Even more on Bayes and conjugate forms

Saad Mneimneh

1 A story about brewing

If we have no a priori knowledge of the variance σ^2 (and mean μ), then

$$s^2 = \frac{S}{n-1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

has an expected value equal to σ^2. s^2, so calculated, is a perfectly good estimate of σ^2, but it is seldom or never equal to σ^2. If $x_1, \ldots, x_n$ are independent and identically normally distributed, William Gosset, while working at Guinness brewery, was the first to point out that if we substitute s^2 for σ^2, we have no right to believe that

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

will still be normally distributed (recall that $(\bar{x} - \mu)\sqrt{n}/\sigma \sim N(0,1)$). Gosset investigated the distribution of t, and published preliminary results under the pseudonym "Student", to avoid official disclosure of Guinness confidential information. He came up with the t-distribution, but the math was incomplete. Later, Fisher perfected the work of "Student" which lead to our current knowledge of t-tests.

2 The t-distribution

Let $Z \sim N(0,1)$ and $V \sim \chi^2_k$ be two independent random variables, and consider the following random variable:

$$t = \frac{Z}{\sqrt{V/k}}$$

The density of t can then be shown to be the following, with parameter k (degrees of freedom):

$$f(t) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})}(1 + \frac{t^2}{k})^{-\frac{k+1}{2}}$$

This is the form that was obtained by Gosset in his fundamental paper. However, the math was incomplete. Fisher proved later that:

- $S/\sigma^2 = (n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$
- $\bar{x}$ and s^2 are independent (this is a nice property of the estimated normal mean and variance)

Thus,

$$t = \frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2/\sigma^2}{n-1}}} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

produces the desired random variable.

We are going to prove the two facts listed above for the case of $n = 2$:

$$\bar{x} = \frac{x_1 + x_2}{2}$$

$$s^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2}{1} = \frac{(x_1-x_2)^2}{2}$$

The last equality is given by replacing $\bar{x}$ with $(x_1+x_2)/2$. Therefore,

$$\frac{(n-1)s^2}{\sigma^2} = \frac{(x_1-x_2)^2}{2\sigma^2} = \left(\frac{x_1-x_2}{\sqrt{2}\sigma}\right)^2$$

Note that $x_1 - x_2$ has a zero mean and a variance of $2\sigma^2$. Therefore, the expression above corresponds to the square of a standard normal variable, which is χ^2_1 distributed.

Now we show the independence of $\bar{x}$ and s^2 (again for the case of $n = 2$). It will be enough to show that $x_1 + x_2$ and $(x_1 - x_2)^2$ are independent (see the expressions for $\bar{x}$ and s^2 respectively). We will achieve this result by showing that the joint density is equal to the product of the individual densities. First observe that $f(x_1, x_2) = f(x_1)f(x_2)$ because x_1 and x_2 are independent. Now define $y_1 = x_1 + x_2$ and $y_2 = (x_1 - x_2)^2$. We need to find $f(y_1, y_2)$.

To do so, we may refer to the following generalization of the single variable case (there must be a one-to-one correspondence between $[y_1, \ldots, y_n]$ and the $[x_1, \ldots, x_n]$; otherwise, we add up the contribution of the different solutions):

$$f(y_1, \ldots, y_n) = f(x_1, \ldots, x_n) \cdot \left|J\right|^{-1}$$

where J is the determinant of the following matrix:

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_1} \cdots \frac{\partial y_1}{\partial x_n} \\ \ddots \\ \frac{\partial y_n}{\partial x_1} \cdots \frac{\partial y_n}{\partial x_n} \end{bmatrix}$$

In our case, we have

$$\begin{bmatrix} \frac{\partial(x_1+x_2)}{\partial x_1} & \frac{\partial(x_1+x_2)}{\partial x_2} \\ \frac{\partial(x_1-x_2)^2}{\partial x_1} & \frac{\partial(x_1-x_2)^2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2(x_1-x_2) & -2(x_1-x_2) \end{bmatrix}$$

The determinant of the above matrix is $J = -4(x_1 - x_2)$; therefore, $|J| = 4|x_1 - x_2| = 4\sqrt{y_2}$.

$$f(y_1, y_2) = \sum_{(x_1,x_2)\text{ solutions}} f(x_1, x_2)|J|^{-1}$$

$$= \sum_{(x_1,x_2)\text{ solutions}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_2-\mu)^2}{2\sigma^2}} \frac{1}{4\sqrt{y_2}}$$

Now to express x_1 and x_2 in terms of y_1 and y_2, we have two solutions: $x_1 = (y_1+\sqrt{y_2})/2, x_2 = (y_1-\sqrt{y_2})/2$ and $x_1 = (y_1-\sqrt{y_2})/2, x_2 = (y_1+\sqrt{y_2})/2$. The two solutions are symmetric in x_1 and x_2, so adding their contribution is equivalent to a multiplcation by 2.

$$f(y_1, y_2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{[(y_1+\sqrt{y_2})/2-\mu]^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{[(y_1-\sqrt{y_2})/2-\mu]^2}{2\sigma^2}} \frac{2}{4\sqrt{y_2}}$$

After some manipulation, we get:

$$f(y_1, y_2) = \frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)} e^{-\frac{(y_1-2\mu)^2}{2(\sqrt{2}\sigma)^2}} \frac{1/(2\sigma^2)}{2\sqrt{\pi}} \big(\frac{y_2}{4\sigma^2}\big)^{1-1/2} e^{-\frac{y_2}{4\sigma^2}} = f(y_1)f(y_2)$$

As an additional excercise, let us now derive the form of the density for $t = Z/\sqrt{V/k}$ (Z and V are independent, $N(0,1)$ and χ^2_k distributed respectively). First, we find the density $f(z, v)$ by multiplying $f(z)$ and $f(v)$.

$$f(z, v) \propto e^{-\frac{z^2}{2}} v^{\frac{k}{2}-1} e^{-\frac{v}{2}}$$

Then, consider the two variables $t = z(k/v)^{1/2}$ and $r = v$.

$$J = \begin{bmatrix} \frac{\partial t}{\partial z} & \frac{\partial t}{\partial v} \\ \frac{\partial r}{\partial z} & \frac{\partial r}{\partial v} \end{bmatrix} = \begin{bmatrix} (\frac{k}{v})^{1/2} & -\frac{z}{2v^2}(\frac{k}{v})^{-1/2} \\ 0 & 1 \end{bmatrix}$$

So $|J|^{-1} = (\frac{v}{k})^{1/2}$. Note that $z^2 = t^2 r/k$. Therefore,

$$f(t, r) = f(z, v)|J|^{-1} \propto e^{-\frac{t^2 r}{2k}} r^{\frac{k-1}{2}} e^{-\frac{r}{2}} = r^{\frac{k+1}{2}-1} e^{-\frac{r}{2}(1+t^2/k)}$$

$$f(t) \propto \int_0^\infty e^{-\frac{r}{2}(1+t^2/k)} r^{\frac{k+1}{2}-1} dr$$

Making a change of variable $x = r(1 + t^2/k)/2$, we get:

$$f(t) \propto \int_0^\infty (1 + \frac{t^2}{k})^{-\frac{k+1}{2}} x^{\frac{k+1}{2}-1} e^{-x} dx$$

$$= (1 + \frac{t^2}{k})^{-\frac{k+1}{2}} \int_0^\infty x^{\frac{k+1}{2}-1} e^{-x} dx = \Gamma\big(\frac{k+1}{2}\big)\big(1 + \frac{t^2}{k}\big)^{-\frac{k+1}{2}}$$

3 The t-test

Consider the problem of deciding on the mean of two groups (see note 6). This time, however, the variance σ^2 for each group is also unknown. Therefore, we have to estimate σ_x^2 and σ_y^2 using s_x^2 and s_y^2 respectively. We form the following two random variables:

$$\frac{(n-1)s_x^2}{\sigma_x^2} \sim \chi^2_{n-1}$$

$$\frac{(m-1)s_y^2}{\sigma_y^2} \sim \chi^2_{m-1}$$

where n and m represent the size of the two groups. Using the properties of the χ^2 distribution, we know that:

$$\frac{(n-1)s_x^2}{\sigma_x^2} + \frac{(m-1)s_y^2}{\sigma_y^2} \sim \chi^2_{n+m-2}$$

We also know that:

$$\frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \sim N(0,1)$$

Therefore, if we further assume that $\sigma_x^2 = \sigma_y^2$, we can form a Student distributed random variable t with $n+m-2$ degrees of freedom, as follows:

$$t = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{(1/n + 1/m)\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}}$$

Testing whether $\mu_x = \mu_y$ would then reduce to checking how extreme t is when $\mu_x - \mu_y = 0$, thus:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(1/n + 1/m)\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}}$$

Example: Consider the following data for the two groups:

group 1

25	17	29	29	26	24	27	33
23	14	21	26	20	27	26	32
20	32	17	23	20	30	26	12

group 2

21	26	28	31	14	27	29	23
18	25	32	23	16	21	17	20
26	23	7	18	29	32	24	19

We compute the following:

$$n = m = 24$$
$$\bar{x} = 24.13$$
$$\bar{y} = 22.88$$
$$s_x^2 = 31.81$$
$$s_y^2 = 37.70$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(1/n + 1/m)\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}} = \frac{1.25}{1.698} = 0.736$$

What is an extreme value for t? We can consult the tables. The degree of freedom is $k = n + m - 2 = 46$. However, in the table we find values corresponding to $k = 40$ and $k = 50$. Let's choose the one that gives a higher value for the extreme (giving the hypothesis a better chance), and that is $k = 40$, and we see that $P(t \geq 1.684) \leq 0.05$. Since $0.736 < 1.684$, it is not considered extreme, and we may adopt the hypothesis that $\mu_x = \mu_y$. In fact, $k = 50$ would lead to the same decision.

What if σ_x^2 and σ_y^2 are not assumed to be equal? Then the t-test is not valid, and we must revert to approximations. This will be left for the reader to investigate if interested.

4 Student as a conjugate prior

We will show in this section that the Student density is a conjugate prior for the case when the mean and the variance of a sample are both unknown. Therefore, let $x_1, \ldots, x_n$ be n independent normally distributed observations,

$$X_i|\mu, \sigma^2 \sim N(\mu, \sigma^2)$$

and consider the priors:

$$S_0/\sigma^2 \sim \chi^2_{k_0}$$
$$\mu|\sigma^2 \sim N(\beta, \sigma^2/n_0)$$

for some S_0, k_0, and n_0.

First we show that such prior is actually Student distributed. For this, consider $f(\mu, \sigma^2) = f(\mu|\sigma^2)f(\sigma^2)$ (note that we need to multiply $f(S_0/\sigma^2)$ by $|d(S_0/\sigma^2)/d\sigma^2|$ to get $f(\sigma^2)$.

$$f(\mu, \sigma^2) \propto \frac{1}{\sigma}e^{-\frac{(\mu-\beta)^2}{2\sigma^2/n_0}} \cdot (\sigma^2)^{-k_0/2-1}e^{-\frac{S_0}{2\sigma^2}} = (\sigma^2)^{-(k_0+1)/2-1}e^{-\frac{S_0+n_0(\mu-\beta)^2}{2\sigma^2}}$$

Then,

$$f(\mu) = \int_0^\infty f(\mu, \sigma^2) d\sigma^2$$
$$\propto \int_0^\infty (\sigma^2)^{-(k_0+1)/2-1} e^{-\frac{S_0+n_0(\mu-\beta)^2}{2\sigma^2}} d\sigma^2$$

Let's make the change of variable $t = \frac{S_0+n_0(\mu-\beta)^2}{2\sigma^2}$, then the integral reduces to:

$$\int_0^\infty t^{(k_0+1)/2-1} e^{-t} dt / [S_0+n_0(\mu-\beta)^2]^{(k_0+1)/2} = \Gamma\Big(\frac{k_0+1}{2}\Big)[S_0+n_0(\mu-\beta)^2]^{-(k_0+1)/2}$$

Therefore,

$$f(\frac{\mu-\beta}{\sqrt{S_0/n_0k_0}}) \propto [S_0 + n_0(\mu-\beta)^2]^{-(k_0+1)/2} \propto \Big[1 + \frac{\Big(\frac{\mu-\beta}{\sqrt{S_0/n_0k_0}}\Big)^2}{k_0}\Big]^{-(k_0+1)/2}$$

which means that the random variable between parenthesis is Student distributed with k_0 degrees of freedom.

Let us now derive the posterior.

$$f(\mu, \sigma^2 | x_1, \ldots, x_n) = f(x_1, \ldots, x_n | \mu, \sigma^2) f(\mu, \sigma^2)$$
$$\propto (\sigma^2)^{-n/2} e^{-\frac{S+n(\bar{x}-\mu)^2}{2\sigma^2}} (\sigma^2)^{-(k_0+1)/2-1} e^{-\frac{S_0+n_0(\mu-\beta)^2}{2\sigma^2}}$$
$$= (\sigma^2)^{-(n+k_0+1)/2-1} e^{-\frac{S+S_0+n(\mu-\bar{x})^2+n_0(\mu-\beta)^2}{2\sigma^2}}$$

where $S = \sum_i (x_i - \bar{x})^2$. This gives $S'/\sigma^2 \sim \chi^2_{n+k_0}$ (integrate with respect to μ), where S' is defined below. By using a change of variable similar to above, we also obtain:

$$f(\frac{\mu-\beta'}{s'/\sqrt{n'}} | x_1, \ldots, x_n) \propto \Big[1 + \frac{\Big(\frac{\mu-\beta'}{s'/\sqrt{n'}}\Big)^2}{k}\Big]^{-(k+1)/2}$$

where

$$k = n + k_0$$
$$n' = n + n_0$$
$$\beta' = (n_0\beta + n\bar{x})/n'$$
$$S' = S_0 + S + n_0\beta^2 + n\bar{x}^2 - n'\beta'^2$$
$$s'^2 = S'/k$$

Note that when $S_0 = 0$, $n_0 = 0$, and $k_0 = -1$, we have

$$n' = n$$
$$\beta' = \bar{x}$$
$$s' = s$$
$$k = n - 1$$

and, therefore, we retrieve the classical results:

$$S/\sigma^2 \sim \chi^2_{n-1}$$
$$\frac{\mu - \bar{x}}{s/\sqrt{n}} \sim t_{n-1}$$

which shows again that the sampling approach is a special case of a more general Bayesian approach. The prior (improper as usual) in this case will be:

$$f(\mu, \sigma^2) = \frac{1}{\sigma^2}$$

Example: In the spirit of Gosset's work, assume that growing wheat has a yield per plot that is believed to be normally distributed. Assume also that the prior distribution of the variance has a mean of 300 and a variance of 25600. As for the mean, it is expected to be around 110 and this information is thought to be worth about 15 observations.

Let's see how we can use this information to set up the prior. Based on our prior for σ^2, we can compute the following:

$$E[\sigma^2] = \frac{S_0}{k_0 - 2} = 300$$
$$Var(\sigma^2) = \frac{2S_0}{(k_0 - 2)^2(k_0 - 4)} = 25600$$

This means that $k_0 = 11$ and $S_0 = 2700$. The other information gives $\beta = 110$ and $n_0 = 15$. Now if we actually observe the following:

$$141, 102, 73, 171, 137, 91, 81, 157, 146, 69, 121, 134$$

then $n = 12$, $\bar{x} = 119$, and $S = 13045$. Using this, the parameters for the posterior will be:

$$k = k_0 + n = 23$$
$$n' = n_0 + n = 27$$
$$\beta' = (n_0\beta + n\bar{x})/n' = 114$$
$$S' = S_0 + S + n_0\beta^2 + n\bar{x}^2 - n'\beta'^2 = 16285$$
$$s' = \sqrt{S'/k} = 26.61$$

Then

$$\frac{\mu - 114}{5.1}$$

is Student distributed with 23 degrees of freedom. Using tables for the Student distribution with $k = 23$, we find that $\mu \in [103, 125]$ with 95% probability.

Example: Let us revisit the question on the means of two groups. Let $\bar{x} - \bar{y} = d$ and $\mu_x - \mu_y = w$. Assume

$$d|w, \sigma^2 \sim N(w, \sigma^2(1/n + 1/m))$$

$$\mu_x|\sigma^2 \sim \mu_y|\sigma^2 \sim N(\beta, \sigma^2/n_0) \Rightarrow w|\sigma^2 \sim N(0, 2\sigma^2/n_0)$$

$$S_0/\sigma^2 \sim \chi^2_{k_0}$$

We would like to obtain a posterior density for $w|d$. We can rewrite the above by making a change of variable. Let $\tau^2 = \sigma^2(1/n + 1/m)$. Then:

$$d|w, \tau^2 \sim N(w, \tau^2)$$

$$w|\tau^2 \sim N(0, \frac{\tau^2}{\frac{n+m}{nm} n_0/2})$$

$$\frac{S_0 \frac{n+m}{nm}}{\tau^2} \sim \chi^2_{k_0}$$

So we can use the following general form (and then replace n_0 and S_0 and β accordingly):

$$d|w, \tau^2 \sim N(w, \tau^2)$$

$$w|\tau^2 \sim N(\beta, \frac{\tau^2}{n_0})$$

$$\frac{S_0}{\tau^2} \sim \chi^2_{k_0}$$

which is equivalent to the previous problem with $n = 1$ ($\bar{d} = d$, $S = 0$). By letting $n_0 = 0$, $S_0 = 1$, $\beta = 0$, and then replacing S_0 with $(n + m)/(nm)$, we get that

$$\frac{w - d}{\sqrt{\frac{1}{1+k_0}}\sqrt{1/n + 1/m}} = \frac{w - d}{v\sqrt{1/n + 1/m}}$$

is Student distributed with $1 + k_0$ degrees of freedom. The interpretation is that when k_0 is large (student density approximates standard normal), w behaves like a normal variable with mean d and some variance $v^2(1/n + 1/m)$ where v^2 is small.