# Math 3236 Statistical Theory

3/16/23

t distribution (Chapter 8.4)

$X_i$, $i = 1 \dots N$, are i.i.d r.v. with

$\mathbb{E}(X_i) = \mu$ $\quad$ $Var(X_i) = \sigma^2$

$$U = \frac{1}{\sqrt{N}}\left(\sum_i \left(\frac{X_i - \mu}{\sigma}\right)\right)$$

pivotal quantities

$$V = \sum_i (X_i - \bar{X})^2$$

where $\bar{X} = \frac{1}{N} \sum_i X_i$

$U \simeq N(0,1)$

$V \simeq \chi^2_{N-1}$

Moreover $U$ and $V$ are

independent.

$$\bar{X} = \frac{\sigma}{\sqrt{N}} U + \mu$$

$$\bar{X} = N\left(\mu, \frac{\sigma^2}{N}\right)$$

---

What happens if I take same natural combination of $U$, $V$?

Theorem. If $X$ is $N(0,1)$ and $Y$ is $\chi^2_m$ (chi square with $m$ dof) then the r.v

$$T = \frac{X}{\left(\frac{Y}{m}\right)^{1/2}}$$

has p.d.f.

$$f_T(t) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{(m\pi)^{1/2}\,\Gamma\left(\frac{m}{2}\right)}\left(1 + \frac{t^2}{m}\right)^{-(m+1)/2}$$

Moreover $T$ is called a $T$ r.v. with $m$ d.o.f.

---

Proof:

$$X, Y$$

$$T = \frac{X}{\left(\frac{Y}{m}\right)^{1/2}} \qquad W = Y$$

We have the j. p.d.f of $X\,Y$.
We first find the j.p.d.f of $T\,W$ and the integrate out $W$ to find the p.d.f. of $T$.

$$f_{X,Y}(x,y) = f_X(x)\,f_Y(y) =$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma\left(\frac{m}{2}\right)} \frac{1}{2^{m/2}} e^{-\frac{x^2}{2}} \quad y^{\frac{m}{2}} e^{-y/2}$$

$$Y = W \qquad X = T\left(\frac{W}{m}\right)^{1/2}$$

Compute the Jacobian

$$\begin{pmatrix} \frac{\partial y}{\partial w} & \frac{\partial y}{\partial t} \\ \frac{\partial x}{\partial w} & \frac{\partial x}{\partial t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{T}{2}\left(\frac{w}{m}\right)^{-1/2} & \left(\frac{w}{m}\right)^{1/2} \end{pmatrix}$$

$$J(w, t) = \left(\frac{w}{m}\right)^{1/2}$$

$$f_{W,T}(w, t) = c\, w^{(m+1)/2 - 1} \exp\left[-\frac{1}{2}\left(1 + \frac{t^2}{m}\right) w\right]$$

$$c = \left[2^{(m+1)/2} (m\pi)^{1/2} \Gamma\left(\frac{m}{2}\right)\right]^{-1}$$

$$\int_0^\infty f_{W,T}(w,t)\,dw =$$

$$c \int_0^\infty w^{\alpha-1} e^{-\beta w}\,dw$$

$$\alpha = \frac{m+1}{2} \qquad \beta = \frac{1}{2}\left(1 + \frac{t^2}{m}\right)$$

$$\int_0^\infty w^{\alpha-1} e^{-\beta w}\,dw$$

$$u = \beta w$$

$$\beta^{-\alpha} \int_0^\infty u^{\alpha-1} e^{-u}\,du = \Gamma(\alpha)\,\beta^{-\alpha}$$

$$f_W(w) = c_f \left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}}$$

$$c_f = \Gamma\left(\frac{m+1}{2}\right) \Big/ (m\pi)^{1/2}\,\Gamma\left(\frac{m}{2}\right)$$

Main application: Let $X_i$ be i.i.d. normal r.v. with

$\mathbb{E}(X_i) = \mu \qquad Var(X_i) = \sigma^2.$

Define

$$\bar{X} = \frac{1}{N} \sum_i X_i$$

$$S^2 = \sum_i (X_i - \bar{X})^2$$

Then

$$T = \frac{\sqrt{N}(\bar{X} - \mu)}{\left(\frac{S^2}{(n-1)}\right)^2} \sim t_{n-1} \text{ d.o.f.}$$

$$T = \frac{\overbrace{\frac{\sqrt{N}}{\sigma}(\bar{X} - \mu)}^{N(0,1)}}{\underbrace{\left(\frac{S^2}{\sigma^2(n-1)}\right)^{1/2}}_{\chi_{n-1}}} \sim t_{n-1}$$

$$\hat{\sigma}' = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Then $\hat{\sigma}'$ is an estimator for $\sigma$. This is not the MLE but is unbiased.

---

$$f_T(t) = c\left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}}$$

If I want to compute

$\mathbb{E}(T^k)$ is not defined

if $k \geq m$. Thus a $T$ r.v. with $n$ d.o.f. has at most $k-1$ moments defined.

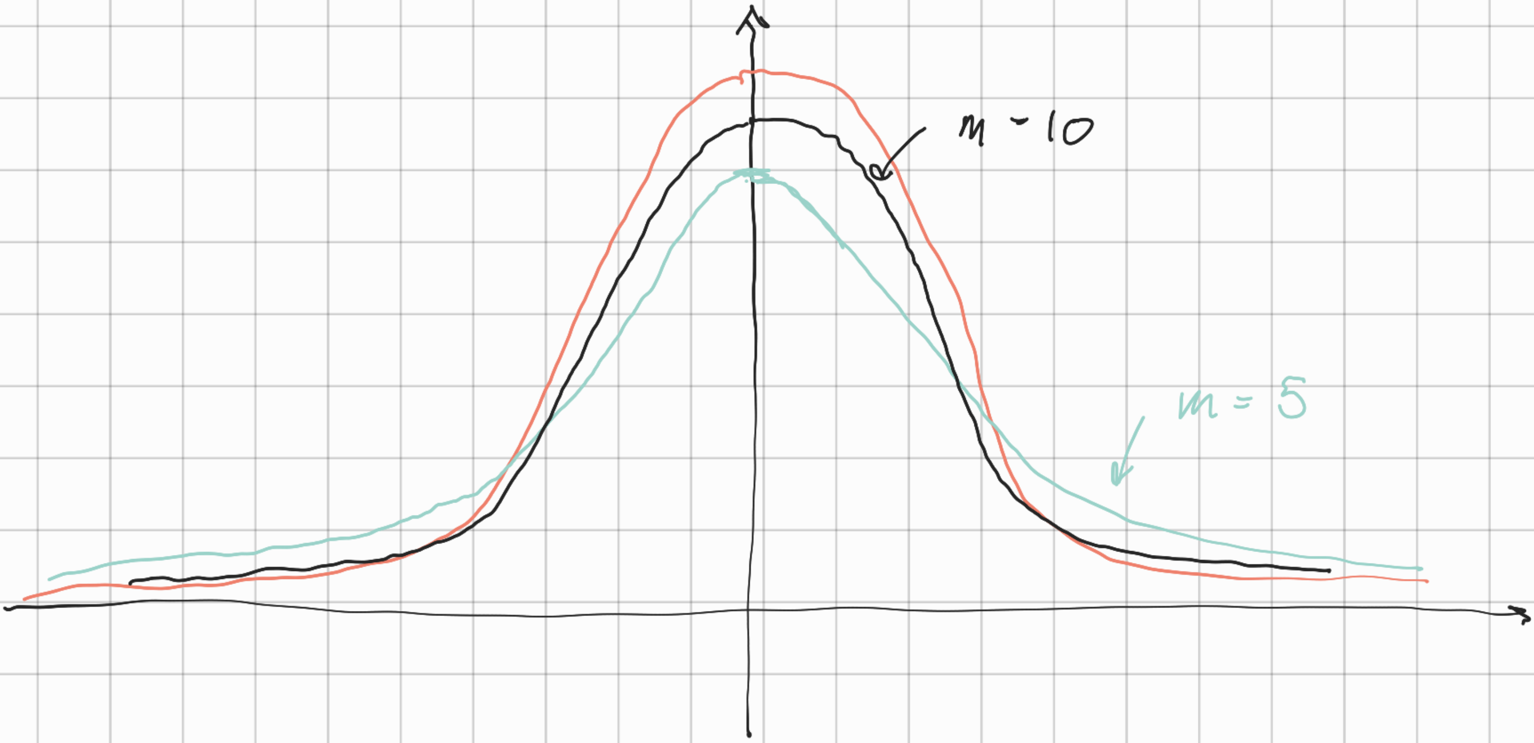
What happens if $m$ is very large

$$\left(1+\frac{x^2}{m}\right)^{-\frac{(m+1)}{2}} \xrightarrow[m\to\infty]{} e^{-\frac{x^2}{2}}$$

$$\frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m}\;\Gamma\left(\frac{m}{2}\right)} \xrightarrow[m\to\infty]{} \frac{1}{\sqrt{2}}$$

$$f_T(t) \xrightarrow[m\to\infty]{} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

$T_m \xrightarrow[m\to\infty]{} Z$ where $Z$ is $\mathcal{N}(0,1)$.




$m \geq 40$ in general $T_m$ and $Z$ are essentially indistinguishable.

For example

$t_{\alpha, m}$ critical value for $t$ distribution with $m$ d.o.f.

$$\mathbb{P}(T \geq T_{\alpha, m}) = \alpha$$

$z_\alpha$ critical value for

Standard Normal distribution

$$\mathbb{P}(Z \geq z_\alpha) = \alpha$$

$t_{m, 0.025}$

| | | |
|---|---|---|
| $m=2$ | $4.303$ | |
| $m=10$ | $2.228$ | |
| $m=20$ | $2.086$ | |
| $m=30$ | $2.042$ | |
| $m=\infty$ | $1.960$ | $z_{0.025}$ |

---

Why we did all This?

$X_i \quad i = 1 \ldots 30$ are i.i.d

$\mathcal{N}(\mu, \sigma^2)$

$x_i \quad i = 1 \ldots 30$

$$\overline{x} = \frac{1}{30}\sum_{i=1}^{30} x_i$$

$$s^2 = \frac{1}{29}\sum_{i=1}^{N}(x_i - \overline{x})^2$$

$$\mathbb{P}\left(|\overline{X} - \mu| \geq \delta\right) \leq \alpha$$

Find $\delta$ such that that inequality is true.

$$\mathbb{P}\left(|\overline{X} - \mu| \geq \delta\right) =$$

$$\mathbb{P}\left(\left|\frac{\sqrt{30}(X - \mu)}{\hat{\sigma}'}\right| \geq \frac{\sqrt{30}}{\sigma'}\delta\right) =$$

$$\hat{\sigma}' = \left(\frac{1}{n-1}\sum_i (X_i - \overline{X})^2\right)^{1/2}$$

$$= \mathbb{P}\left(|T| \geq \frac{\sqrt{30}}{\hat{\sigma}'}\delta\right)$$

$$\mathbb{P}\left(|T| \geq \frac{\sqrt{30}}{\hat{\sigma}'}\delta\right) = \alpha$$

Is equivalent To

$$\mathbb{P}\left(T \geq \frac{\sqrt{30}}{\hat{\sigma}'}\delta\right) = \frac{\alpha}{2}$$

because t-distribution is symmetric

$$\frac{\sqrt{30}}{\hat{\sigma}'}\delta = t_{29,\alpha/2}$$

$$\delta = \frac{\hat{\sigma}'}{\sqrt{30}} t_{29,\alpha/2}$$

$$\mathbb{P}\left(|\bar{X} - \mu| \geq \frac{\hat{\sigma}'}{\sqrt{30}} t_{29,\alpha/2}\right) = \alpha$$

Confidence Interval

With prob $1-\alpha$ I have

$$\bar{x} - \frac{\hat{\sigma}'}{\sqrt{n}} t_{n-1, \alpha/2} \le \mu \le \bar{x} + \frac{\hat{\sigma}'}{\sqrt{n}} t_{n-1, \alpha/2}$$

Confidence interval.