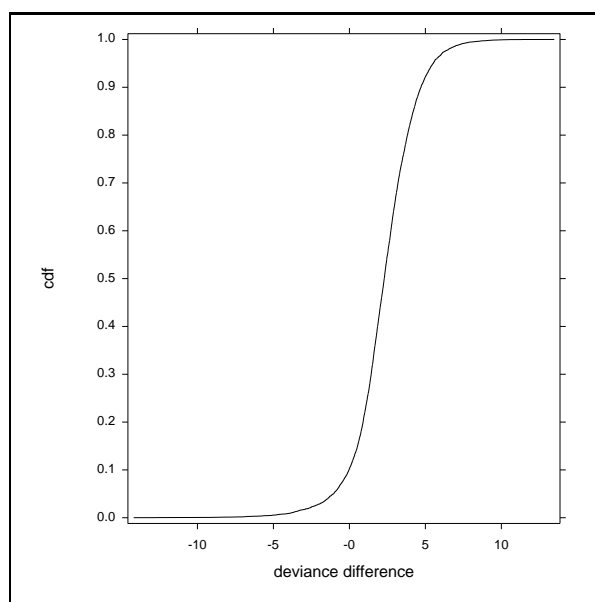


Figure 3.1: Posterior cdf of DD

The empirical probability that $DD < 0$ (that is, that the null hypothesis has higher likelihood than the alternative) is .103, with simulation standard error 0.003. This is very close to the p -value, as expected from the Dempster fundamental relation. The empirical 95% equal-tailed credible interval for DD is $(-2.258, 6.266)$. The equivalent interval for the likelihood ratio is $(0.044, 3.09)$ and that for the posterior probability of the null hypothesis (with equal prior

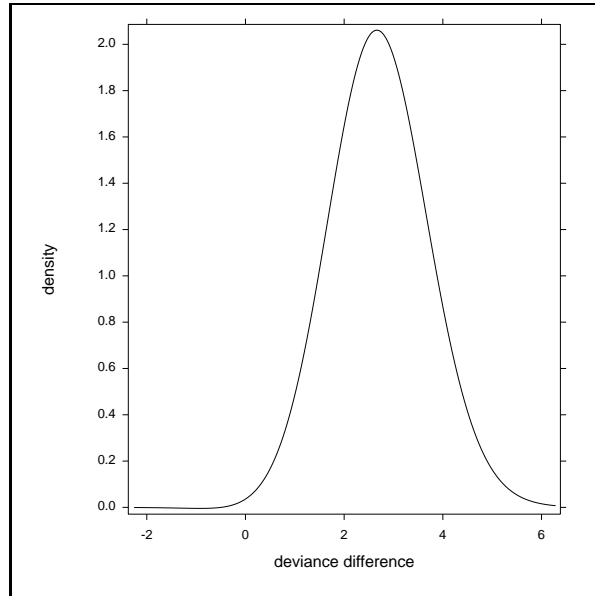


Figure 3.2: Posterior density of DD

probabilities) is (0.042, 0.756). The sample data are consistent with a wide range of values for μ , and hence for the likelihood ratio.

3.2 Two samples, equal variances

3.2.1 Credible interval

The two samples of sizes n_j , $j = 1, 2$ provide sample means \bar{y}_j and the pooled within-sample sum of squares

$$WSS = \sum_{i=1}^{n_1} (y_i - \bar{y}_1)^2 + \sum_{i=n_1+1}^{n_1+n_2} (y_i - \bar{y}_2)^2.$$

The diffuse prior distribution for μ_1, μ_2, σ is $1/\sigma^2$. As for the one-sample case, the joint posterior distribution of (μ_1, μ_2, σ) can be factored in the form

$$\mu_j \mid \sigma, \mathbf{y} \sim \text{independent } N(\bar{y}_j, \sigma^2/n_j), \quad WSS/\sigma^2 \mid \mathbf{y} \sim \chi_{n-2}^2,$$

so that

$$\mu_1 - \mu_2 \mid \sigma, \mathbf{y} \sim N(\bar{y}_1 - \bar{y}_2, \sigma^2(1/n_1 + 1/n_2)),$$

and

$$[\mu_1 - \mu_2 - (\bar{y}_1 - \bar{y}_2)] / [\sigma \sqrt{1/n_1 + 1/n_2}] \mid \sigma, \mathbf{y} \sim N(0, 1).$$

Since this distribution does not depend on σ , the distribution holds unconditionally as well, and hence

$$t = [\mu_1 - \mu_2 - (\bar{y}_1 - \bar{y}_2)] / [s\sqrt{1/n_1 + 1/n_2}] \mid \mathbf{y} \sim t_{n-2}$$

where $n = n_1 + n_2$, $s^2 = WSS/(n-2)$. The $100(1-\alpha)\%$ credible interval for $\mu_1 - \mu_2$ corresponds exactly, for the diffuse prior, to the usual $100(1-\alpha)\%$ t frequentist confidence interval.

As in the one-sample case, simulation methods can be used to generate a random sample from the posterior distribution of $\mu_1 - \mu_2$. We do not give details.

3.2.2 Model comparisons

The null hypothesis is $H_1 : \mu_1 = \mu_2$, the alternative is $H_2 : \mu_1 \neq \mu_2$. A Bayesian problem occurs here: we need to specify the form of the *nuisance parameter* corresponding to the parameter of interest, since we have to express the likelihood ratio in terms of the parameters defined under the alternative model: under the null hypothesis the common mean has to be expressed in terms of μ_1 and μ_2 . In the frequentist maximized likelihood ratio approach this is not necessary, since maximization carries over to any transformations of the parameters.

We choose wherever possible the *orthogonal parametrization* (in the information matrix). This gives near-independence of the parameter of interest and the nuisance parameter(s) in large samples (Cox and Reid 1987). It is easily verified that if we define $\theta = \mu_1 - \mu_2$ and $\phi = (n_1\mu_1 + n_2\mu_2)/(n_1 + n_2)$, the information matrix in θ, ϕ and σ is diagonal, so in large samples the joint posterior distribution of these parameters will have near joint independence. So we define

$$\mu = \frac{n_1\mu_1 + n_2\mu_2}{n}$$

as the common mean under the null hypothesis.

The likelihood under the alternative hypothesis is

$$L(\mu_1, \mu_2, \sigma) = \frac{1}{[\sqrt{2\pi}\sigma]^n} \exp \left\{ -\frac{1}{2\sigma^2} [WSS + n_1(\bar{y}_1 - \mu_1)^2 + n_2(\bar{y}_2 - \mu_2)^2] \right\}$$

and the likelihood ratio and deviance difference are

$$\begin{aligned} LR_{12} &= \exp \left\{ -\frac{1}{2\sigma^2} [n_1(\bar{y}_1 - \mu)^2 - n_1(\bar{y}_1 - \mu_1)^2 + n_2(\bar{y}_2 - \mu)^2 - n_2(\bar{y}_2 - \mu_2)^2] \right\} \\ DD_{12} &= \frac{1}{\sigma^2} [n_1(\bar{y}_1 - \mu_1)^2 - n_1(\bar{y}_1 - \mu)^2 + n_2(\bar{y}_2 - \mu_2)^2 - n_2(\bar{y}_2 - \mu)^2] \\ &= \frac{1}{\sigma^2} \frac{n_1 n_2}{n} \{ (\bar{y}_1 - \bar{y}_2)^2 - [(\mu_1 - \mu_2) - (\bar{y}_1 - \bar{y}_2)]^2 \} \\ &= \frac{n_1 n_2}{n} \frac{(\bar{y}_1 - \bar{y}_2)^2}{s^2} \cdot \frac{s^2}{\sigma^2} - Z^2 \\ &= t^2 W / (n-2) - Z^2 \end{aligned}$$

where t is the two-sample t -statistic, $W = WSS/\sigma^2$ has a χ_{n-2}^2 distribution and

$$Z = \sqrt{\frac{n_1 n_2}{n}} \frac{(\mu_1 - \mu_2) - (\bar{y}_1 - \bar{y}_2)}{\sigma}$$

has a standard normal distribution independent of W . These results exactly parallel those for the one-sample case.

3.2.3 Example

We use the example in Gönen et al (2005), which we also adapted for the one-sample case above. In samples of $n_1 = 10$ and $n_2 = 11$ we observe sample means $\bar{y}_1 = 5.0$ and $\bar{y}_2 = -0.2727$, with sample standard deviations of $s_1 = 8.74$ and $s_2 = 5.90$. The two-sample t -statistic is 1.634, with a p -value of 0.1187. The 95% credible interval for $\mu_1 - \mu_2$ is $\bar{y}_1 - \bar{y}_2 \pm 2.262s\sqrt{\frac{n}{n_1 n_2}}$, here $(-1.025, 12.479)$. This is identical to the 95% confidence interval (with the diffuse priors used).

We make 10,000 draws from the posterior distribution of DD as described in Section 2. The cdf of these values is shown in Figure 3.3, and a smooth approximation to the density in Figure 3.4.

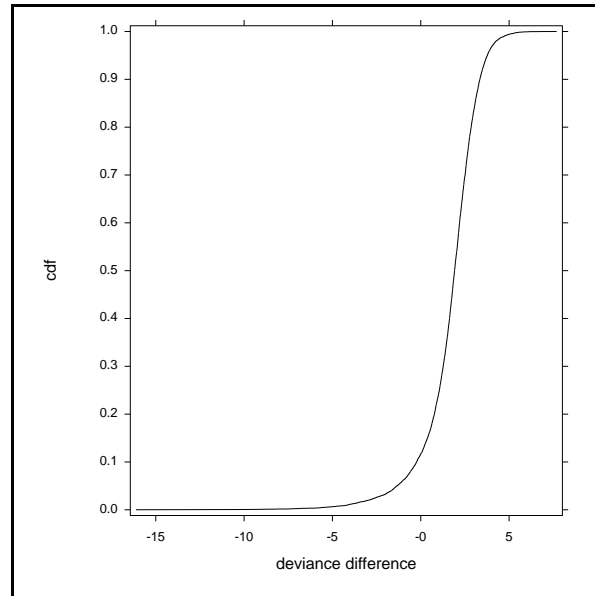


Figure 3.3: Posterior cdf of DD

The empirical probability that $DD < 0$ is .117, with simulation standard error 0.003. This is again very close to the p -value, for the same reason given above. The empirical 95% equal-tailed credible interval for DD is $(-2.576, 4.148)$. This interval is less wide than for the one-sample t -test because the larger sample size gives a better definition of σ . The equivalent interval for the

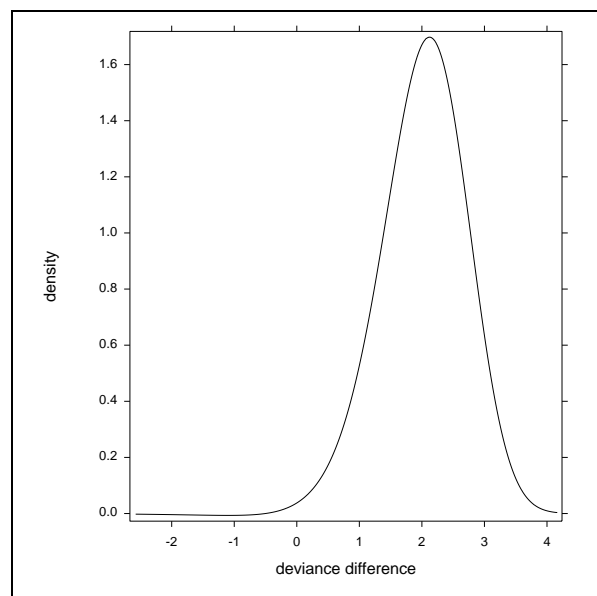


Figure 3.4: Posterior density of DD

likelihood ratio is (0.126, 3.626), and that for the posterior probability of the null hypothesis is (0.112, 0.784).

3.3 Two samples, different variances

3.3.1 Credible interval

The two samples of sizes n_j , $j = 1, 2$ now provide sample means \bar{y}_j , and sums of squares $RSS_j = \sum_{i=1}^{n_j} (y_i - \bar{y}_j)^2$, $j = 1, 2$. The non-informative prior distribution for $\mu_1, \mu_2, \sigma_1, \sigma_2$ is $1/(\sigma_1\sigma_2)$. As for the one-sample case, the joint posterior distribution of $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ can be factored in the form

$$\mu_j \mid \sigma_j, \mathbf{y} \sim N(\bar{y}_j, \sigma_j^2/n), \quad RSS_j/\sigma_j^2 \mid \mathbf{y} \sim \chi_{n_j-1}^2,$$

so that

$$\mu_1 - \mu_2 \mid \sigma_1, \sigma_2, \mathbf{y} \sim N(\bar{y}_1 - \bar{y}_2, \sigma_1^2/n_1 + \sigma_2^2/n_2).$$

Integrating out σ_1 and σ_2 to obtain the exact analytic posterior is complicated (the Behrens-Fisher distribution for the difference between scaled t variables), but simulation methods can be used very easily to generate a sample from the posterior distribution of $\mu_1 - \mu_2$:

- make M random draws of RSS_j/σ_j^2 , $j = 1, 2$ from their marginal $\chi_{n_j-1}^2$ distributions, and hence M random draws $\sigma_j^{[m]}$ of σ_j ;

- for each $\sigma_j^{[m]}$, make a random draw $\mu_j^{[m]}$ of μ_j from its conditional normal $N(\bar{y}_j, \sigma_j^{[m]2}/n_j)$ distribution;
- compute $\delta^{[m]} = \mu_1^{[m]} - \mu_2^{[m]}$.

Alternatively (Tanner 1996 p. 29) we can make M random draws from t_1 and t_2 and convert them to μ_1 and μ_2 , and then δ .

Credible intervals for δ can be computed as above from the empirical distribution of the $\delta^{[m]}$. The credible intervals in this case correspond exactly, for the noninformative priors, to the fiducial Behrens-Fisher intervals (Tanner 1996 pp. 20-21).

3.3.2 Model comparison

The null hypothesis is $H_1 : \mu_1 = \mu_2$, the alternative is $H_2 : \mu_1 \neq \mu_2$. As in the previous case of equal variances, we need to define the common mean under the null hypothesis. It is easily verified that the information orthogonalizing choice for the nuisance parameter is

$$\mu = \left[\frac{n_1\mu_1}{\sigma_1^2} + \frac{n_2\mu_2}{\sigma_2^2} \right] / \left[\frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2} \right].$$

The likelihood under the alternative hypothesis is

$$\frac{1}{[\sqrt{2\pi}]^n \sigma_1^{n_1} \sigma_2^{n_2}} \exp \left\{ -\frac{1}{2\sigma_1^2} [RSS_1 + n_1(\bar{y}_1 - \mu_1)^2] - \frac{1}{2\sigma_2^2} [RSS_2 + n_2(\bar{y}_2 - \mu_2)^2] \right\}$$

and the likelihood ratio and deviance difference are

$$\begin{aligned} LR_{12} &= \exp \left\{ -\frac{n_1}{2\sigma_1^2} [(\bar{y}_1 - \mu)^2 - (\bar{y}_1 - \mu_1)^2] - \frac{n_2}{2\sigma_2^2} [(\bar{y}_2 - \mu)^2 - (\bar{y}_2 - \mu_2)^2] \right\} \\ DD_{12} &= \frac{n_1}{\sigma_1^2} [(\bar{y}_1 - \mu_1)^2 - (\bar{y}_1 - \mu)^2] + \frac{n_2}{\sigma_2^2} [(\bar{y}_2 - \mu)^2 - (\bar{y}_2 - \mu_2)^2] \\ &= \{ (\bar{y}_1 - \bar{y}_2)^2 - [(\mu_1 - \mu_2) - (\bar{y}_1 - \bar{y}_2)]^2 \} / \left[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right] \\ &= \frac{(\bar{y}_1 - \bar{y}_2)^2}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \cdot \frac{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} - Z^2 \\ &= t^{*2} \cdot \frac{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} - Z^2 \end{aligned}$$

where

$$Z = \frac{(\mu_1 - \mu_2) - (\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

has a standard normal distribution independent of σ_1^2 and σ_2^2 ,

$$t^* = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

is the Welch t^* statistic proposed as an approximate t -test alternative to the Behrens-Fisher test, and $s_j^2 = RSS_j/(n_j - 1)$. The distribution of DD_{12} has no closed form density, but can be directly simulated by generating M independent values of the two χ^2 variates RSS_1/σ_1^2 and RSS_2/σ_2^2 and of Z , and forming the M values of DD_{12} .

3.3.3 Example

We use the same example but now do not assume the variances are equal. We compare the Bayes analysis with the Welch approximation which uses t^* with fractional degrees of freedom ν :

$$\nu = \left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right]^2 / \left[\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1} \right].$$

The t^* statistic is 1.742, with approximate degrees of freedom 15.59, and approximate p -value of 0.102.

We make 10,000 random draws from the posterior distribution of DD_{12} as described in Section 2. The cdf of these values is shown in Figure 3.5, and a smooth approximation to the density in Figure 3.6.

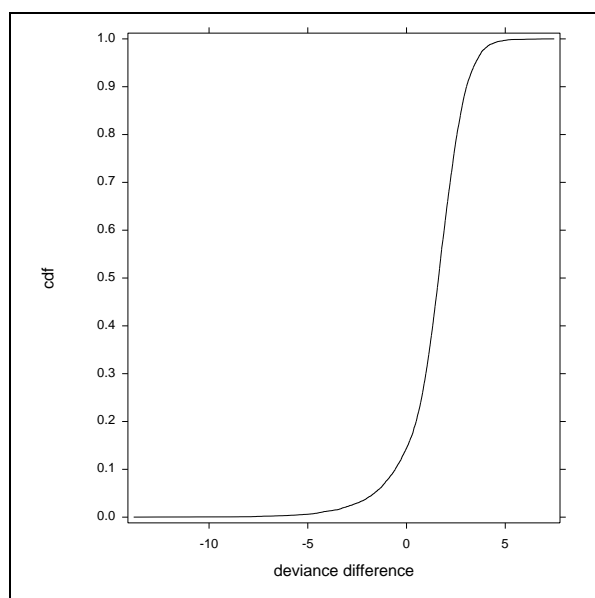


Figure 3.5: Posterior cdf of DD

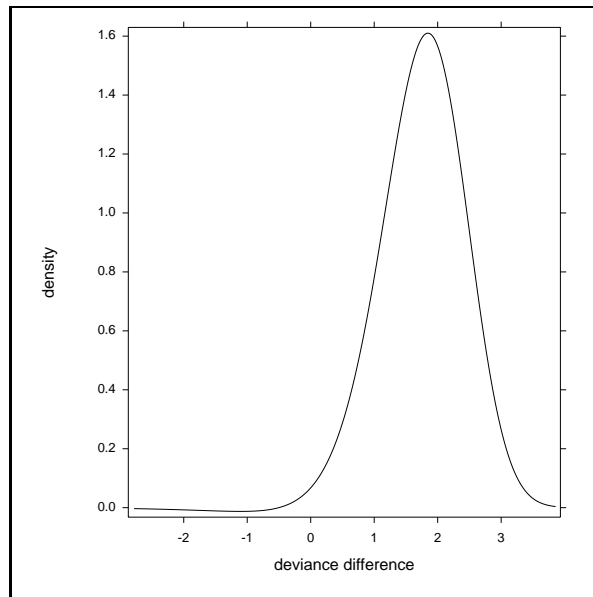


Figure 3.6: Posterior density of DD

The empirical probability that $DD < 0$ is .144, with simulation standard error 0.004. The empirical 95% equal-tailed credible interval for DD is $(-2.782, 3.842)$, and the intervals for the likelihood ratio and posterior probability of the null hypothesis are $(0.147, 4.02)$ and $(0.128, 0.801)$. These intervals give even less evidence against the null hypothesis, as expected from the weaker model assumption of different variances. The empirical probability that $LR < 1/9$ (i.e. that $DD > 2 \log 9 = 4.394$) is 0.009.

The Bayesian version gives somewhat more support to the null hypothesis than the Welch test.

3.4 The Gönen et al two-sample test

The Bayesian test proposed by Gönen et al reparametrises from (μ_1, μ_2, σ) to (μ, δ, σ) , where $\mu = (\mu_1 + \mu_2)/2$ and δ is the effect size: $\delta = (\mu_2 - \mu_1)/\sigma$. They use non-informative priors for the common mean and σ : $\pi(\mu, \sigma) = c/\sigma^2$. The (independent) prior for the effect size is normal, with mean and variance hyper-parameters λ and σ_δ^2 . These have to be set by the user of the test.

For the computation of the Bayes factor, the likelihood under each hypothesis is integrated over μ and σ ; the difficulty that this prior is improper is finessed by the device of using a proper prior which is made to approach the limiting improper prior, and requiring this limiting process to be the same under each hypothesis, so that the integrating constant in the proper prior cancels out.

The form of the Bayes factor resulting is (Gönen et al p. 253)

$$BF = \frac{T_\nu(t|0, 1)}{T_\nu(t|n_\delta^{1/2}\lambda, 1 + n_\delta\sigma_\delta^2)},$$

where t is the value of the frequentist two-sample test statistic, $n_\delta^{-1} = n_1^{-1} + n_2^{-1}$, and $T_\nu(\cdot|a, b)$ denotes the noncentral t probability density function having location a , scale $b^{1/2}$ and degrees of freedom ν . The integrated likelihood under the alternative, and the Bayes factor, depend explicitly on the values of the hyper-parameters.

For the hyper-parameter values $\lambda = 0$ and $\sigma_\delta = 1/3$, the Bayes factor is 0.791, equivalent to a posterior probability of H_0 of 0.442. The posterior probability is sensitive to both λ and σ_δ : Figures 1 and 2 in Gönen et al show that this probability varies from 0.25 to 1.0 over the ranges of parameter values $\lambda \in (-1, 1)$, $\sigma_\delta \in (0.01, 1.00)$ considered. For comparison, the 95% credible interval for this posterior probability from Section 3.2.1 is (0.112, 0.784).

3.4.1 Informative prior for the effect size

The choice of the effect size as the parameter of interest, and its informative prior, greatly complicate the analysis, because the three new model parameters are no longer information-orthogonal. The hyper-parameters in the effect size prior have to be specified by the user; it is certainly not clear from its formulation how informative the prior is relative to the information in the likelihood. We can evaluate the latter fairly easily.

From the standard analysis for the two-sample problems earlier we have that for diffuse priors on μ_1 and μ_2 , the posterior distribution of $\mu_2 - \mu_1$ given σ is $N(\bar{y}_2 - \bar{y}_1, \sigma^2(1/n_1 + 1/n_2))$, and therefore the posterior distribution of δ given σ is $N((\bar{y}_2 - \bar{y}_1)/\sigma, 1/n_1 + 1/n_2)$. So the conditional (given σ) variance of δ in the example is $1/10 + 1/11 = 0.10909$; the unconditional variance must be larger since this is increased by the variance of the conditional mean. The conditional posterior standard deviation is 0.330.

The range of prior standard deviations considered by Gönen et al is (0.01, 1.00). For one-third of this range the prior for δ is more informative than the likelihood; for the example they quote of $\lambda = 0$, $\sigma_\delta = 1/3$, the prior will be more informative than the likelihood so the posterior mean for δ will be shrunk towards zero. This is reflected in the small Bayes factor of 0.791.

We are not arguing against the formation by subject-matter experts of informative priors for the effect size. But it is important for the expert, and even more for the novice, to know how the information in the prior and the likelihood are combined in the final analysis. For this reason we use wherever possible *diffuse* priors, to allow the information *in the data* to be assessed.

3.5 The normal variance test

We wish to test a null hypothesis $H_1 : \sigma^2 = \sigma_1^2$ against a general alternative, in the normal model $Y \sim N(\mu, \sigma^2)$, given data $\mathbf{y} = (y_1, \dots, y_n)$. The likelihoods under the null and alternative hypotheses, and their ratio and the deviance difference, are

$$\begin{aligned}
 L_1 = L(\mu, \sigma_1) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2} \frac{(y_i - \mu)^2}{\sigma_1^2}\right\} \\
 L_2 = L(\mu, \sigma) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2} \frac{(y_i - \mu)^2}{\sigma^2}\right\} \\
 LR_{12} = L_1/L_2 &= \left[\frac{\sigma^2}{\sigma_1^2}\right]^{n/2} \exp\left\{-\frac{1}{2} \left[\frac{RSS}{\sigma_1^2} + \frac{n(\bar{y} - \mu)^2}{\sigma_1^2} - \frac{RSS}{\sigma^2} - \frac{n(\bar{y} - \mu)^2}{\sigma^2}\right]\right\} \\
 D_{12} = -2 \log LR_{12} &= n \log \left[\frac{RSS/\sigma_1^2}{RSS/\sigma^2}\right] + \frac{RSS}{\sigma_1^2} + \frac{n(\bar{y} - \mu)^2}{\sigma^2} \left[\frac{RSS/\sigma_1^2}{RSS/\sigma^2}\right] \\
 &\quad - \frac{RSS}{\sigma^2} - \frac{n(\bar{y} - \mu)^2}{\sigma^2} \\
 &= n(\log w_1 - \log W) + w_1 - W + Z^2(w_1/W - 1),
 \end{aligned}$$

where $w_1 = RSS/\sigma_1^2$, $W = RSS/\sigma^2$ and $Z^2 = n(\bar{y} - \mu)^2/\sigma^2$. There is no exact posterior distribution for D_{12} , but it is easily simulated, as for the Bayesian t-tests.

An alternative approach often used for frequentist testing is through the *marginal* or *restricted* likelihood for σ^2 based only on RSS . Non-Bayesian justifications for ignoring the additional term in the likelihood which depends on μ and σ have been heuristic, relying on the claim of lack of “available information” (Kalbfleisch and Sprott 1970) about σ from the \bar{y} component, in the “absence of knowledge” about μ .

If we use the marginal likelihood for the same test, we have the marginal null and alternative likelihoods, and their ratio and deviance difference:

$$\begin{aligned}
 M_1 = M(\sigma_1) &= \frac{1}{\sigma_1^\nu} \exp\left\{-\frac{RSS}{2\sigma_1^2}\right\} \\
 M_2 = M(\sigma) &= \frac{1}{\sigma^\nu} \exp\left\{-\frac{RSS}{2\sigma^2}\right\} \\
 MLR_{12} = M_1/M_2 &= \left[\frac{\sigma^2}{\sigma_1^2}\right]^{\nu/2} \exp\left\{-\frac{RSS}{2} \left[\frac{1}{\sigma_1^2} - \frac{1}{\sigma^2}\right]\right\} \\
 MD_{12} = -2 \log MLR_{12} &= \nu \log \left[\frac{\sigma^2}{\sigma_1^2}\right] + \frac{RSS}{\sigma_1^2} - \frac{RSS}{\sigma^2} \\
 &= \nu \log \left[\frac{RSS/\sigma_1^2}{RSS/\sigma^2}\right] + \frac{RSS}{\sigma_1^2} - \frac{RSS}{\sigma^2} \\
 &= \nu(\log w_1 - \log W) + w_1 - W,
 \end{aligned}$$

where $\nu = n - 1$. This differs from D_{12} by the term

$$\log(w_1/W) + Z^2(w_1/W - 1).$$

It is easily seen that if $w_1 < W$, both D_{12} and MD_{12} are negative, while if $w_1 > W$, both D_{12} and MD_{12} are positive. So the posterior probability that the null hypothesis is better supported than the alternative *is the same under both likelihoods*: they lead to the same conclusion about the posterior probability that the likelihood ratio is greater than 1. However the posterior probabilities for *strong* support are not the same, since the percentiles of the likelihood deviance distribution are always more extreme than those of the marginal likelihood deviance distribution.

So the full likelihood *always provides stronger evidence against the null hypothesis* than the marginal likelihood, *except for the case of a simple preference for $L_{12} > 1$ or $L_{12} < 1$* , when they provide the *same* evidence. Figure 3.7 shows an example with $\sigma_1 = 3$ and unbiased standard deviation $s = 4$, in a sample of $n = 20$ using $M = 10,000$ draws. The full likelihood curve is solid, the marginal likelihood curve is dotted. The probability that the likelihood ratio $L_{21} > 1$ is 0.933 for both likelihoods (corresponding to a p -value of 0.067), but the probability that $L_{21} > 9$ is 0.853 for the full likelihood, but 0.846 for the marginal likelihood.

Using the likelihood distribution resolves the argument over whether the full or the marginal likelihood should be used. From the posterior likelihood distribution point of view, one should always use the full likelihood (as in all other cases!), since the evidence for a simple preference is the same as for the marginal likelihood, but the evidence for a strong preference is always greater.

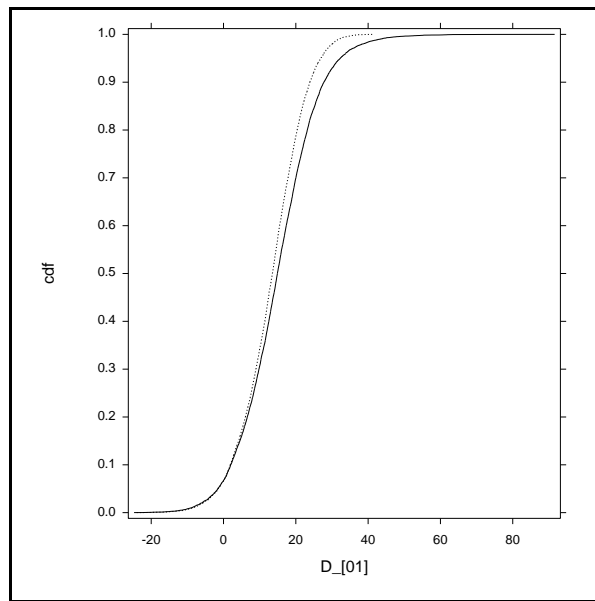


Figure 3.7: Posterior distributions of D_{12} and MD_{12} , variance test