IC105: Probability and Statistics **2021-22-M**

Lecture 21: Distributions Based on Sampling from Normal Distribution

Instructor: Dr. Kuldeep Kumar Kataria *Scribe:*

Theorem 21.1. *Let $X_1, X_2, \ldots, X_n$ $(n \geq 2)$ be a random sample from $N(\mu, \sigma^2)$ distribution, where $-\infty < \mu < \infty$ and $\sigma > 0$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ denote the sample mean and sample variance, respectively. Then*

(i) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$*;*

(ii) $\bar{X}$ *and* S^2 *are independent r.v.'s;*

(iii) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$*;*

(iv) $E(S^2) = \sigma^2$, $\mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}$, $E(S) = \sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma(\frac{n-1}{2})}\sigma$.

Proof. (i) Follows from last theorem by taking $k = n$, $a_i = \frac{1}{n}$, $\mu_i = \mu$, $\sigma_i^2 = \sigma$, $i = 1, 2, \ldots, n$.

(ii) Let $Y_i = X_i - \bar{X}$, $i = 1, 2, \ldots, n$ and let $\underline{Y} = (Y_1, Y_2, \ldots, Y_n)$. Then

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n}(X_i - \bar{X}) = \sum_{i=1}^{n} X_i - n\bar{X} = 0$$

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} Y_i^2 \quad \text{(a function of } \underline{Y}\text{)}$$

The joint m.g.f. of $(\underline{Y}, \bar{X})$ is given by

$$M_{\underline{Y},\bar{X}}(\underline{t}) = E\left(e^{\sum_{i=1}^{n} t_i Y_i + t_{n+1}\bar{X}}\right), \ \ \underline{t} = (t_1, t_2, \ldots, t_n, t_{n+1}) \in \mathbb{R}^{n+1}$$

$$\begin{aligned}
\sum_{i=1}^{n} t_i Y_i + t_{n+1}\bar{X} &= \sum_{i=1}^{n} t_i(X_i - \bar{X}) + t_{n+1}\bar{X} \\
&= \sum_{i=1}^{n} t_i X_i + \frac{(t_{n+1} - \sum_{i=1}^{n} t_i)}{n}\sum_{i=1}^{n} X_i \\
&= \sum_{i=1}^{n}\left(t_i - \bar{t} + \frac{t_{n+1}}{n}\right) X_i, \ \text{where } \bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i \\
&= \sum_{i=1}^{n} u_i X_i, \ \ \text{where } u_i = t_i - \bar{t} + \frac{t_{n+1}}{n}, \ i = 1(1)n.
\end{aligned}$$

Then $\sum_{i=1}^n u_i = t_{n+1}$ and $\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (t_i - \bar{t})^2 + \frac{t_{n+1}^2}{n}$.

$$\begin{aligned}
M_{\underline{Y},\bar{X}}(\underline{t}) &= E\left(e^{\sum_{i=1}^n u_i X_i}\right) \\
&= \prod_{i=1}^n M_{X_i}(u_i) \\
&= \prod_{i=1}^n e^{\mu u_i + \frac{1}{2}\sigma^2 u_i^2} \\
&= e^{\mu \sum_{i=1}^n u_i + \frac{\sigma^2}{2}\sum_{i=1}^n u_i^2} \\
&= e^{\mu t_{n+1} + \frac{\sigma^2}{2}\{\sum_{i=1}^n (t_i-\bar{t})^2 + \frac{t_{n+1}^2}{n}\}} = \left\{e^{\mu t_{n+1} + \frac{\sigma^2 t_{n+1}^2}{2n}}\right\}\left\{e^{\frac{\sigma^2}{2}\sum_{i=1}^n (t_i-\bar{t})^2}\right\}
\end{aligned}$$

$$\begin{aligned}
& M_{\underline{Y}}(t_1, t_2, \dots, t_n) = M_{\underline{Y},\bar{X}}(t_1, t_2, \dots, t_n, 0) = e^{\frac{\sigma^2}{2}\sum_{i=1}^n (t_i-\bar{t})^2}, (t_1, t_2, \dots, t_n) \in \mathbb{R}^n \\
& M_{\bar{X}}(t_{n+1}) = M_{\underline{Y},\bar{X}}(0, \dots, 0, t_{n+1}) = e^{\mu t_{n+1} + \frac{\sigma^2 t_{n+1}^2}{2n}},\ t_{n+1} \in \mathbb{R} \\
\implies & M_{\underline{Y},\bar{X}}(\underline{t}) = M_{\underline{Y}}(t_1, t_2, \dots, t_n) M_{\bar{X}}(t_{n+1}), \forall\, \underline{t} = (t_1, t_2, \dots, t_n, t_{n+1}) \in \mathbb{R}^{n+1} \\
\implies & \underline{Y} \text{ and } \bar{X} \text{are independent} \\
\implies & \sum_{i=1}^n (X_i - \bar{X})^2 \text{ and } \bar{X} \text{ are independent.}
\end{aligned}$$

(iii) Let $Z_i = \frac{X_i - \mu}{\sigma}$, $i = 1, 2, \dots, n$. Then $Z_1, Z_2, \dots, Z_n$ are iid $N(0,1)$ r.v.'s. Also let $Z = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$ (using (i)). Let

$$W = Z^2 = \frac{n(\bar{X}-\mu)^2}{\sigma^2} \text{ and } T = \frac{(n-1)S^2}{\sigma^2}.$$

Then by (ii), W and T are independent r.v.s. Also $W \sim \chi_1^2$ and $V = \sum_{i=1}^n Z_i^2 \sim \chi_n^2$.

$$\begin{aligned}
V = \sum_{i=1}^n Z_i^2 &= \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 \\
&= \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu)^2 \\
&= \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2 + \frac{n(\bar{X}-\mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2} = T + W.
\end{aligned}$$

This implies

$$\begin{aligned}
& M_V(t) = M_T(t) M_W(t) \\
\implies & M_T(t) = \frac{M_V(t)}{M_W(t)} = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} = (1-2t)^{-(n-1)/2},\ \ t < \frac{1}{2} \to \text{m.g.f. of } \chi_{n-1}^2 \\
\implies & T = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.
\end{aligned}$$

(iv) $T = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_v$, where $v = n-1$. Thus

$$E(T^s) = \int_0^\infty t^s \frac{1}{2^{v/2}\Gamma(v/2)} e^{-t/2} t^{v/2-1} \mathrm{d}t$$
$$= \frac{1}{2^{v/2}\Gamma(v/2)} \int_0^\infty e^{-t/2} t^{\frac{v+2s}{2}-1} \mathrm{d}t = \frac{2^{\frac{v+2s}{2}}\Gamma\left(\frac{v+2s}{2}\right)}{2^{v/2}\Gamma(v/2)} = \frac{2^s\Gamma\left(\frac{v+2s}{2}\right)}{\Gamma(v/2)},\ s > -\frac{v}{2}.$$

This implies

$$\frac{(n-1)^s}{\sigma^{2s}} E(S^{2s}) = \frac{2^s\Gamma\left(\frac{v+2s}{2}\right)}{\Gamma(v/2)}$$
$$\implies E(S^r) = \left(\frac{2}{n-1}\right)^{r/2} \frac{\Gamma\left(\frac{v+r}{2}\right)}{\Gamma(v/2)} \sigma^r,\ \ r > v$$
$$\implies E(S^r) = \left(\frac{2}{n-1}\right)^{r/2} \frac{\Gamma\left(\frac{n-1+r}{2}\right)}{\Gamma((n-1)/2)} \sigma^r,\ \ r > v$$
$$\implies E(S) = \left(\frac{2}{n-1}\right)^{1/2} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma((n-1)/2)} \sigma$$
$$\implies E(S^2) = \left(\frac{2}{n-1}\right) \frac{\Gamma\left(\frac{n-1}{2}+1\right)}{\Gamma((n-1)/2)} \sigma^2 = \sigma^2$$
$$\implies E(S^4) = \left(\frac{2}{n-1}\right)^2 \frac{\Gamma\left(\frac{n-1}{2}+2\right)}{\Gamma((n-1)/2)} \sigma^4 = \frac{n+1}{n-1}\sigma^4$$
$$\mathrm{Var}(S^2) = E(S^4) - (E(S^2))^2 = \frac{2\sigma^4}{n-1}.$$

This completes the proof. □

Remark 21.2. *Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution having p.m.f. / p.d.f. f. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Let $E(X_1) = \mu$ and $\mathrm{Var}(X_1) = \sigma^2$. Then $E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \mu$, $\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{\sigma^2}{n}$.*

$$E\left[(n-1)S^2\right] = E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right]$$
$$\implies (n-1)E(S^2) = E\left[\sum_{i=1}^n X_i^2 - n\bar{X}^2\right]$$
$$= \sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2)$$
$$= n\left[E(X_1^2) - E(\bar{X}^2)\right]$$
$$= n\left[\mathrm{Var}(X_1) + (E(X_1))^2 - \mathrm{Var}(\bar{X}) - (E(\bar{X}))^2\right]$$
$$= n(\sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2) = (n-1)\sigma^2 \implies E(S^2) = \sigma^2.$$

For this reason $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is called sample variance and not $S_1^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$. Note that $E(S_1^2) = E\left(\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{n-1}{n}\sigma^2 < \sigma^2$, i.e., S_1^2 underestimates σ^2.

21.1. Distributions Based on Sampling from Normal Distribution

Definition 21.3. *(a) For a positive integer m, a random variable X is said to have the student t-distribution with m degrees of freedom (written as $X \sim t_m$) if the p.d.f. of X is given by*

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

(b) For positive integers n_1 and n_2 a random variable X is said to have the Snedecor F distribution with (n_1, n_2) degrees of freedom (written as $X \sim F_{n_1,n_2}$) if its p.d.f. is given by

$$f(x|n_1, n_2) = \begin{cases} \dfrac{\left(\dfrac{n_1}{n_2}\right)\left(\dfrac{n_1 x}{n_2}\right)^{\dfrac{n_1}{2}-1}}{B(n_1/n_2, n_2/2)} \left(1 + \frac{n_1 x}{n_2}\right)^{-(n_1+n_2)}, \ x > 0, \\ 0, \ \textit{otherwise}. \end{cases}$$

Remark 21.4. *(a) Note that*

$$\begin{aligned} X \sim t_m &\implies f(x|m) = f(-x|m), \ \forall\, x \\ &\implies X \overset{d}{=} -X \\ &\implies \textit{ distribution of } X \textit{ is symmetric about } 0 \implies m_e = 0 \textit{ and } E(X) = 0, \textit{ provided it exists}. \end{aligned}$$

(b) $X \sim t_m \implies f(x|m) \uparrow m(-\infty, 0), \downarrow (0, \infty) \implies m_0 = 0$.

(c) t_1 distribution is nothing but Cauchy distribution with p.d.f.

$$f(x|1) = \frac{1}{\pi}\frac{1}{1+x^2}, \ -\infty < x < \infty \implies E(X) \textit{ does not exist}.$$

(d) Let $X \sim F_{n_1,n_2}$. Then,

$$f(x|n_1, n_2) = \begin{cases} \dfrac{\frac{n_1}{n_2}}{B(n_1/2, n_2/2)} \left(\dfrac{\frac{n_1 x}{n_2}}{1+\frac{n_1 x}{n_2}}\right)^{\frac{n_1}{2}-1} \left(\dfrac{1-\frac{n_1 x}{n_2}}{1+\frac{n_1 x}{n_2}}\right)^{n_2-1} \left(1 + \dfrac{n_1 x}{n_2}\right)^{-2}, \ x > 0, \\ 0, \ \textit{otherwise}. \end{cases}$$

$$\implies Y = \frac{\frac{n_1 X}{n_2}}{1+\frac{n_1 X}{n_2}} \sim Be\left(n_1/2, n_2/2\right).$$

Theorem 21.5. *(a) Let $Z \sim N(0,1)$ and let $Y \sim \chi^2_m$, $m \in \{1, 2, \ldots\}$ be independent random variables. Then*

$$T = \frac{Z}{\sqrt{Y/m}} \sim t_m.$$

(b) For positive integers n_1 and n_2, let $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$ be independent random variables, then

$$U = \frac{X_1/n_1}{X_2/n_2} \sim F_{n_1,n_2}.$$

(c) Let $X \sim t_m$. Then $E(X^2)$ is not finite if $r \in \{m, m+1, \ldots\}$. For $r \in \{1, 2, \ldots, m-1\}$ $(m \geq r+1)$

$$E(X^r) = \begin{cases} 0, \ \textit{if } r \textit{ is odd}, \\ \dfrac{m^{r/2} r! \Gamma((m-r)/2)}{2^r (r/2)! \Gamma(m/2)}, \ \textit{if } r \textit{ is even}. \end{cases}$$

(d) If $X \sim t_m$ *then*

$$\begin{aligned}
\textit{Mean} &= \mu_1' = E(X) = 0, \;\; m = 2, 3, \ldots, \\
\operatorname{Var}(X) &= \mu_2 = E((X-\mu_1')^2) = \frac{m}{m-2}, \;\; m \in \{3, 4, \ldots\}, \\
\textit{Coefficient of skewness} &= \beta_1 = 0, \; m = 4, 5, 6, \ldots, \\
\textit{Kurtosis} &= \nu_1 = \frac{3(m-2)}{m-4}, \;\; m \in \{5, 6, \ldots\}.
\end{aligned}$$

(e) Let n_1*,* n_2 *and* r *be positive integers, and let* $X \sim F_{n_1,n_2}$*. Then, for* $n_2 \in \{1, 2, \ldots, 2r\}$ *and* $r \geq \frac{n_2}{2}$*, it follows that* $E(X^r)$ *is not finite. For* $n_2 \in \{2r+1, 2r+2, \ldots\}$ *and* $r \geq \frac{n_2-1}{2}$*, we have*

$$E(X^r) = \left(\frac{n_2}{n_1}\right)^r \prod_{i=1}^{r} \left(\frac{n_1 + 2(i-1)}{n_2 - 2i}\right).$$

(f) If $X \sim F_{n_1,n_2}$ *then*

$$\begin{aligned}
\textit{Mean} &= \mu_1' = E(X) = \frac{n_2}{n_2 - 2}, \; \textit{if } n_2 \in \{3, 4, \ldots\}, \\
\operatorname{Var}(X) &= \mu_2 = E((X-\mu_1')^2) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_1-2)^2(n_2-4)}, \;\; \textit{if } n_2 \in \{5, 6, \ldots\}, \\
\textit{Coefficient of skewness} &= \beta_1 = \frac{2(2n_1+n_2-2)}{n_2-6}\sqrt{\frac{2(n_2-4)}{n_1(n_1+n_2-2)}}, \;\; n_2 \in \{7, 8, \ldots\}, \\
\textit{Kurtosis} &= \nu_1 = \frac{12\left[(n_2-2)^2(n_2-4) + n_1(n_1+n_2-2)(5n_2-22)\right]}{n_1(n_2-6)(n_2-8)(n_1+n_2-2)}.
\end{aligned}$$

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

$$f_{Y,Z}(y,z) = f_Y(y)f_Z(z) = \begin{cases} \dfrac{1}{2^{(m+1)/2}\Gamma(m/2)\sqrt{\pi}} e^{-\frac{y+z^2}{2}} y^{\frac{m}{2}-1}, \; y > 0, \; -\infty < z < \infty, \\ 0, \; \text{otherwise}. \end{cases}$$

Let $U = \sqrt{\frac{Y}{m}}$. $S_{Y,Z} = (0,\infty) \times \mathbb{R}$. Let $\underline{h} = (h_1, h_2) : (0,\infty) \times \mathbb{R} \to \mathbb{R}^2$ where $h_1(y,z) = \frac{z}{\sqrt{y/m}}$ and $h_2(y,z) = \sqrt{y/m}$. The transformation $\underline{h} : S_{Y,Z} \to \mathbb{R}$ is 1-1 with inverse transformation $\underline{h}^{-1} = (h_1^{-1}, h_2^{-1})$, where $h_1^{-1}(t,u) = mu^2$, $h_2^{-1}(t,u) = tu$, $J = \begin{vmatrix} 0 & 2mu \\ u & t \end{vmatrix} = -2mu^2$.

$$\underline{h}(S_{Y,Z}) = \{(t,u) : mu^2 > 0, \; -\infty < tu < \infty\} = \{(t,u) : u > 0, \; t \in \mathbb{R}\} = \mathbb{R} \times (0,\infty).$$

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

$$\begin{aligned}
f_{T,U}(t,u) &= f_{Y,Z}(h_1^{-1}(t,u), h_2^{-1}(t,u))|J| I_{h(S_{Y,Z})}(t,u) \\
&= \frac{1}{2^{(m+1)/2}\Gamma(m/2)\sqrt{\pi}} e^{-\frac{mu^2+t^2u^2}{2}} (mu^2)^{\frac{m}{2}-1} |2mu^2| I_{\mathbb{R}\times(0,\infty)}(t,u) \\
&= \frac{m^{m/2}u^m}{2^{(m-1)/2}\Gamma(m/2)\sqrt{\pi}} e^{-\frac{(m+t^2)u^2}{2}} I_{\mathbb{R}}(t) I_{(0,\infty)}(u).
\end{aligned}$$

The marginal p.d.f. of T is

$$\begin{aligned}
f_T(t) &= \int_{-\infty}^{\infty} f_{T,U}(t,u)\mathrm{d}u \\
&= \frac{m^{m/2}}{2^{(m-1)/2}\Gamma(m/2)\sqrt{\pi}} \int_{-\infty}^{\infty} u^m e^{-\frac{(m+t^2)u^2}{2}}\mathrm{d}u \\
&= \frac{1}{\Gamma(m/2)\sqrt{m\pi}} \frac{1}{(1+t^2/m)^{\frac{m+1}{2}}} \int_{-\infty}^{\infty} y^{\frac{m-1}{2}} e^{-y}\mathrm{d}y \ \ (u^2 = y) \\
&= \frac{\Gamma(\frac{m+1}{2})}{\Gamma(m/2)\sqrt{m\pi}} \frac{1}{(1+t^2/m)^{\frac{m+1}{2}}}, \quad t \in \mathbb{R} \longrightarrow \text{p.d.f. of } t_m.
\end{aligned}$$

(b) The joint p.d.f. of $\underline{X} = (X_1, X_2)$ is given by

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2) = \frac{1}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} e^{-\frac{x_1+x_2}{2}} x_1^{\frac{n_1}{2}-1} x_2^{\frac{n_2}{2}-1} I_{(0,\infty)\times(0,\infty)}(x_1,x_2).$$

Let $V = \frac{X_2}{n_2}$. $S_{\underline{X}} = (0,\infty)\times(0,\infty)$. Consider the transformation: $\underline{h} = (h_1, h_2) : (0,\infty)\times(0,\infty) \to \mathbb{R}$ defined by $h_1(x_1,x_2) = \frac{x_1/n_1}{x_2/n_2}$ and $h_2(x_1,x_2) = \frac{x_2}{n_2}$ so that $U = h_1(X_1,X_2)$ and $V = h_2(X_1,X_2)$.

The transformation $\underline{h} : (0,\infty)\times(0,\infty) \to \mathbb{R}^2$ is 1-1 with inverse transformation $\underline{h}^{-1} = (h_1^{-1}, h_2^{-1})$, where

$$h_1^{-1}(u,v) = n_1uv, h_2^{-1}(u,v) = n_2v, \ \ J = \begin{vmatrix} n_1v & n_1u \\ 0 & n_2 \end{vmatrix} = n_1n_2v,$$

$$\underline{h}(S_{\underline{X}}) = \{(u,v) : n_1uv > 0, \ n_2v > 0\} = \{(u,v) : u > 0, \ v > 0\} = (0,\infty)\times(0,\infty).$$

Thus, the joint p.d.f. of (U,V) is given by

$$\begin{aligned}
f_{U,V}(u,v) &= f_{X_1,X_2}(h_1^{-1}(u,v), h_2^{-1}(u,v))|J| I_{\underline{h}(S_{\underline{X}})}(u,v) \\
&= \frac{n_1^{n_1/2} n_2^{n_2/2}}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} e^{-\frac{(n_2+n_1u)v}{2}} u^{\frac{n_1}{2}-1} v^{\frac{n_1+n_2}{2}-1} I_{(0,\infty)}(u) I_{(0,\infty)}(v)
\end{aligned}$$

The marginal p.d.f. of U is given by

$$\begin{aligned}
f_U(u) &= \int_{-\infty}^{\infty} f_{U,V}(u,v)\mathrm{d}v \\
&= \frac{n_1^{n_1/2} n_2^{n_2/2}}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} u^{\frac{n_1}{2}-1} \int_0^{\infty} e^{-\frac{(n_2+n_1u)v}{2}} v^{\frac{n_1+n_2}{2}-1}\mathrm{d}v \\
&= \frac{\Gamma(\frac{n_1+n_2}{2})}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} \frac{(n_1u/n_2)^{\frac{n_1}{2}-1}}{(1+n_1u/n_2)^{\frac{n_1+n_2}{2}}} I_{(0,\infty)} \longrightarrow \text{p.d.f. of } F_{n_1,n_2}.
\end{aligned}$$

(c) Fix $m \in \{1,2,\dots\}$. Then $X \stackrel{d}{=} \frac{Z}{\sqrt{Y/m}}$ where $Z \sim N(0,1)$ and $Y \sim \chi_m^2$ are independent. This implies that

$$\begin{aligned}
E(X^r) &= E\left(\frac{Z}{\sqrt{Y/m}}\right)^r = m^{r/2}E(Z^rY^{-r/2}) = m^{r/2}E(Z^r)E(Y^{-r/2}) \ \ (Y \text{ and } Z \text{ are independent}) \\
E(Z^r) &= \begin{cases} 0, \ \text{if } r = 1,3,5,\dots, \\ \frac{r!}{2^{r/2}(r/2)!}, \ \ \text{if } r = 2,4,6\dots. \end{cases} \\
E(Y^{-r/2}) &= \frac{1}{2^{m/2}(m/2)!} \int_0^{\infty} y^{\frac{m-r}{2}-1} e^{-y/2}\mathrm{d}y = \infty, \ \ \text{if } r \geq m.
\end{aligned}$$

For $r < m$, we have

$$E(Y^{-r/2}) = \frac{2^{\frac{m-r}{2}}\Gamma(\frac{m-r}{2})}{2^{m/2}\Gamma(m/2)} = \frac{\Gamma(\frac{m-r}{2})}{2^{r/2}\Gamma(m/2)}$$
$$\implies E(X^r) = \begin{cases} 0, \text{ if } r \text{ is odd and } r < m, \\ \dfrac{m^{r/2} r! \Gamma(m-r)/2}{2^r (r/2)! \Gamma(m/2)}, \text{ if } r \text{ is even and } \ r < n. \end{cases}$$

(d) Exercise.

(e) Fix $n_1, n_2 \in \mathbb{N}$. Then

$$X \overset{d}{=} \frac{X_1/n_1}{X_2/n_2} = \frac{n_2}{n_1}\frac{X_1}{X_2},$$

where $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$ are independent. For $r \in \mathbb{N}$,

$$E(X^r) = \left(\frac{n_2}{n_1}\right)^r E\left(\frac{X_1^r}{X_2^r}\right) = \left(\frac{n_2}{n_1}\right)^r E(X_1^r) E\left(\frac{1}{X_2^r}\right),$$
$$E(X_1^r) = \frac{1}{2^{n_1/2}\Gamma(n_1/2)}\int_0^\infty x^{\frac{n_1+2r}{2}-1} e^{-x/2} \mathrm{d}x$$
$$= \frac{2^{\frac{n_1+2r}{2}}\Gamma(\frac{n_1+2r}{2})}{2^{n_1/2}\Gamma(n_1/2)} = \prod_{i=1}^r (n_1 - 2(i-1)), \ \ r \in \{1, 2, \dots\},$$
$$E\left(\frac{1}{X_2^2}\right) = \begin{cases} \dfrac{2^{\frac{n_1-2r}{2}}\Gamma(\frac{n_2-2r}{2})}{2^{n_2/2}\Gamma(n_2/2)}, \ n_2 > 2r, \\ \infty, \text{ if } n_2 \leq 2r \end{cases} = \begin{cases} \prod_{i=1}^r (n_2 - 2i), \ n_2 > 2r, \\ \infty, \text{ if } n_2 \leq 2r, \end{cases}$$
$$\implies E(X^r) = \begin{cases} \left(\frac{n_2}{n_1}\right)^r \prod_{i=1}^r \left(\frac{n_1+2(i-1)}{n_2-2i}\right), \ n_2 > 2r, \\ \infty, \text{ if } n_2 \leq 2r. \end{cases}$$

(f) Exercise. $\square$

Corollary 21.6. *Let $X_1, X_2, \dots, X_n$ ($n \geq 2$) be a random sample from $N(\mu, \sigma^2)$ distribution, where $\mu \in \mathbb{R}$ and $\sigma > 0$. Let*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \ \ S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$$

denote the sample mean and sample variance, respectively. Then,

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}.$$

Proof. We know that

$$\bar{X} \sim N(\mu, \sigma^2/n) \text{ and } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \text{ independent}$$
$$\implies \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1) \text{ and } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \text{ independent}$$
$$\implies \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{S^2/\sigma^2}} \sim t_{n-1}, \text{ that is, } \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}.$$

This completes the proof. $\square$

Corollary 21.7. *Let $X_1, X_2, \ldots, X_m$ ($m \geq 2$) and $Y_1, Y_2, \ldots, Y_n$ ($n \geq 2$) be independent random samples (that is, $\underline{X} = (X_1, X_2, \ldots, X_m)$ and $\underline{Y} = (Y_1, Y_2, \ldots, Y_m)$ are independent) from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ distribution, respectively where $\mu_i \in \mathbb{R}$, $i = 1, 2$, and $\sigma_i > 0$, $i = 1, 2$. Let*

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \;\; \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \;\; S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2, \;\; S_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Then, (a) $\dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0, 1)$,

(b) $\dfrac{1}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(m-1)S_1^2/\sigma_1^2 + (n-1)S_1^2/\sigma_2^2}{m+n-2}}} \sim t_{m+n-2}$,

(c) $\dfrac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F_{m-1, n-1}$.

Proof. $\bar{X} \sim N(\mu_1, \sigma_1^2/m)$, $\bar{Y} \sim N(\mu_2, \sigma_2^2/n)$, $\frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2_{m-1}$ and $\frac{(n-1)S_2^2}{\sigma_2^2}$ are independent r.v.s. Thus,

$$\begin{aligned}
&\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma_1^2/m + \sigma_2^2/n\right) \\
&\frac{(m-1)S_1^2}{\sigma_1^2} + \frac{(n-1)S_2^2}{\sigma_2^2} \sim t_{m+n-2} \\
\implies &\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0, 1) \text{ and } \frac{1}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(m-1)S_1^2/\sigma_1^2 + (n-1)S_1^2/\sigma_2^2}{m+n-2}}} \sim t_{m+n-2}.
\end{aligned}$$

This completes the proof. □

Remark 21.8. *(a) Note that*

$$\begin{aligned}
&X \sim t_m \\
\implies &X \stackrel{d}{=} \frac{N(0,1)}{\sqrt{\chi_m^2/m}} \Big\rangle \textit{independent} \\
\implies &X^2 \stackrel{d}{=} \frac{(N(0,1))^2}{\chi_m^2/m} \Big\rangle \textit{independent} \\
&\quad = \frac{\chi_1^2}{\chi_m^2/m} \Big\rangle \textit{independent} \stackrel{d}{=} F_{1,m}.
\end{aligned}$$

Thus, $X \sim t_m \implies X^2 \sim F_{1,m}$.

(b) Note that

$$\begin{aligned}
&X \sim F_{n_1, n_2} \\
\implies &X \stackrel{d}{=} \frac{\chi_{n_1}^2/n_1}{\chi_{n_2}^2/n_2} \Big\rangle \textit{independent} \\
\implies &\frac{1}{X} \stackrel{d}{=} \frac{\chi_{n_2}^2/n_2}{\chi_{n_1}^2/n_1} \Big\rangle \textit{independent} \stackrel{d}{=} F_{n_2, n_1}.
\end{aligned}$$

Thus, $X \sim F_{n_1, n_2} \implies \frac{1}{X} \sim F_{n_2, n_1}$.

(c) $X \sim t_m \implies$ *Kurtosis =* $\nu_1 = \frac{3(m-2)}{m-4},\ m > 4 \implies t_m$ *distribution* $(m > 4)$ *is symmetric and leptokurtic (that is, it has sharper peak and longer fatter tails compared to* $N(0,1)$ *distribution). As* $m \to \infty, \nu_1 \to \infty$*. This suggests that for large d.f.* m*,* t_m *distribution behaves like* $N(0,1)$ *distribution.*

(d) For various values of $m \in \mathbb{N}$ *and* $\alpha \in (0,1)$*, the d.f. of* t_m *is tabulated in various text books.*

(e) For fixed $n_1 \in \mathbb{N}$*,* $n_2 \in \mathbb{N}$ *and* $\alpha \in (0,1)$ *let* $f_{n_1,n_2,\alpha}$ *be the* $(1-\alpha)$*-th quantile of* $X \sim F_{n_1,n_2}$*. Thus*

$$P(X \le f_{n_1,n_2,\alpha}) = 1-\alpha \implies P\left(\frac{1}{X} \le \frac{1}{f_{n_1,n_2,\alpha}}\right) = \alpha \implies f_{n_2,n_1,1-\alpha} = \frac{1}{f_{n_1,n_2,\alpha}} \ \left(\textit{as } \frac{1}{X} \sim F_{n_2,n_1}\right).$$

Example 21.9. *Let* $X_1, X_2, \ldots, X_n$ *be a random sample from* $N(\mu, \sigma^2)$ *distribution, where* $\mu \in \mathbb{R}$*,* $\sigma > 0$ *and* $n \ge 2$*. Let*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ \textit{ and } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

be the sample mean and sample variance, respectively. Evaluate $E\left(\frac{\bar{X}}{S}\right)$*, for* $n > 2$*.*

Solution: We have

$$\begin{aligned}
&\bar{X} \sim N(\mu, \sigma^2/n) \ \text{ and } Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \text{ are independent} \\
\implies E\left(\frac{\bar{X}}{S}\right) &= \frac{\sqrt{n-1}}{\sigma} E(\bar{X} Y^{-1/2}) \\
&= \frac{\sqrt{n-1}}{\sigma} E(\bar{X}) E(Y^{-1/2}) \ \text{ (indepedence)} \\
&= \frac{\sqrt{n-1}}{\sigma} \mu \int_0^\infty \frac{e^{-y/2} y^{\frac{n-2}{2}-1}}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} \mathrm{d}y \\
&= \frac{\sqrt{n-1}}{\sigma} \mu \frac{2^{\frac{n-2}{2}} \Gamma(\frac{n-2}{2})}{2^{\frac{n-1}{2}} \Gamma(\frac{n-1}{2})} = \frac{\sqrt{(n-1)/2}}{\sigma} \frac{\Gamma(\frac{n-2}{2})}{\Gamma(\frac{n-1}{2})} \frac{\mu}{\sigma}.
\end{aligned}$$

Example 21.10. *Let* $Z_1, Z_2, \ldots, Z_n$ *be iid* $N(0,1)$ *r.v.s and let* $a_i, b_i \in \mathbb{R}$*,* $i = 1, \ldots, n$ *be such that* $\sum_{i=1}^n a_i^2 > 0$*,* $\sum_{i=1}^n b_i^2 > 0$ *and* $\sum_{i=1}^n a_i b_i = 0$*. Show that*

(a) $Y_1 = \sqrt{\frac{\sum_{i=1}^n b_i^2}{\sum_{i=1}^n a_i^2}} \frac{\sum_{i=1}^n a_i Z_i}{|\sum_{i=1}^n b_i Z_i|} \sim t_1$*;*

(b) $Y_2 = \frac{\sum_{i=1}^n b_i^2}{\sum_{i=1}^n a_i^2} \left(\frac{\sum_{i=1}^n a_i Z_i}{\sum_{i=1}^n b_i Z_i}\right)^2 \sim F_{1,1}$*;*

(c) $Y_3 = \sqrt{\frac{\sum_{i=1}^n b_i^2}{\sum_{i=1}^n a_i^2}} \frac{\sum_{i=1}^n a_i Z_i}{\sum_{i=1}^n b_i Z_i} \sim t_1$*.*

Solution: Linear combination of $Z_1, Z_2, \ldots, Z_n$:

$$
\begin{aligned}
& c_1 \sum_{i=1}^{n} a_i Z_i + c_2 \sum_{i=1}^{n} b_i Z_i \quad \text{(univariate normal distribution)} \\
\implies & \left(\sum_{i=1}^{n} a_i Z_i, \sum_{i=1}^{n} b_i Z_i \right) \sim N_2 \\
& E\left(\sum_{i=1}^{n} a_i Z_i \right) = 0, \;\; \mathrm{Var}\left(\sum_{i=1}^{n} a_i Z_i \right) = \sum_{i=1}^{n} a_i^2, \\
& E\left(\sum_{i=1}^{n} b_i Z_i \right) = 0, \;\; \mathrm{Var}\left(\sum_{i=1}^{n} b_i Z_i \right) = \sum_{i=1}^{n} b_i^2, \\
& \mathrm{Cov}\left(\sum_{i=1}^{n} a_i Z_i, \sum_{i=1}^{n} b_i Z_i \right) = \sum_{i=1}^{n} a_i b_i = 0, \\
\implies & \left(\sum_{i=1}^{n} a_i Z_i, \sum_{i=1}^{n} b_i Z_i \right) \sim N_2 \left(0, 0, \sum_{i=1}^{n} a_i^2, \sum_{i=1}^{n} b_i^2, 0 \right) \\
\implies & \sum_{i=1}^{n} a_i Z_i \sim N\left(0, \sum_{i=1}^{n} a_i^2 \right) \;\text{ and }\; \sum_{i=1}^{n} b_i Z_i \sim N\left(0, \sum_{i=1}^{n} b_i^2 \right) \;\text{ are independent} \\
\implies & \frac{\sum_{i=1}^{n} a_i Z_i}{\sqrt{\sum_{i=1}^{n} a_i^2}} \sim N(0,1) \;\text{ and }\; \frac{\sum_{i=1}^{n} b_i Z_i}{\sqrt{\sum_{i=1}^{n} b_i^2}} \sim N(0,1) \;\text{ are independent.}
\end{aligned}
$$

(a)

$$
\begin{aligned}
& \frac{\sum_{i=1}^{n} a_i Z_i}{\sqrt{\sum_{i=1}^{n} a_i^2}} \sim N(0,1) \text{ and } \frac{\left(\sum_{i=1}^{n} b_i Z_i\right)^2}{\sum_{i=1}^{n} b_i^2} \sim \chi_1^2 \text{ are independent} \\
\implies & \frac{\sum_{i=1}^{n} a_i Z_i / \sqrt{\sum_{i=1}^{n} a_i^2}}{\sqrt{\frac{\left(\sum_{i=1}^{n} b_i Z_i\right)^2}{\sum_{i=1}^{n} b_i^2}}} \sim t_1, \;\; \text{that is, } Y_1 \sim t_1.
\end{aligned}
$$

(b) Since $t_1^2 \stackrel{d}{=} F_{1,1}$, the result follows on using (a).

(c) $F_{Y_3}(y) = P(Y_3 \le y) = P\left(\frac{Z_1}{Z_2} \le y\right)$, $y \in \mathbb{R}$, (since $Y_3 \stackrel{d}{=} t_1 \stackrel{d}{=} \frac{Z_1}{\sqrt{y}}$), where $Z \sim N(0,1)$ are independent. Clearly,

$$
\begin{aligned}
F_{Y_3}(y) &= P\left(\frac{Z_1}{|Z_2|} \le y, \; Z_2 > 0 \right) + P\left(-\frac{Z_1}{|Z_2|} \le y, \; Z_2 < 0 \right) \\
&= P\left(\frac{Z_1}{|Z_2|} \le y, \; Z_2 > 0 \right) + P\left(\frac{Z_1}{|Z_2|} \le y, \; Z_2 < 0 \right) \;\; ((Z_1, Z_2) \stackrel{d}{==} (-Z_1, Z_2)) \\
&= P\left(\frac{Z_1}{|Z_2|} \le y \right), \; \forall\, y \in \mathbb{R}
\end{aligned}
$$

$$
\implies Y_3 \stackrel{d}{=} \frac{Z_1}{|Z_2|} \sim t_1, \; (\text{by } (a)) \implies Y_3 \sim t_1.
$$