Finding Student t Distribution

Maths Note | Larry Cui

May 9, 2022

This is a note on heavy stuff. Facing variables from a normal distribution, if we want to test the null hypothesis, we rely on the the statistic:

$$Z = \frac{\overline{Y} - \mu_0}{\sigma/\sqrt{n}}$$

Thanks to the central limit theorem, we know the above distribution is normal, and we can use $f(z)$ to find the probabilities we need.

However, in reality we may never know the variance of the population, and we have only a limited sample size due to economical reason as well as other restraints.

Our best guess about the variance is S^2, also derived from sample:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

and we use `t-ratio` to test our inference,

$$T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}$$

It turns out the pdf of t ratio is not exactly the same as normal distribution. Oxford graduate William Sealy Gossett published a paper in 1908 in which he derived a formula for the pdf, and later in 1924 R. A. Fisher presented a rigorous mathematical derivation of Gossett's pdf. Consequently, the pdf is now known as the *Student t distribution*.

It involves a lot of work to get the pdf of t ratio. We will start with some background function ideas, and from section 2 we start proving, while we also attach some lemmas and their proofs to appendix at the end of the note.

1 Review of some functions

1.1 error function

The error function is defined as

$$\text{erf}(x) = \frac{1}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt$$

It has no analytical integrand. From the definition, we have its derivative

$$\mathrm{erf}(x)' = \frac{1}{\sqrt{\pi}} e^{-x^2}$$

1.2 gamma function

$$\Gamma(r) = \int_0^\infty y^{r-1} e^{-y}\, dy$$

We've discussed some features of the gamma function in previous chapter (4.6). Here we prove $\Gamma(1/2) = \sqrt{\pi}$. First of all we prove its square value and the root will present itself.

$$\Gamma(1/2)^2 = \left(\int_0^\infty t^{-1/2} e^{-t}\, dt\right)^2$$

Let $x^2 = t$, then the above equation becomes

$$\begin{aligned}
\left(\int_0^\infty t^{-1/2} e^{-t}\, dt\right)^2 &= \left(2\int_0^\infty e^{-x^2}\, dx\right)^2 = \left(\int_{-\infty}^\infty e^{-x^2}\, dx\right)\left(\int_{-\infty}^\infty e^{-y^2}\, dy\right) && \triangleright \text{ symmetrical integral} \\
&= \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-(x^2+y^2)}\, dx\, dy = \int_0^{2\pi} \int_0^\infty e^{-r^2} r\, dr\, d\theta && \triangleright \text{ to polar coordinates} \\
&= \pi
\end{aligned}$$

As a result, we can see that $\Gamma(1/2) = \sqrt{\pi}$.

1.3 additive property

Based on gamma function, we have gamma distribution with parameters (r, λ)

$$f_Y(y) = \underbrace{\frac{\lambda^r}{\Gamma(r)}}_{\text{constant}} y^{r-1} e^{-\lambda y}, \quad y \geqslant 0$$

Additivity Suppose two independent variables U has the gamma pdf with parameters r and λ, V with s and the same λ,

$$f_U(u) = \underbrace{\frac{\lambda^r}{\Gamma(r)}}_{\text{constant}} u^{r-1} e^{-\lambda u}, \quad u \geqslant 0 \qquad \text{and} \qquad f_V(v) = \underbrace{\frac{\lambda^r}{\Gamma(s)}}_{\text{constant}} v^{s-1} e^{-\lambda v}, \quad v \geqslant 0$$

Then $U + V$ has a gamma pdf with parameters $r + s$ and λ:

$$f_{U+V}(u+v) = \underbrace{\frac{\lambda^{r+s}}{\Gamma(r+s)}}_{\text{constant}} (u+v)^{r+s-1} e^{-\lambda(u+v)}, \quad (u+v) \geqslant 0$$

Proof We proved it half way in chapter 4.6 before we learn beta function. Now we use beta to fully prove this lemma. Let $t = u + v$, then $t \geqslant 0$ and the distribution function of t is

$$\begin{aligned}
f_{U+V}(t) &= \int_0^t f_U(u) f_V(t-u)\, du && \triangleright \text{ pdf of } t \text{ is joint pdf of } u \text{ and } v \\
&= \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} \int_0^t u^{r-1} e^{-\lambda u} (t-u)^{s-1} e^{-\lambda(t-u)}\, du \\
&= \frac{\lambda^{r+s} e^{-\lambda t}}{\Gamma(r)\Gamma(s)} \int_0^t u^{r-1} (t-u)^{s-1}\, du \\
&= \frac{\lambda^{r+s} e^{-\lambda t} t^{r+s-1}}{\Gamma(r)\Gamma(s)} \int_0^1 x^{r-1} (1-x)^{s-1}\, dx && \triangleright \text{ Let } x = u/t, x \in [0,1]
\end{aligned}$$

Now we know that the integrand part of the above equation is beta function,

$$\text{Beta}(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\, dx = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

put it back and cancel gamma parts, we have

$$f_{U+V}(t) = \frac{\lambda^{r+s}}{\Gamma(r+s)} t^{r+s-1} e^{-\lambda t}$$

2 Chi square distribution

Definition The pdf of $U = \sum_{i=1}^m Z_i^2$, where $Z_1, Z_2, \dots, Z_m$ are independent *standard normal* random variables, is called the chi square distribution with m *degrees of freedom*.

It turns out that chi square distribution is nothing but a special case of gamma distribution.

pdf of chi square Let $U = \sum_{i=1}^m Z_i^2$, where $Z_1, Z_2, \dots, Z_m$ are independent standard normal random variables, then U has a gamma distribution with $r = m/2$ and $\lambda = 1/2$ so that

$$f_U(u) = \frac{1}{2^{m/2}\Gamma(m/2)} u^{(m/2)-1} e^{-u/2}, \qquad u \geqslant 0 \tag{2.1}$$

Proof Let's first study the case when $m = 1$ so $U = Z^2$. We find cdf of the Z^2 and then differentiate it to get the pdf of it.

$$F_{Z^2}(u) = P(Z^2 \leqslant u) = P(-\sqrt{u} \leqslant Z \leqslant \sqrt{u}) = 2 \cdot P(Z \leqslant \sqrt{u})$$

Since Z is standard normal, we have Z cdf for less than $\sqrt{u}$,

$$F_{Z^2}(u) = 2 \cdot P(Z \leqslant \sqrt{u}) = 2 \cdot \frac{1}{\sqrt{2\pi}} \int_0^{\sqrt{u}} e^{-z^2/2}\, dz$$

If we differentiate both sides wrt u,

$$f_{Z^2}(u) = \frac{2}{\sqrt{2\pi}} \frac{d}{du} \int_0^{\sqrt{u}} e^{-z^2/2}\, dz = \frac{2}{\sqrt{2\pi}} \frac{d\sqrt{u}}{du} e^{-\sqrt{u}^2/2} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{u}} e^{-u/2}$$

Use the knowledge of $\Gamma(1/2) = \sqrt{\pi}$, we can re-write

$$f_{Z^2}(u) = \frac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} u^{\frac{1}{2}-1} e^{-\frac{u}{2}}$$

The above is a gamma distribution with $r = \lambda = \frac{1}{2}$.

Once we have the pdf for single Z^2, we know the pdf of the sum of m-tuple Z^2 is also a gamma distribution, with $r = m/2$ and the same $\lambda = 1/2$, according to the additive property of gamma distribution:

$$f_U(u) = \frac{(1/2)^{m/2}}{\Gamma(m/2)} u^{(m/2)-1} e^{-u/2}$$

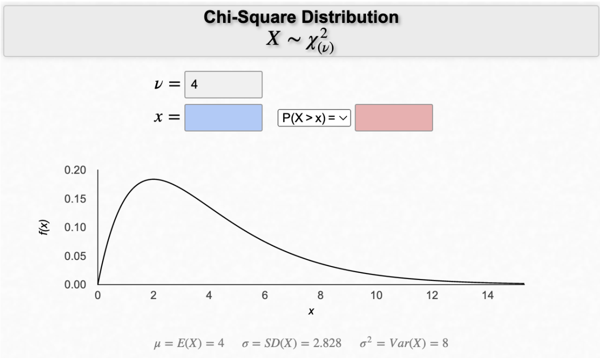

Figure 1: chi square with $m = 4$

3 S^2 is chi square

Okay, we are now one step closer. Let's turn to study the t^2 ratio:

$$t^2 = \frac{(\overline{Y} - \mu)^2}{S^2/n}$$

theorem Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2, then

a. S^2 and $\overline{Y}$ are *independent*

b. the following term has a *chi square* distribution with $n-1$ degrees of freedom:

$$(n-1)S^2/\sigma^2 = \sum_{i=1}^{n} (Y_i - \overline{Y})^2/\sigma^2 \tag{3.1}$$

Proof We standardize Y_i by setting $X_i = (Y_i - \mu)/\sigma$. Let A be a n square orthogonal matrix whose lase row is $(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$. If we define vector $\vec{X} = [X_1, X_2, \ldots, X_n]$, we can get another vector $\vec{Z} = A\vec{X}$. Note the last component of $\vec{Z}$, $Z_n = \frac{1}{\sqrt{n}}(X_1 + X_2 + \cdots + X_n) = \sqrt{n}\overline{X}$.

By the property of orthogonal matrix (see appendix), we observe by intuition:

(1) X_i is mapping to Z_i in a new orthonormal coordinates, the shape of $\vec{X}$ doesn't change, it's the same as $\vec{Z}$.

(2) From (1), we know the same shape of vector $\vec{X}$ is popping up in A orthonormal coordinate, and it's projection on the new coordinate is just the components of vector $\vec{Z}$. As $\vec{X}$'s component X_i is standard normal, so is $\vec{Z}$'s.

Here below we give rigorous proof.

Suppose $\vec{Z}$ belongs to a set D, we have

$$P(\vec{Z} \in D) = P(A\vec{X} \in D) = P(\vec{X} \in A^{-1}D) \qquad \triangleright\ A^{-1}A\vec{X}=\vec{X}$$

If we integrate probability of $f_X(x)$ over set $A^{-1}D$, we get

$$P(\vec{X} \in A^{-1}D) = \int_{A^{-1}D} f_X(x_1) \cdots f_X(x_n)\, dx_1 \cdots dx_n$$

Jacobian Determinant:

Now we want to integrate z_i over D instead of x_i over $A^{-1}D$, we need Jacobian determinant (see appendix) for dz, i.e.,

$$\det(J) = \left| \frac{\partial(x_1, \ldots x_n)}{\partial(z_1, \ldots, z_n)} \right|$$

However, we cannot get this determinant directly from Jacobian matrix for z. But we notice the following relationship between dx_i and dz_i:

$$dx_1 \cdots dx_n = \left| \frac{\partial(x_1, \ldots x_n)}{\partial(z_1, \ldots, z_n)} \right| dz_1 \cdots dz_n$$

and

$$\left| \frac{\partial(z_1, \ldots z_n)}{\partial(x_1, \ldots, x_n)} \right| dx_1 \cdots dx_n = dz_1 \cdots dz_n$$

so we have

$$\left| \frac{\partial(x_1, \ldots x_n)}{\partial(z_1, \ldots, z_n)} \right| = \left| \frac{\partial(z_1, \ldots z_n)}{\partial(x_1, \ldots, x_n)} \right|^{-1}$$

It's obvious to find $\frac{\partial z_i}{\partial x_i}$. From $\vec{Z} = A\vec{X}$, we have

$$z_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n$$

Differentiating z_i wrt $x_1, x_2, \ldots, x_n$, since x is independent variable, we have i^{th} row of the Jacobian matrix as,

$$[a_{i1}, a_{i2}, \ldots, a_{in}]$$

We can conclude that the Jacobian matrix is the same as orthogonal matrix A:

$$\det(J) = \left| \frac{\partial(z_1, \ldots z_n)}{\partial(x_1, \ldots, x_n)} \right| = \det(A) = 1$$

then
$$\left|\frac{\partial(x_1,\dots x_n)}{\partial(z_1,\dots,z_n)}\right| = \left|\frac{\partial(z_1,\dots z_n)}{\partial(x_1,\dots,x_n)}\right|^{-1} = 1$$

Now we continue working on the integrand wrt variable x,

$$\begin{aligned} P(\vec{Z}\in D) = P(\vec{X}\in A^{-1}D) &= \int_{A^{-1}D} f_X(x_1)\cdots f_X(x_n)\,dx_1\cdots dx_n \\ &= \int_D f_X(x_1)\cdots f_X(x_n)\left|\frac{\partial(x_1,\dots x_n)}{\partial(z_1,\dots,z_n)}\right| dz_1\cdots dz_n \\ &= \int_D f_X(g(z_1))\cdots f_X(g(z_n))\cdot 1\cdot dz_1\cdots dz_n \end{aligned}$$

where $g(z_i)$ is mapping z back to x. Because $\vec{X} = A^{-1}\vec{Z}$, we know function g is row multiplication of inverse matrix A^{-1}.

Now as f_X is the pdf of standard normal distribution, we have the above equation further developed as

$$P(\vec{Z}\in D) = \int_D (2\pi)^{-n/2} e^{-(1/2)(x_1^2+x_2^2+\cdots+x_n^2)}\,dz_1\cdots dz_n$$

But we already know that $\left|\vec{X}\right| = \left|\vec{Z}\right|$, so

$$\begin{aligned} x_1^2+x_2^2+\cdots+x_n^2 &= z_1^2+z_2^2+\cdots+z_n^2 \\ \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n Z_i^2 \end{aligned}$$

then
$$P(\vec{Z}\in D) = \int_D (2\pi)^{-n/2} e^{-(1/2)(z_1^2+z_2^2+\cdots+z_n^2)}\,dz_1\cdots dz_n$$

implying that Z_i is also `independent standard normal`.

As the last component of vector $\vec{Z}$ is $\sqrt{n}\overline{X}$, so

$$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^{n-1} Z_i^2 + n\overline{X}^2$$

Now let's see what we can get from X^2,

$$\begin{aligned} \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n X_i^2 - 2n\overline{X}^2 + 2n\overline{X}^2 \\ &= \sum_{i=1}^n (X_i-\overline{X})^2 + n\overline{X}^2 \end{aligned}$$

We have two conclusions from above equations:

(1) $\sum_{i=1}^n (X_i-\overline{X})^2 = \sum_{i=1}^{n-1} Z_i^2$;

(2) As Z_i is independent distribution, so $Z_n = n\overline{X}^2$ is independent to $\sum_{i=1}^{n-1} Z_i^2$, so we know $n\overline{X}^2$, and of course $\overline{X}$, are also independent to $\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Recalling at the beginning, we let $X_i = (Y_i - \mu)/\sigma$, so

$$\begin{aligned} \sum_{i=1}^{n} X_i &= \frac{1}{\sigma}\sum_{i=1}^{n}(Y_i - \mu) \\ n\overline{X} &= \frac{1}{\sigma}(n\overline{Y} - n\mu) \\ \sigma\overline{X} + \mu &= \overline{Y} \end{aligned}$$

and

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(\sigma X_i + \mu - \sigma\overline{X} - \mu)^2 = \sigma^2\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Coming all this long way, we finally arrive at the result:

(1) Just as $\overline{X}$ and $\sum_{i=1}^{n}(X_i - \overline{X})^2$ are independent, $\overline{Y}$ and $\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ are also independent.

(2) $\frac{n-1}{\sigma^2}S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n-1} Z_i^2$, so it's a chi square distribution with $n-1$ degrees.

4 F distribution

We will see the square of t ratio is a special case of F distribution. But first of all, what's F distribution?

Definition Suppose that U and V are independent chi square variables with n and m degrees of freedom. Let $W = \frac{V/m}{U/n}$, we call the distribution of W an F distribution with m and n degrees of freedom.

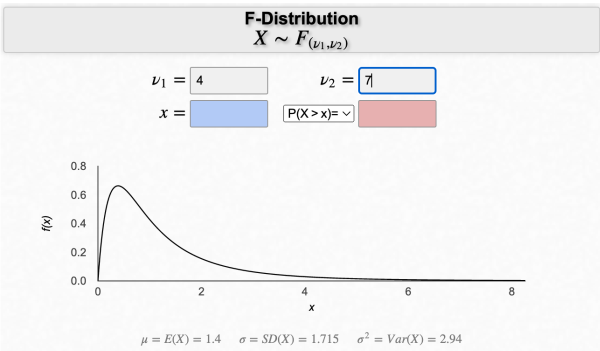

Figure 2: F distribution with $m = 4, n = 7$

F distribution Suppose $F_{m,n}(W) = \dfrac{V/m}{U/n}$ denotes an F variable with m, n degrees of freedom. The pdf of $F_{m,n}$ has the form

$$f_{F_{m,n}}(w) = \frac{\Gamma(\frac{m+n}{2}) m^{m/2} n^{n/2} w^{(m/2)-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})(n+mw)^{(m+n)/2}}, \qquad w \geqslant 0 \tag{4.1}$$

Proof We begin by finding the pdf for V/U before they are divided by m, n respectively. From Eq. 2.1, we have

$$f_V(v) = \frac{1}{2^{m/2}\Gamma(m/2)} v^{(m/2)-1} e^{-v/2} \qquad \text{and} \qquad f_U(u) = \frac{1}{2^{n/2}\Gamma(n/2)} u^{(n/2)-1} e^{-u/2}$$

Now we let $W' = V/U$. Here below is a little lemma for quotient pdf.

lemma Let $w = y/x$, the pdf of w is

$$f(w) = \int x f(x) f(wx) dx$$

So we have pdf of W'

$$\begin{aligned}
f_{V/U}(w') &= \int_0^\infty u f_U(u) f_V(uw')\, du \\
&= \int_0^\infty u \frac{1}{2^{n/2}\Gamma(n/2)} u^{(n/2)-1} e^{-u/2} \frac{1}{2^{m/2}\Gamma(m/2)} (uw')^{(m/2)-1} e^{-uw'/2}\, du \\
&= \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} w'^{(m/2)-1} \int_0^\infty u^{\frac{n+m}{2}-1} e^{-[(1+w')/2]u}\, du
\end{aligned}$$

Now we let $y = \dfrac{1+w'}{2} u$, the integrand above can be written as

$$\begin{aligned}
\int_0^\infty u^{\frac{n+m}{2}-1} e^{-[(1+w')/2]u}\, du &= \int_0^\infty \left(\frac{2y}{1+w'}\right)^{\frac{n+m}{2}-1} e^{-y} \cdot \frac{2}{1+w'}\, dy \\
&= \left(\frac{2}{1+w'}\right)^{\frac{n+m}{2}} \int_0^\infty y^{\frac{n+m}{2}-1} e^{-y}\, dy \\
&= \left(\frac{2}{1+w'}\right)^{\frac{n+m}{2}} \Gamma\left(\frac{n+m}{2}\right) && \triangleright \text{ integrand is a gamma func.}
\end{aligned}$$

then

$$\begin{aligned}
f_{V/U}(w') &= \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} w'^{(m/2)-1} \left(\frac{2}{1+w'}\right)^{\frac{n+m}{2}} \Gamma\left(\frac{n+m}{2}\right) \\
&= \frac{\Gamma(\frac{n+m}{2})}{\Gamma(n/2)\Gamma(m/2)} \cdot \frac{w'^{(m/2)-1}}{(1+w')^{\frac{n+m}{2}}}
\end{aligned}$$

Comment since $\dfrac{1}{\text{Beta}(\frac{n}{2},\frac{m}{2})} = \dfrac{\Gamma(\frac{n+m}{2})}{\Gamma(n/2)\Gamma(m/2)}$, we can also write

$$f_{V/U}(w') = \frac{1}{\text{Beta}(\frac{n}{2},\frac{m}{2})} \cdot \frac{w'^{(m/2)-1}}{(1+w')^{\frac{n+m}{2}}}$$

As $W = \dfrac{V/m}{U/n}$ and $W' = \dfrac{V}{U}$, we have $W = W'\dfrac{n}{m}$. To use pdf for W' to represent W's, we need another lemma first.

lemma Let $y = ax$, and pdf for x is $f_X(x)$, then $f_Y(y)$ is $\dfrac{1}{|a|} f_X\left(\dfrac{y}{x}\right)$.

Proof If we integrate x to get the cdf of y, it's like this:

$$F(y) = \int_0^{ax} f_X(x)\,dx$$

Notice that $dy = a\,dx$, then differentiate both sides to get the pdf for y,

$$F(y)' = f_Y(y) = \left[\int_0^y f_X\left(\frac{y}{a}\right)\frac{1}{a}\,dy\right]' = \frac{1}{|a|} f_X\left(\frac{y}{x}\right)$$

Follow this route, we know that,

$$\begin{aligned} f(w) &= \frac{m}{n} f_{V/U}(W \cdot \frac{m}{n}) \\ &= \frac{m}{n} \frac{1}{\text{Beta}(\frac{n}{2},\frac{m}{2})} \cdot \frac{(wm/n)^{(m/2)-1}}{(1+wm/n)^{(n+m)/2}} \\ &= \frac{1}{\text{Beta}(\frac{n}{2},\frac{m}{2})} \cdot \frac{m^{m/2} n^{n/2} w^{(m/2)-1}}{(n+mw)^{(m+n)/2}} \end{aligned}$$

thus proven.

5 F to t ratios

Our goal is to find the pdf of $t = \dfrac{\overline{Y}-\mu}{S/\sqrt{n}}$. It turns out, however, that t just belongs to a family of quotients known as t ratios. By finding the pdf for that family, we can easily find the pdf of t as well.

Definition t ratios family: Let Z be a standard normal, and let U_n be a chi square random variable with n degrees of freedom and independent of Z. The `student t ratio` with n degrees of freedom is denoted T_n,

$$T_n = \frac{Z}{\sqrt{U_n/n}}$$

lemma $f_{T_n}(t)$ is `symmetric` for all t. From this symmetrical property, and based on F distribution, we have the t ratio pdf below.

pdf of t ratio with n df

$$f_{T_n}(t) = \frac{1}{\sqrt{n}\text{Beta}\left(\frac{1}{2}, \frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \tag{5.1}$$

Proof We can suppose $t > 0$. Due to the symmetry of the pdf, we can find $P(t < 0)$ by $1 - P(t > 0)$.

$$\begin{aligned} F_{T_n}(t) &= \frac{1}{2} + P(0 \leqslant T_n \leqslant t) && \triangleright \text{ symmetrical on 0} \\ &= \frac{1}{2} + \frac{1}{2}P(-t \leqslant T_n \leqslant t) \\ &= \frac{1}{2} + \frac{1}{2}P(0 \leqslant T_n^2 \leqslant t^2) \\ &= \frac{1}{2} + \frac{1}{2}F_{T_n^2}(t^2) && \triangleright \text{ when F dist. can kick in} \end{aligned}$$

then we can differentiate both sides to get the pdf for t,

$$f_{T_n}(t) = F'_{T_n}(t) = \frac{1}{2} \cdot 2t \cdot f_{T_n^2}(t^2) = t \cdot f_{T_n^2}(t^2)$$

But as Z^2 is divided by none, and U is divided by n, we know T_n^2 has a F-distribution pdf with $m = 1, n = n$ degrees of freedom,

$$T_n^2 = \frac{Z^2}{U/n}$$

so

$$f_{T_n^2}(t^2) = \frac{1}{\text{Beta}(\frac{n}{2}, \frac{1}{2})} \cdot \frac{n^{n/2}(t^2)^{-\frac{1}{2}}}{(n + t^2)^{(1+n)/2}}$$

and

$$\begin{aligned} f_{T_n}(t) = t \cdot f_{T_n^2}(t^2) &= t \cdot \frac{1}{\text{Beta}(\frac{n}{2}, \frac{1}{2})} \cdot \frac{n^{n/2}(t^2)^{-\frac{1}{2}}}{(n + t^2)^{(1+n)/2}} \\ &= \frac{1}{\text{Beta}(\frac{n}{2}, \frac{1}{2})} \cdot \frac{n^{n/2}}{(n + t^2)^{(1+n)/2}} \\ &= \frac{1}{\sqrt{n}\text{Beta}(\frac{n}{2}, \frac{1}{2})} \cdot \frac{n^{(1+n)/2}}{(n + t^2)^{(1+n)/2}} \\ &= \frac{1}{\sqrt{n}\text{Beta}(\frac{n}{2}, \frac{1}{2})} \cdot \left(1 + \frac{t^2}{n}\right)^{-(1+n)/2} \end{aligned}$$

6 final step

We have only one missing piece to the whole puzzle: how to verify our ratio is in fact a student t ratio so that the above pdf can apply. Let's give the conclusion first then the proof process.

theorem Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2, then

$$T_{n-1} = \frac{\overline{Y} - \mu}{S/\sqrt{n}}$$

has a student t distribution with $(n-1)$ degrees of freedom.

Proof In order to prove that T_{n-1} is a student t distribution, we need to prove that $\overline{Y} - \mu$ is a standard normal, and the square of denominator S^2/n is a chi square with some degrees of freedom.

Let's divide both numerator and denominator by $\sigma/\sqrt{n}$, then:

(1) for the numerator, it becomes $\frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}$. We know this is a standard normal by CLT.

(2) for the denominator, we have,

$$\frac{S/\sqrt{n}}{\sigma/\sqrt{n}} = \frac{S}{\sigma} = \sqrt{\frac{S^2}{\sigma^2}}$$

We know that $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$, so

$$\sqrt{\frac{S^2}{\sigma^2}} = \sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sigma^2(n-1)}}$$

By Eq. 3.1, it's evident that $\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sigma^2}$ is a chi square with $n-1$ degrees of freedom. By dividing $(n-1)$ again, the whole fraction is exactly a form of student t ratio with $n-1$ degrees.

Take a second look at student t distribution! As degree $n \to \infty$, S is asymptotically close to σ. As a result, the distribution of T_{n-1} is more and more like a standard normal. The following graph show this point [1].

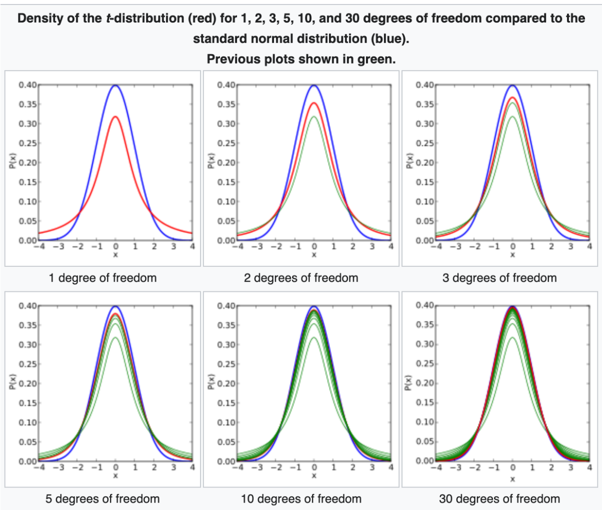

Figure 3: Compare of t with normal

[1] Graph from Wikipedia.

0 Appendix

0.1 orthogonal matrix

Definition A square matrix P is called orthogonal if it is invertible and if

$$P^{-1} = P^T$$

From this definition, we can deduce some interesting properties of the orthogonal matrix.

theorem A square n matrix P is orthogonal if and only if its column vectors for an orthonormal set.

Proof let $\mathbf{p}_i$ stands for column i of matrix P, so $P = [\mathbf{p}_1 \mathbf{p}_2 \cdots \mathbf{p}_n]$. The product of $P^T P$ has the form

$$P^T P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{n1} \\ p_{12} & p_{22} & \cdots & p_{n2} \\ \vdots & \vdots & & \vdots \\ p_{1n} & p_{2n} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{p}_1 \cdot \mathbf{p}_1 & \mathbf{p}_1 \cdot \mathbf{p}_2 & \cdots & \mathbf{p}_1 \cdot \mathbf{p}_n \\ \mathbf{p}_2 \cdot \mathbf{p}_1 & \mathbf{p}_2 \cdot \mathbf{p}_2 & \cdots & \mathbf{p}_2 \cdot \mathbf{p}_n \\ \vdots & \vdots & & \vdots \\ \mathbf{p}_n \cdot \mathbf{p}_1 & \mathbf{p}_n \cdot \mathbf{p}_2 & \cdots & \mathbf{p}_n \cdot \mathbf{p}_n \end{bmatrix}$$

We now suppose the column vectors of P form an orthonormal set, so $\mathbf{p}_i \cdot \mathbf{p}_i = 1$, elsewhere is 0, and $P^T P = I_n$.

But as $P^{-1} P = I$, we conclude that $P^T = P^{-1}$, thus prove orthonormal leads to orthogonal.

theorem Matrix P of square n is orthogonal if and only if for any vector of length n, $\vec{\beta}$

$$\left|P\vec{\beta}\right| = \left|\vec{\beta}\right|$$

Proof we denote $\vec{\beta} = [b_1, b_2, \ldots, b_n]$, so

$$P\vec{\beta} = [b_1 \mathbf{p}_1 + b_2 \mathbf{p}_2 + \cdots b_n \mathbf{p}_n]$$

Notice that we are now in a new space with $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ as its new coordinates. What's b_i? they are nothing but components' scalar of old coordinates! We use these scalars to construct components in new coordinates, and $b_1 \mathbf{p}_1, b_2 \mathbf{p}_2, \ldots, b_n \mathbf{p}_n$ are new components, squaring of which is the original length,

$$\left|P\vec{\beta}\right|^2 = (b_1 \mathbf{p}_1)^2 + (b_2 \mathbf{p}_2)^2 + \cdots (b_n \mathbf{p}_n)^2 = b_1^2 + b_2^2 + \cdots + b_n^2 = \left|\vec{\beta}\right|^2$$

We cancelled out $\mathbf{p}_i^2$ because orthogonal means orthonormal, each column vector length is 1.

theorem If matrix P is orthogonal, then $\det(P) = 1$.

Proof we have three lemmas to verify this theorem.

lemma 1: $\det(AB) = \det(A)\det(B) \longrightarrow 1 = \det(P)\det(P^{-1})$

lemma 2: $\det(P) = \det\left(P^T\right)$

lemma 3: $\det(P^{-1}) = \det\left(P^T\right)$

0.2 Jacobian transformation

We are familiar with certain coordinate changes like this one:

$$\iint_D f(x,y)\,dx\,dy = \iint_{D_0} f(r\cos\theta, r\sin\theta) r\,dr\,d\theta$$

In fact, the above is just a special case of Jacobian transformation. If we define $x = f_x(u,v)$ and $y = f_y(u,v)$, a square from uv coordinates to x, y plane may look like this,

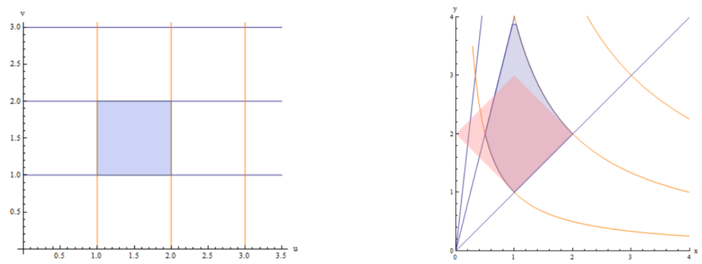

Figure 4: change from uv to xy coordinates

The deformed shape is close to a parallelogram. From the lower-most point, we want to find two vectors (the two sides of the parallelogram), and use them to calculate the area of the shape. You may notice it's not exactly a standard parallelogram, but when we look into very tiny piece of shapes, the approximation would be quite nice.

Assume the lowest point is (u_0, v_0), and $\vec{a}$ for up-right direction, $\vec{b}$ for up-left direction.

$$\begin{aligned}
\vec{a} &= [f_x(u_0 + \Delta u, v_0) - f_x(u_0, v_0), f_y(u_0 + \Delta u, v_0) - f_y(u_0, v_0)] \\
&\doteq \left[\Delta u \frac{\partial x}{\partial u}, \Delta u \frac{\partial y}{\partial u}\right] \\
\vec{b} &= [f_x(u_0, v_0 + \Delta v) - f_x(u_0, v_0), f_y(u_0, v_0 + \Delta v) - f_y(u_0, v_0)] \\
&\doteq \left[\Delta v \frac{\partial x}{\partial v}, \Delta v \frac{\partial y}{\partial v}\right]
\end{aligned}$$

How to calculate the area of parallelogram from vectors $\vec{a}, \vec{b}$? Some algebraic work will yield answer: the parallelogram area is the absolute value of the determinant of the following matrix,

$$\begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$$

$$\textbf{Jacobian determinant} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

As a convention, we often use $\left|\frac{\partial x, y}{\partial u, v}\right|$ to denote the Jacobian determinant. Let $\Delta A = dx\,dy$ denote every piece of shape in xy plane, then

$$\Delta A \doteq \left|\frac{\partial(x, y)}{\partial(u, v)}\right| \Delta u \Delta v$$

In the end, we can integrate the shape in uv coordinates

$$\iint_D f(x, y)\,dx\,dy = \iint_{D_0} f(f_x(u, v), f_y(u, v)) \left|\frac{\partial(x, y)}{\partial(u, v)}\right| du\,dv$$

Comment be aware of the region D_0, it may often change also. To get the correct integral, you need to make sure the distorted region cope with the relationship between new and old coordinates.

Jacobian determinant rule also apply to 3D or even higher dimensions. Again we pick a tiny piece of block from xyz space and see what distorted shape it is in the new uvw coordinates.

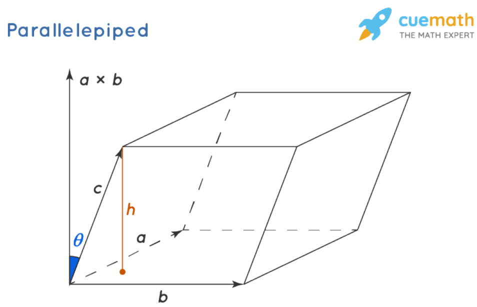

Figure 5: 3D shape distorted

Following the same manner, we need to find three vectors $\vec{a}, \vec{b}$ and $\vec{c}$, and we have

$$\Delta V = dx\,dy\,dz = a \cdot (b \times c) = \det \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} & \frac{\partial z}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} & \frac{\partial z}{\partial v} \\ \frac{\partial x}{\partial w} & \frac{\partial y}{\partial w} & \frac{\partial z}{\partial w} \end{bmatrix} \Delta u \Delta v \Delta w$$

so

$$\iiint_D f(x, y, z)\,dx\,dy\,dz = \iiint_{D_0} f(f_x, f_y, f_z) \left|\frac{\partial(x, y, z)}{\partial(u, v, w)}\right| du\,dv\,dw$$