Lec-46

Recall: Infinite population, random sample of size n by iids $X_1, X_2, \dots, X_n$.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \rightarrow \text{sample mean}$$

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \rightarrow \text{sample variance}$$

Distribution: Sampling distribution (given by the common distⁿ. of the iids X_i's)

Thm: If $\bar{X}$ is the mean of a random sample of size n from a normal population with mean μ and the variance σ^2, its sampling distribution is a normal distribution with the mean μ and the variance $\frac{\sigma^2}{n}$.

In other words, $\dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ has the std. normal distribution.

Remark: The above is an important result. But the major difficulty in applying it is that in most realistic applications, the population standard deviation is unknown. This makes it necessary to replace σ with an estimate, usually the value of the sample standard deviation S.

$\therefore$ We end up studying the exact distribution of $\dfrac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ for random samples from normal population.

Th^m: If Y and Z are independent random variables, Y has a chi-square distribution with ν degrees of freedom, and Z has the std. normal distribution, then the distribution of

$$T = \frac{Z}{\sqrt{Y/\nu}}$$ is given by

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad \text{for } -\infty<t<\infty$$

and it is called the <u>t-distribution</u> with ν degrees of freedom.

Proof: $\because$ Y and Z are indep, their joint pdf is given by

$$f(y,z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,\frac{1}{\Gamma\left(\frac{\nu}{2}\right)2^{\nu/2}}\,y^{\nu/2-1}e^{-\frac{y}{2}}, \quad y>0, \; -\infty<z<\infty$$

Now, $t = \dfrac{z}{\sqrt{y/\nu}} \Rightarrow z = t\sqrt{y/\nu}$

$$\therefore \frac{dz}{dt} = \sqrt{y/\nu}$$

[**Recall:** (Thm (A) from Lec-24)

$X \to$ cont. ran. var., $Y = u(X)$

$f(x) \to$ pdf of x.

If $y = u(x) \to$ (1) Differentiable, (2) either increasing or decreasing $\forall$ values within the range of x

for which $f(x) \neq 0$.

then for those values of x, the eqn $y = u(x)$ can be uniquely solved for x to give $x = w(y)$ & for the corresponding values of y, the pdf of $Y = u(X)$ is:

$$g(y) = \begin{cases} f[w(y)]\,|w'(y)| & \text{if } u'(x) \neq 0 \\ 0 & \text{o.w.} \end{cases}$$

So here Z & $T = u(Z) = \dfrac{Z}{\sqrt{Y/\nu}}$

$\therefore$ Joint density of Y and T (by directly applying the above thm):

$$g(y,t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\nu/2)\, 2^{\nu/2}} e^{-\frac{t^2 y}{\nu 2}} y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}} \frac{y^{\frac{1}{2}}}{\sqrt{\nu}}$$

$$= \begin{cases} \dfrac{1}{\sqrt{2\pi\nu}} \dfrac{1}{\Gamma\left(\frac{\nu}{2}\right) 2^{\nu/2}} y^{\frac{\nu-1}{2}} e^{-\frac{y}{2}\left(1+\frac{t^2}{\nu}\right)}, & y > 0,\ -\infty < t < \infty \\ 0 & \text{o.w.} \end{cases}$$

Let $w = \dfrac{y}{2}\left(1 + \dfrac{t^2}{\nu}\right)$ & integrating we finally get:

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\;\Gamma\left(\frac{\nu}{2}\right)} \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad -\infty < t < \infty$$

→ Also known as Student-t distribution

(W. S. Gosset) Pen name "Student."

Main application of the t-distribution:

Thm: If $\bar{X}$ and S^2 are the mean and the variance of a random sample of size n from a normal population with the mean μ and the variance σ^2, then

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

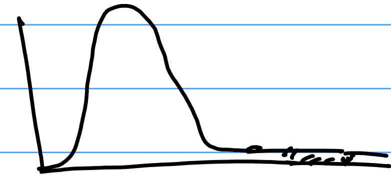
has the t-distribution with $(n-1)$ degrees of freedom.

Example ①: In 16 one-hour test runs, the gasoline consumption of an engine averaged 16.4 gallons with a std. deviation of 2.1 gallons. Test the claim that the average gasoline consumption of this engine is 12.0 gallons per hour.

Soln: Substituting $n=16$, $\mu = 12\cdot 0$, $\bar{x} = 16\cdot 4$, $s = 2\cdot 1$ in the above Thm, we get

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{16\cdot 4 - 12}{\frac{2\cdot 1}{\sqrt{16}}} = 8\cdot 38$$

"From the table" (which will be uploaded) we know that $\nu = 15$ the probability of getting a value of T greater than $2\cdot 947$ is $0\cdot 005$. So the probability of getting a value greater than 8 must be negligible. Thus it would seem reasonable to conclude that the true average hourly gasoline consumption of the engine exceeds $12\cdot 0$ gallon.



Another distribution that plays an important role in connection with sampling from normal populations is the F-distribution, named after Sir Ronald A. Fisher → one of the most prominent statisticians of the last century.

Thm: If U and V are independent random variables having chi-square distributions with ν_1 and ν_2 degrees of freedom, then

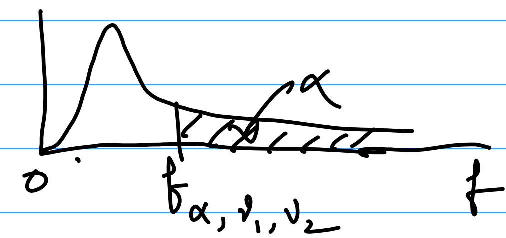
$$F = \frac{U/\nu_1}{V/\nu_2}$$

is a random variable having an <u>F-distribution</u>, that is, a random variable whose pdf is given by:

$$g(f) = \begin{cases} \dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\dfrac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} f^{\frac{\nu_1}{2}-1} \left(1+\dfrac{\nu_1}{\nu_2}f\right)^{-\frac{1}{2}(\nu_1+\nu_2)} & \text{for } f>0 \\ 0 & \text{o.w.} \end{cases}$$

Table: <u>Notation:</u> $f_{\alpha,\nu_1,\nu_2} :\Rightarrow$ Area to the right under the curve of the F distribution with ν_1 and ν_2 degrees of freedom is equal to α.

That is, f_{α,ν_1,ν_2} is s.t. $P(F \geq f_{\alpha,\nu_1,\nu_2}) = \alpha$



Applications:

If we are interested in comparing the variances σ_1^2 and σ_2^2 of two normal populations; for instance, in problems in which we want to estimate the ratio $\frac{\sigma_1^2}{\sigma_2^2}$ or perhaps to test whether

$\sigma_1^2 = \sigma_2^2$.

We base such inferences on <u>independent random samples</u> of size n_1 and n_2 from the two populations and thm (main application of χ^2-dist.) from Lec - Lec 9

(<u>Recall</u>: If $\bar{X}$ and S^2 of a ran. sample of size n,
mean — variance

from a normal population with mean μ & std. dev. σ, then ① $\bar{X}$ and S^2 are indep,
② $\frac{(n-1)S^2}{\sigma^2}$ has a χ^2-dist. with $(n-1)$ degrees of freedom).

according to which we have

$$\chi_1^2 = \frac{(n_1-1)\,S_1^2}{\sigma_1^2} \quad \text{and} \quad \chi_2^2 = \frac{(n_2-1)\,S_2^2}{\sigma_2^2}$$

are values of random variables having chi-square distributions with (n_1-1) and (n_2-1) degrees of freedom respectively.

By "independent random samples" we mean that the n_1+n_2 random variables constituting the two random samples are all independent, so that the two chi-square random variables are independent and the substitution for their values for U and V in the ~~...~~ Th^m for F-dist^n yields:

Th^m: If S_1^2 and S_2^2 are the variances of independent random samples of size n_1 and n_2 from normal populations with the variances σ_1^2 and σ_2^2, then

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

is a ran. var. having an F-dist^n. with (n_1-1) and (n_2-1) degrees of freedom.