Chapter 6

Important continuous population distributions

6.1 Chi-square distribution

Definition: If $Z_1, Z_2, \ldots, Z_\nu$ are ν independent standard Normal, $N(0,1)$, random variables

then

$$W = \sum_{i=1}^{\nu} Z_i^2 \sim \chi^2_\nu$$

has a Chi-square distribution with ν degrees of freedom.

Properties:

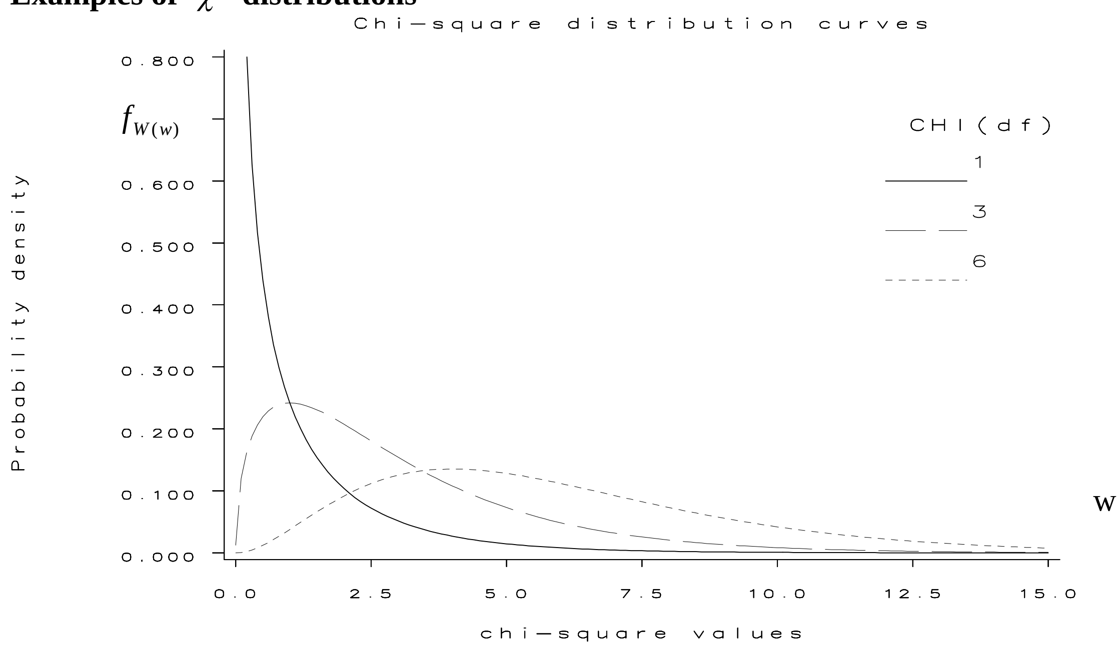
i) ν is the only parameter of the distribution

ii) range of W is $(0,\infty)$

iii) $E(W) = \nu$, $V(W) = 2\nu$ [Note variance is useful to remember later]

iv) If $W_1 \sim \chi^2_{\nu_1}$ and $W_2 \sim \chi^2_{\nu_2}$ and W_1 and W_2 are independent then $W_1 + W_2 \sim \chi^2_{\nu_1+\nu_2}$

Examples of χ^2 distributions



Theorem 1: Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample from a $N(\mu, \sigma^2)$ population then

$$W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{where} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 .$$

6.2 t distribution

Definition: If $Z \sim N(0,1)$ and $W \sim \chi^2_\nu$, and Z and W are independent then

$$T = \frac{Z}{\sqrt{\frac{W}{\nu}}} \sim t_\nu$$

has a t distribution with ν degrees of freedom.

Properties:

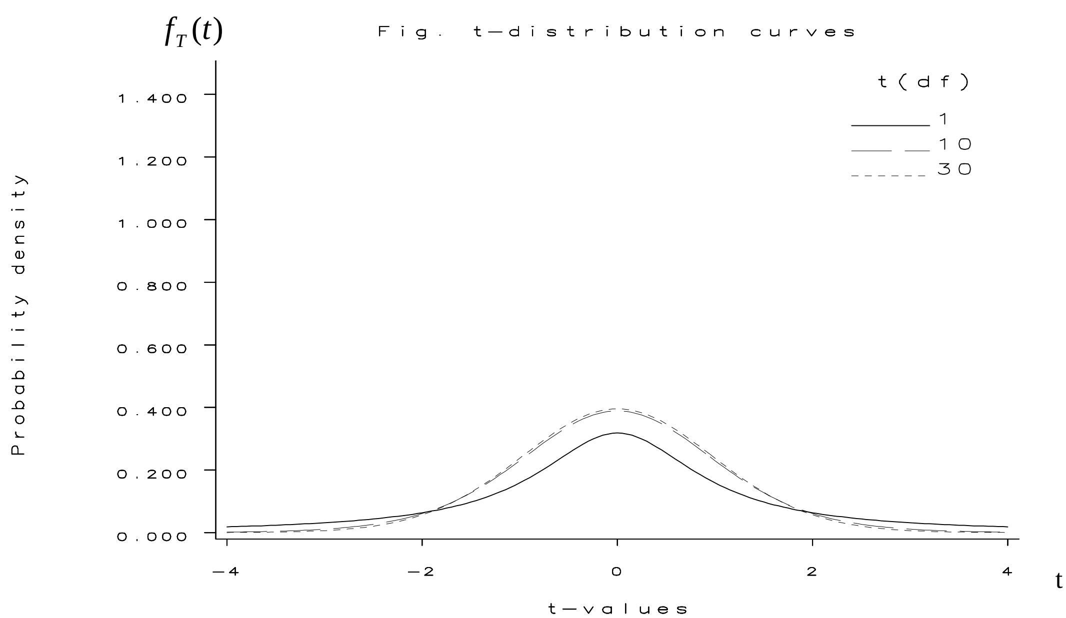
i) ν is the parameter of the distribution

i) range of T is (-∞,∞)

iii) distribution is symmetric about zero

iv) as $\nu \to \infty$ distribution of $T \to N(0,1)$

Examples of t-distributions:



Theorem 2: Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample from a population with $X \sim N(\mu, \sigma^2)$

then
$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$
where
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Proof:

We know that $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \Rightarrow Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$.

Also from the results on the χ^2 distribution we have that

$$W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \qquad \text{where} \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 .$$

Also $\bar{X}$ and S^2 are independent, so Z and W are independent, so

$$\therefore T = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2 (n-1)}}} \sim t_{n-1}$$
$$i.e.\ T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

Note: that since the t-statistic $T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ depends only on μ (and not on σ^2) we can use the above result for tests and C.I.'s for μ, when σ is unknown.

Exercise on χ^2 -*t* - Distributions

1. Suppose that X has a χ^2 distribution with 5 degrees of freedom.
   Using tables, determine the values x_1 and x_2 for which
   $$P(X < x_1) = P(X > x_2) = 0.025$$
   Show these values on a sketch of the probability density function.

2. A χ^2_2 distribution simplifies to an Exponential distribution with mean μ =2.
   Find the c.d.f. of an Ex(2) distribution and verify the values in the χ^2 table for ν=2.

3. Let X ~ t_{30}. Use tables to determine the values x_1 and x_2 such that
   $$P(-x_1 \le X \le x_1) = 0.90$$
   $$P(X > x_2) = 0.995$$
   Compare with the corresponding values from standardised normal distribution tables.

PRACTICAL 5 Continuous Distributions

Q1 The heights X of elderly women are stored in column C1 of ELDERLY.MTW
[From : Hand et al. 'Small data set']

a) Open MINITAB and input the saved worksheet ELDERLY.MTW by
   > File > Open Worksheet
   and finding the file J: \SCTMS \SOM \MA109 \PRACTICALS \ELDERLY.MTW

b) Obtain **descriptive statistics** for the variable X (height), including a graphical summary with a fitted Normal distribution by :

   > Stat > Basic Statistics > Display Descriptive Statistics

   | Variables **'C1'** |

   > Graphs | O Graphical Summary | OK > OK

c) Obtain a Normal **probability plot** to check whether X can be modelled using a Normal distribution.

\> Graph > Probability Plot

| Variables **'C1'** |

\> Distribution > Normal | OK > OK

Does a Normal distribution provide a suitable model for the distribution of X ?

d) Use the MINITAB commands for calculating pdf, cdf and inverse cdf values
[i.e **> Calc > Probability Distributions > Normal** etc]
to calculate the following for the Normal model for the variable X having the same mean and standard deviation as the sample :

i) $p(X \le 160)$, $p(X \le 170)$ and hence state $p(160 \le X \le 170)$

ii) quartiles Q_1 and Q_3 and hence state the Semi-Interquartile Range (SIR) for *X*.

iii) find x_1 and x_2 such that $p(X \le x_1) = 0.025$ and $p(X > x_2) = 0.025$ and hence state a 95% interval for *X*.

iv) find a 99% interval for *X*.

Q2 The diameter X of cherry trees was recorded and stored in column C1 of worksheet CHERRY.MTW. [From : Hand et al. 'Small data set'].

a) Open MINITAB and input the saved worksheet CHERRY.MTW by
\> File > Open Worksheet and finding the file CHERRY.MTW from J: \etc

b) Obtain **descriptive statistics** for the variable X (diameter), including a graphical summary with a fitted Normal distribution .

c) Obtain a Normal **probability plot** to check whether X can be modelled using a Normal distribution.
Does a Normal distribution provide a suitable model for the distribution of X ?

d) Use the **alternative probability plots** in MINITAB to see if you can find a suitable alternative model for the distribution of X, i.e. by > Graph > Probability Plot etc.

Q3 If $W \sim \chi^2_5$ use **> Calc > Probability Distributions > etc.** to find

i) $p(W > 6)$, $p(W > 10)$

ii) find w_1 and w_2 such that $p(W \le w_1) = 0.025$ and $p(W > w_2) = 0.025$ and hence state a 95% interval for *W*.
[Note in this example $w_1 = \chi^2_{5,0.975}$ and $w_2 = \chi^2_{5,0.025}$
where the standard notation is that if
$p(W > w) = \alpha$ where $W \sim \chi^2_\nu$ then $w = \chi^2_{\nu,\alpha}$]

Q4 If $T \sim t_{10}$ find

i) $p(X > 2)$, $p(X > 3)$

ii) find t such that $p(T > t) = 0.05$
[Note in this example $t = t_{10,0.05}$
where the standard notation is that if
$p(T > t) = \alpha$ where $t \sim t_\nu$ then $t = t_{\nu,\alpha}$]

Q5 To compare graphically the pdf of a standard Normal distribution N(0,1) with t distributions with $\nu = 1$ and 5 degrees of freedom.

i) Open MINITAB and type the following values into column 1 :
-3, -2.75, -2.5, -2.25, -2, -1.75, -1.5, -1.25, -1, -0.75, -0.5, -0.25,
0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3

ii) Calculate the corresponding of the pdf f(z) of Z~N(0,1) in column 2
[i.e use > **Calc** > **Probability Distributions** > **Normal** etc]

iii) Calculate the corresponding of the pdf f(t) of $t \sim t_1$ in column 3
[i.e use > **Calc** > **Probability Distributions** > **t distribution** etc]

iv) Calculate the corresponding of the pdf f(t) of $t \sim t_5$ in column 4
[i.e use > **Calc** > **Probability Distributions** > **t distribution** etc]

v) Graph C2-C4 against C1. Annotate the graph with a title, axis labels etc.

vi) Comment on the graphical comparison of the standard Normal with the t distribution