<u>The Chi-Square Goodness-of-Fit Test</u> [Navidi section 6.10; Devore chapter 14]

One simple test to determine whether or not a given random sample is consistent with some probability distribution is to compare the numbers of observations in each of k intervals with the numbers that would be expected if the probability distribution is correct.

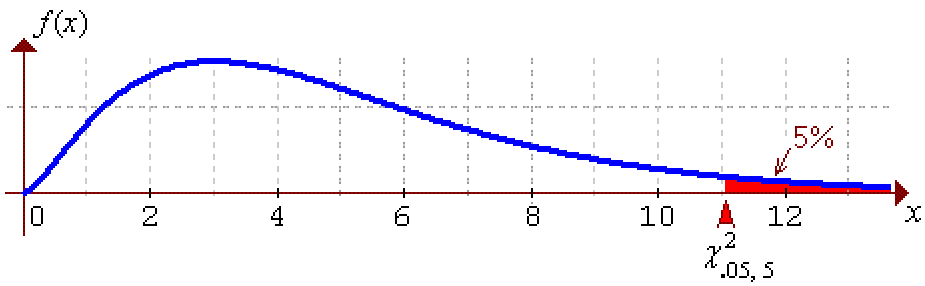
Let the observed values be $\{o_1, o_2, \ldots, o_k\}$.
The null hypothesis is $\mathcal{H}_o: \; p_1 = p_{1o}, \, p_2 = p_{2o}, \ldots, \, p_k = p_{ko}$, (that is, the probability distribution is correct), where p_i is the probability that the random quantity will fall in the i^{th} interval.
The number of values expected in the i^{th} interval when the null hypothesis is true is $e_i = n\,p_i$,
where $n = \sum_{i=1}^{k} e_i = \sum_{i=1}^{k} O_i$ is the total number of observations.
The chi-square test statistic is

$$\boxed{\chi^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i}}$$

Provided that all expected values are sufficiently large (about 5 or greater), this test statistic follows a chi-square distribution with (k – 1) degrees of freedom, to a good approximation.

Clearly, the closer all observed values are to the expected values, the lower the value of χ^2 will be. If the test statistic exceeds $\chi^2_{\alpha, k-1}$, (the value of the chi-square distribution with (k – 1) degrees of freedom, above which α of the probability lies), then there is sufficient evidence to reject the null hypothesis in favour of the alternative hypothesis that the sample did not come from the hypothesized probability distribution.

The probability density function for χ^2_5 is shown here:



Example 14.01

A standard six-sided die is rolled 60 times, with results as shown. Can one conclude, at a 5% level of significance, that the die is loaded (that is, the six faces are not all equally likely)?

Score:	1	2	3	4	5	6
Observed:	8	9	7	10	8	18

Score	o_i	e_i	$(o_i - e_i)$	$(o_i - e_i)^2$	$(o_i - e_i)^2 / e_i$
1	8				
2	9				
3	7				
4	10				
5	8				
6	18				

The test statistic is

$$\chi^2 = \sum_{i=1}^{6} \frac{(o_i - e_i)^2}{e_i} =$$

Compare the test statistic to the critical value

Note that the chi-square goodness-of-fit test is more complicated if the parameters of the probability distribution have to be estimated from the data. We shall not explore such a situation in this course.

Below is an extract from a chi-square table [Navidi Table A.6; Devore Table A.7].
The full table and a calculator for the chi-square distribution are available as Excel spreadsheet files at "`www.engr.mun.ca/~ggeorge/4421/demos/`".

	α				
ν	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**
1	2.70554	3.84146	5.02389	6.63490	7.87944
2	4.60517	5.99146	7.37776	9.21034	10.59663
3	6.25139	7.81473	9.34840	11.34487	12.83816
4	7.77944	9.48773	11.14329	13.27670	14.86026
5	9.23636	11.07050	12.83250	15.08627	16.74960
6	10.64464	12.59159	14.44938	16.81189	18.54758
7	12.01704	14.06714	16.01276	18.47531	20.27774
8	13.36157	15.50731	17.53455	20.09024	21.95495
9	14.68366	16.91898	19.02277	21.66599	23.58935
10	15.98718	18.30704	20.48318	23.20925	25.18818
11	17.27501	19.67514	21.92005	24.72497	26.75685
12	18.54935	21.02607	23.33666	26.21697	28.29952
13	19.81193	22.36203	24.73560	27.68825	29.81947
14	21.06414	23.68479	26.11895	29.14124	31.31935
15	22.30713	24.99579	27.48839	30.57791	32.80132
16	23.54183	26.29623	28.84535	31.99993	34.26719
17	24.76904	27.58711	30.19101	33.40866	35.71847
18	25.98942	28.86930	31.52638	34.80531	37.15645
19	27.20357	30.14353	32.85233	36.19087	38.58226
20	28.41198	31.41043	34.16961	37.56623	39.99685

The Chi-Square Test for Independence

Example 14.02

The lengths of cables from a production line are classified as too short, OK and too long.
The diameters of those cables are classified as too thin, OK and too thick.

Measurements of 300 cables produce the following table

		Diameter			
		Too thin	OK	Too thick	Total
Length	Too short	4	18	53	75
	OK	32	77	26	135
	Too long	64	20	6	90
	Total	100	115	85	300

Can one conclude, at a 5% level of significance, that the diameters are not independent of the lengths of the cables?

Let the observed value in row i, column j be labelled o_{ij}.

Let the sum of the values in row i be labelled $o_{i\bullet} = \sum_{j=1}^{3} o_{ij} = o_{i1} + o_{i2} + o_{i3}$

Let the sum of the values in column j be labelled $o_{\bullet j} = \sum_{i=1}^{3} o_{ij} = o_{1j} + o_{2j} + o_{3j}$

The grand total number of observations is

$$n = o_{\bullet\bullet} = \sum_{j=1}^{3} o_{\bullet j} = \sum_{i=1}^{3} o_{i\bullet} = \sum_{i=1}^{3}\sum_{j=1}^{3} o_{ij} = o_{11} + o_{12} + \ldots + o_{32} + o_{33}$$

Recall that if and only if two events A, B are independent, then $\mathrm{P}[AB] = \mathrm{P}[A] \times \mathrm{P}[B]$

The event “length too short” is independent of the event “diameter too thin” iff

$$\mathrm{P}[\text{"length too short"} \cap \text{"diameter too thin"}] = \mathrm{P}[\text{"length too short"}] \times \mathrm{P}[\text{"diameter too thin"}]$$

This leads to

$$\mathrm{E}[\#\ \text{"length too short"} \cap \text{"diameter too thin"}] = n \times \mathrm{P}[\text{"length too short"} \cap \text{"diameter too thin"}]$$

$$= n \times \frac{\#(\text{"length too short"})}{n} \times \frac{\#(\text{"diameter too thin"})}{n}$$

$$\Rightarrow \quad e_{11} = \frac{o_{1\bullet} \times o_{\bullet 1}}{n} = \frac{75 \times 100}{300} = 25$$

iff $O_{1\bullet} =$ “length too short” is independent of $O_{\bullet 1} =$ “diameter too thin”

Example 14.02 (continued)

The other expected values are generated in the same way:

$$\boxed{e_{ij} = \frac{o_{i\bullet} \times o_{\bullet j}}{o_{\bullet\bullet}}}$$

We are testing, at $\alpha = 5\,\%$,

$\mathcal{H}_o$: the diameter and length are independent

vs.

$\mathcal{H}_A$: the diameter and length are not independent

If $\mathcal{H}_o$ is true, then the expected values are:

		Diameter			
		Too thin	OK	Too thick	Total
Length	Too short	25	28.75	21.25	75
	OK	45	51.75	38.25	135
	Too long	30	34.50	25.5	90
	Total	100	115	85	300

The test statistic is

$$\chi^2 = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\left(o_{ij} - e_{ij}\right)^2}{e_{ij}} = \frac{(4-25)^2}{25} + \frac{(18-28.75)^2}{28.75} + \frac{(53-21.25)^2}{21.25} + \ldots + \frac{(6-25.50)^2}{25.50}$$

Note that one observed value is less than 5, but that doesn't matter. What does matter is: all nine expected values are greater than 5. Tedious calculation results in $\chi^2 = 148.6\ldots$

The number of degrees of freedom is

Compare the test statistic to the critical value

[End of Chapter 14]

[Space for Additional Notes]