Lecture Notes on Student's t-distribution and F distribution.

1. Student's t-distribution

(1). The motivation

Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Then $\overline{X} \sim N(\mu, \sigma^2)$. If σ^2 is known, then $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, so that statistical inference on μ can be made. If σ^2 is unknown, we may use S^2 to approximate σ^2, but then we need to determine the distribution of $\frac{\overline{X}-\mu}{S/\sqrt{n}}$.

Notice that $\frac{\overline{X}-\mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/n-1}}$, where $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

We can consider the random variable $Z/\sqrt{U/r}$, where $Z \sim N(0,1)$, $U \sim \chi^2(r)$.

(2). The definition of $t(r)$

Def. Let $Z \sim N(0,1)$, $U \sim \chi^2(r)$, (independent!) and $T = Z/\sqrt{U/r}$ $U \perp T$. The distribution of T is called the Student's t-distribution of degree of freedom r.

(3). The distribution of $t(r)$.

Thm. The probability density function of T is

$$f(x) = \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})\sqrt{r\pi}} \left(1+\frac{x^2}{r}\right)^{-\frac{r+1}{2}}, \quad -\infty < x < \infty.$$

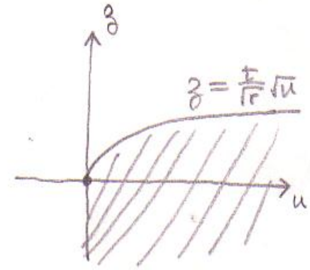
If $r>1$, $E[T]=0$. If $r>2$, $var(T) = \frac{r}{r-2}$.

Note. No $M(t)$ exists.

Proof. The joint p.d.f. of (Z, U) is given by

$$g(z,u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \cdot \frac{1}{\Gamma(\frac{r}{2}) 2^{r/2}} u^{\frac{r}{2}-1} e^{-u/2}, \quad -\infty < z < \infty,\ 0 < u < \infty.$$

Let $F(x) = P[T \le x]$. Then



$$F(x) = P\left[\frac{Z}{\sqrt{U/r}} \le x\right] = P\left[Z \le \frac{x}{\sqrt{r}}\sqrt{U}\right]$$

$$= \int_0^\infty \int_{-\infty}^{\frac{x}{\sqrt{r}}\sqrt{u}} g(z,u)\,dz\,du$$

$$= \int_0^\infty \int_{-\infty}^{\frac{x}{\sqrt{r}}\sqrt{u}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{1}{\Gamma(\frac{r}{2})\, 2^{r/2}} u^{\frac{r}{2}-1} e^{-\frac{u}{2}}\,dz\,du$$

$$= \frac{1}{\sqrt{\pi}\,\Gamma(\frac{r}{2})} \int_0^\infty \left[\int_{-\infty}^{\sqrt{\frac{u}{r}}x} e^{-\frac{z^2}{2}} dz\right] \frac{1}{2^{\frac{r+1}{2}}} u^{\frac{r}{2}-1} e^{-\frac{u}{2}} du$$

$$f(x) = F'(x) = \frac{1}{\sqrt{\pi}\,\Gamma(\frac{r}{2})} \int_0^\infty \left[e^{-\frac{\left(\sqrt{\frac{u}{r}}x\right)^2}{2}} \cdot \sqrt{\frac{u}{r}}\right] \cdot \frac{1}{2^{\frac{r+1}{2}}} u^{\frac{r}{2}-1} e^{-\frac{u}{2}} du$$

$$= \frac{1}{\sqrt{\pi}\,\Gamma(\frac{r}{2})} \int_0^\infty \frac{1}{\sqrt{r}\, 2^{\frac{r+1}{2}}} u^{\frac{r+1}{2}-1} e^{-(1+\frac{x^2}{r})\frac{u}{2}} du$$

$$= \frac{1}{\sqrt{r\pi}\,\Gamma(\frac{r}{2})} \int_0^\infty \frac{1}{2^{\frac{r+1}{2}}} u^{\frac{r+1}{2}-1} e^{-(1+\frac{x^2}{r})\frac{u}{2}} du$$

$$\left[\text{Let } y = (1+\tfrac{x^2}{r})u, \quad dy = (1+\tfrac{x^2}{r})du\right]$$

$$= \frac{1}{\Gamma(\frac{r}{2})\sqrt{r\pi}} \int_0^\infty \frac{1}{2^{\frac{r+1}{2}}} \left[\frac{y}{1+\frac{x^2}{r}}\right]^{\frac{r+1}{2}-1} e^{-\frac{y}{2}} \frac{dy}{1+\frac{x^2}{r}}$$

$$= \frac{1}{\Gamma(\frac{r}{2})\sqrt{r\pi}} \left(1+\frac{x^2}{r}\right)^{-\frac{r+1}{2}} \cdot \Gamma\left(\frac{r+1}{2}\right) \int_0^\infty \frac{1}{\Gamma(\frac{r+1}{2})\, 2^{\frac{r+1}{2}}} y^{\frac{r+1}{2}-1} e^{-\frac{y}{2}} dy$$

$$= \frac{1}{\Gamma(\frac{r}{2})\sqrt{r\pi}} \left(1+\frac{x^2}{r}\right)^{-\frac{r+1}{2}} \Gamma\left(\frac{r+1}{2}\right) = \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})\sqrt{r\pi}} \left(1+\frac{x^2}{r}\right)^{-\frac{r+1}{2}}.$$

<u>Note</u>. It follows that $\lim_{r\to\infty} T = Z \sim N(0,1)$.

2. The F distribution.

(1) The definition of $F(r_1, r_2)$

To compare the variances of two normal distribution $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, we consider the quotient $W = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \dfrac{\frac{[n_1-1]S_1^2}{\sigma_1^2} \Big/ (n_1-1)}{\frac{[n_2-1]S_2^2}{\sigma_2^2} \Big/ (n_2-1)}$

Definition. Let $U \sim \chi^2(r_1)$, and $V \sim \chi^2(r_2)$, $U \perp V$. The distribution of $W = \dfrac{U/r_1}{V/r_2}$ is called the F distribution with r_1 and r_2 degree of freedom, denoted by $F(r_1, r_2)$.

(2). The distribution of $F(r_1, r_2)$.

Thm. The probability density function $f(x)$ of $F(r_1, r_2)$ is given by

$$f(x) = \frac{\Gamma\left(\frac{r_1+r_2}{2}\right)}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1} \left(1+\frac{r_1}{r_2}x\right)^{-\frac{r_1+r_2}{2}}, \quad x > 0.$$

If $r_2 > 2$, $E(W) = \dfrac{r_2}{r_2-2}$. If $r_2 > 4$, $Var(W) = \dfrac{2r_2^2(r_1+r_2-2)}{r_1(r_2-2)^2(r_2-4)}$.

Proof. The joint density function of (U, V) is

$$g(u,v) = \frac{u^{\frac{r_1}{2}-1} e^{-u/2}}{\Gamma\left(\frac{r_1}{2}\right) 2^{r_1/2}} \frac{v^{\frac{r_2}{2}-1} e^{-\frac{v}{2}}}{\Gamma\left(\frac{r_2}{2}\right) 2^{r_2/2}}, \quad 0 < u, v < \infty.$$

For $x > 0$, $F(x) = P[W \le x] = P\left[\dfrac{U/r_1}{V/r_2} \le x\right] = P\left[U \le \dfrac{r_1}{r_2} V x\right]$

$$= \int_0^\infty \int_0^{\frac{r_1}{r_2}vx} g(u,v)\,du\,dv$$

$$= \frac{1}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \int_0^\infty \left[\int_0^{\frac{r_1}{r_2}vx} u^{\frac{r_1}{2}-1} e^{-\frac{u}{2}} du\right] \frac{1}{2^{\frac{r_1+r_2}{2}}} v^{\frac{r_2}{2}-1} e^{-\frac{v}{2}} dv$$

$$f(x) = F'(x) = \frac{1}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \int_0^\infty \left[\left(\frac{r_1}{r_2}vx\right)^{\frac{r_1}{2}-1} e^{-\frac{\frac{r_1}{r_2}vx}{2}} \cdot \frac{r_1}{r_2} v\right] \frac{1}{2^{\frac{r_1+r_2}{2}}} v^{\frac{r_2}{2}-1} e^{-\frac{v}{2}} dv$$

$$= \frac{1}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})} \int_0^\infty \left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1} \frac{1}{2^{\frac{r_1+r_2}{2}}} v^{\frac{r_1+r_2}{2}-1} e^{-(1+\frac{r_1}{r_2}x)\frac{v}{2}} dv$$

$$= \frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1}}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})} \int_0^\infty \frac{1}{2^{\frac{r_1+r_2}{2}}} v^{\frac{r_1+r_2}{2}-1} e^{-(1+\frac{r_1}{r_2}x)\frac{v}{2}} dv$$

$$\left[\text{Let } y = (1+\tfrac{r_1}{r_2}x)v, \quad dy = (1+\tfrac{r_1}{r_2}x)\,dv\right]$$

$$= \frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1}}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})} \int_0^\infty \frac{1}{2^{\frac{r_1+r_2}{2}}} \left(\frac{y}{1+\frac{r_1}{r_2}x}\right)^{\frac{r_1+r_2}{2}-1} e^{-\frac{y}{2}} \frac{dy}{(1+\frac{r_1}{r_2}x)}$$

$$= \frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1}}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})} \left(1+\frac{r_1}{r_2}x\right)^{-\frac{r_1+r_2}{2}} \Gamma\left(\frac{r_1+r_2}{2}\right) \int_0^\infty \frac{1}{\Gamma(\frac{r_1+r_2}{2}) 2^{\frac{r_1+r_2}{2}}} y^{\frac{r_1+r_2}{2}-1} e^{-\frac{y}{2}} dy$$

$$= \frac{\Gamma(\frac{r_1+r_2}{2})}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})} \left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1} \left(1+\frac{r_1}{r_2}x\right)^{-\frac{r_1+r_2}{2}}, \quad x > 0.$$

Notes.

(1). $W \sim F(r_1, r_2)$ if and only if $\frac{1}{W} \sim F(r_2, r_1)$.

(2). If we use the notation $F_\alpha(r_1, r_2)$ to denote the $(1-\alpha)\%$ percentile, i.e. $P[W > F_\alpha(r_1, r_2)] = \alpha$, then

$F_\alpha(r_2, r_1) = 1/F_{1-\alpha}(r_1, r_2)$.

$F_{1-\alpha}(r_1, r_2) = 1/F_\alpha(r_2, r_1)$.