X_n … independent.

But we can derive the dist of $X_i - \bar{X}_n$:

$$X_i - \bar{X}_n = X_i - \frac{1}{n}\sum_{i=1}^{n} X_i = \left(1-\frac{1}{n}\right)X_i - \frac{1}{n}\sum_{j\neq i} X_j \sim N\left(0, \frac{n-1}{n}\sigma^2\right)$$

$$\begin{cases} E[X_i - \bar{X}_n] = E[X_i] - E[\bar{X}_n] = \mu - \mu = 0 \\ var[X_i - \bar{X}_n] = var\left[\left(1-\frac{1}{n}\right)X_i - \frac{1}{n}\sum_{j\neq i} X_j\right] \overset{ind}{=} \left(1-\frac{1}{n}\right)^2 var[X_i] + \left(\frac{1}{n}\right)^2 \sum_{j\neq i} var[X_j] \end{cases}$$

$$= \left(1-\frac{1}{n}\right)^2\sigma^2 + \left(\frac{1}{n}\right)^2(n-1)\sigma^2$$

$$= \left(\frac{(n-1)^2}{n^2} + \frac{n-1}{n^2}\right)\sigma^2$$

$$= \frac{n^2-2n+1+n-1}{n^2}\sigma^2 = \frac{n^2-n}{n^2}\sigma^2$$

$$= \frac{n-1}{n}\sigma^2$$

So $M_{X_i - \bar{X}_n}(t) = e^{\frac{(n-1)\sigma^2 t^2}{n \cdot 2}}$

We are left to show that (*) is equivalent to

$$\mu\left(\frac{s-t}{n}+t\right) + \frac{\sigma^2}{2}\left(\frac{s-t}{n}+t\right)^2 + \mu\left(\frac{s-t}{n}\right)(n-1) + \frac{\sigma^2}{2}\left(\frac{s-t}{n}\right)^2(n-1)$$

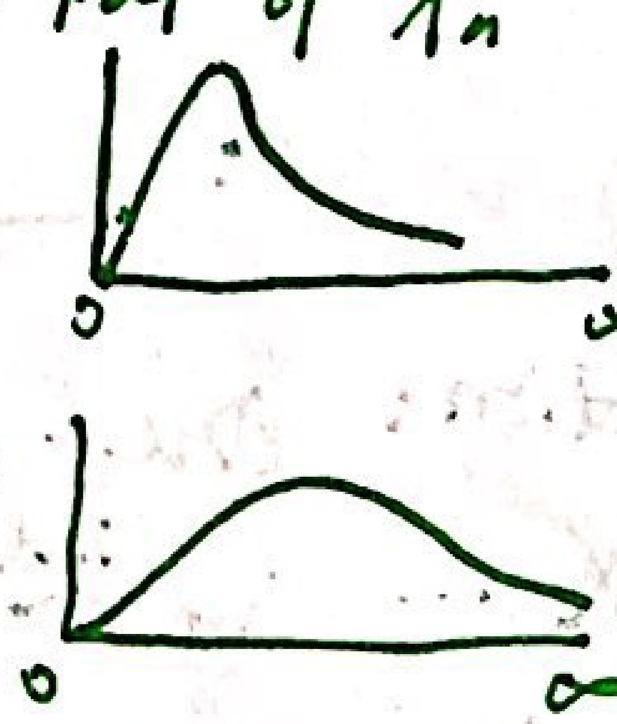
$$\overset{?}{=} \mu s + \frac{\sigma^2 s^2}{2n} + \frac{(n-1)\sigma^2 t^2}{2n}$$

$$\mu\left(\frac{s-t}{n}\right) + \mu t + \mu\left(\frac{s-t}{n}\right)(n-1) = \mu\left(\frac{s-t}{\cancel{n}}\right)\cdot\cancel{n} + \mu t = \mu s$$

$$\frac{\sigma^2}{2}\left(\frac{s-t}{n}+t\right)^2 + \frac{\sigma^2}{2}\left(\frac{s-t}{n}\right)^2(n-1) = \frac{\sigma^2}{2}\left[\left(\frac{s-t}{n}\right)^2 + 2\frac{(s-t)t}{n} + t^2\right] + \frac{\sigma^2}{2}\left(\frac{s-t}{n}\right)^2(n-1) \quad (*)$$

$Z_1, \dots, Z_n \overset{iid}{\sim} N(0,1)$ $\quad E[Z_i]=0$, $var[Z_i]=1$

pdf of χ^2_n



① $\sum_{i=1}^{n} Z_i \sim N(0, n)$ ② $\frac{1}{n}\sum_{i=1}^{n} Z_i = \bar{Z}_n \sim N(0, \frac{1}{n})$

③ $\sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$ $\quad E\left[\sum_{i=1}^{n} Z_i^2\right] = n$, $var\left[\sum_{i=1}^{n} Z_i^2\right] = 2n$

$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

① $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$ ② $\frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$

③ $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_n$

$Z_1^2, \dots, Z_n^2 \overset{iid}{\sim} \chi^2_1$, $var[Z_i^2] = 2 < \infty$ so by CLT, $\frac{\sum_{i=1}^{n} Z_i^2 - n}{\sqrt{2n}} \overset{approx}{\sim} N(0,1)$

$\Leftrightarrow \sum_{i=1}^{n} Z_i^2 \overset{approx}{\sim} N(n, 2n)$

$(*) = \frac{\sigma^2}{2}\left(\frac{[illegible]}{n}\right)\cdot n + \frac{\sigma^2}{2}\left[[illegible] \right] = [illegible] + \frac{\sigma^2}{2n}\left[2(st-t^2)+nt^2\right]$

$$= \frac{\sigma^2}{2n}\left[s^2 - \cancel{2st} + t^2 + \cancel{2st} - 2t^2 + nt^2\right] = \frac{\sigma^2}{2n}\left[s^2 + (n-1)t^2\right]$$

So $\overline{X}_n$ and $X_i - \overline{X}_n$ are independent.

<u>Corollary</u>: Let $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$. Then $\overline{X}_n$ and S_n^2 are independent.

Pf: Since $\overline{X}_n$ ~~and~~ is independent of $X_1 - \overline{X}_n, \cdots, X_n - \overline{X}_n$, and S_n^2 is a function of $X_1 - \overline{X}_n$ ~~and~~, $\cdots$, $X_n - \overline{X}_n$. Hence, $\overline{X}_n$ and S_n^2 are independent.

<u>Theorem 2</u> $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$ $\qquad \frac{(n-1)S_n^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{\sigma^2}$

Pf: Recall that $\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(\boxed{X_i - \overline{X}_n} + \boxed{\overline{X}_n - \mu})^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left[(X_i - \overline{X}_n)^2 + 2(X_i - \overline{X}_n)(\overline{X}_n - \mu) + (\overline{X}_n - \mu)^2\right]}{\sigma^2}$$

$$= \frac{(n-1)S_n^2 + 2(\overline{X}_n - \mu)\underbrace{\boxed{\sum_{i=1}^{n}(X_i - \overline{X}_n)}}_{=0} + n(\overline{X}_n - \mu)^2}{\sigma^2}$$

$$= \frac{(n-1)S_n^2}{\sigma^2} + \underbrace{\frac{n(\overline{X}_n - \mu)^2}{\sigma^2}}_{\sim \chi^2_1}$$

Note
$\sum_{i=1}^{n}(X_i - \overline{X}_n)$
$= \left(\sum_{i=1}^{n} X_i\right) - n\overline{X}_n$
$= 0$

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

Recall $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
$\Leftrightarrow \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$
$\Rightarrow \left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2 = \frac{n(\overline{X}_n - \mu)^2}{\sigma^2} \sim \chi^2_1$

Denote $W = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2}$

$U = \frac{(n-1)S_n^2}{\sigma^2}$, $V = \frac{n(\overline{X}_n - \mu)^2}{\sigma^2}$

We have $W \sim \chi^2_n$, $V \sim \chi^2_1$. U and V are independent.

$W = U + V$

So $M_W(t) = M_U(t) \cdot M_V(t)$

Recall that the MGF of χ^2_n is $(1-2t)^{-\frac{n}{2}}$

so $(1-2t)^{-\frac{n}{2}} = M_U(t)\cdot(1-2t)^{-\frac{1}{2}}$

$\Rightarrow M_U(t) = (1-2t)^{-\frac{n-1}{2}}$

So $U \sim \chi^2_{n-1}$

t distribution

Let $Z \sim N(0,1)$, $U \sim \chi^2_n$, and Z and U are independent.

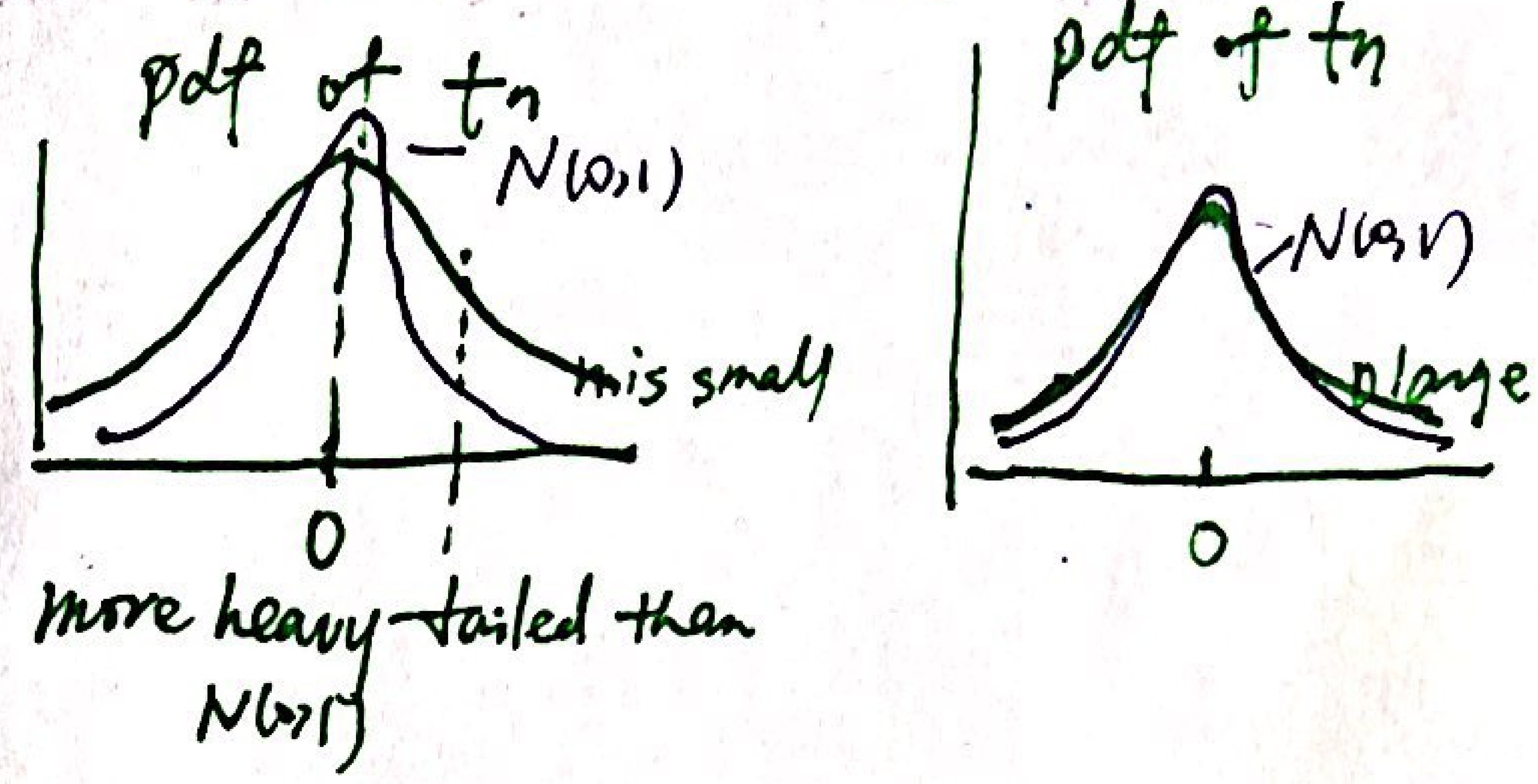
Then the distribution of $\frac{Z}{\sqrt{\frac{U}{n}}}$ is defined as the $\boxed{t_n}$ — the Student's t distribution with n degrees of freedom.

Properties of $T \sim t_n$:

① $E[T] = 0$

② $var[T] = \frac{n}{n-2}$

③ $T \xrightarrow{d} N(0,1)$



Theorem 3 Let $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

Then $\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}$.

Pf: From Theorem 2, $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$.

We also have $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ and $\overline{X}_n$ and S_n^2 are independent

Hence $\frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\frac{(n-1)S_n^2}{\sigma^2}}{n-1}}} \sim t_{n-1}$ by the definition of t distribution

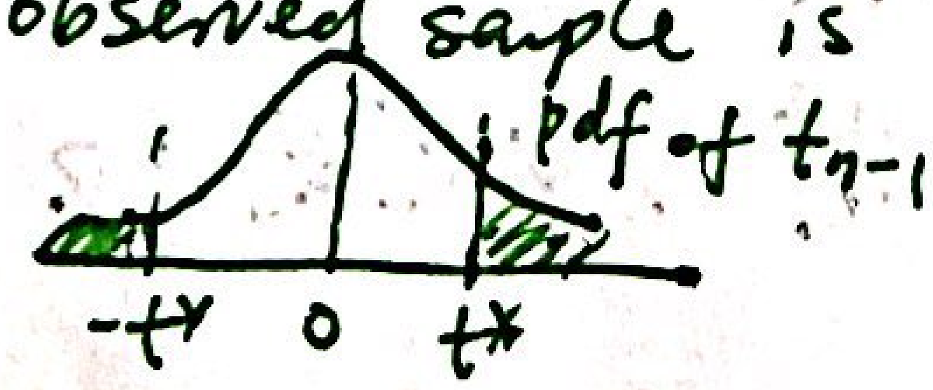
$$= \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\frac{S_n}{\sigma}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \cdot \frac{\sigma}{S_n} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \quad \square$$

Application: One-sample t test

$X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. $H_0: \mu = c$ vs $H_1: \mu \neq c$

test statistic: $T = \frac{\overline{X}_n - c}{S_n/\sqrt{n}} \overset{H_0}{\sim} t_{n-1}$

Say the realization of T in our observed sample is t^*

p-value $= \mathbb{P}_{H_0}(|T| \geq |t^*|) =$ 

Let $U \sim \chi^2_m$, $V \sim \chi^2_n$, and U and V are independent.

Then the distribution of $\dfrac{U/m}{V/n}$ ~~has~~ is defined as $F_{m,n}$ — the F distribution with $\boxed{m}$ numerator degrees of freedom and $\boxed{n}$ denominator degrees of freedom.