Derived Distributions to Statistics

Aman Chawla
LIDS

Introduction

- Seen Applied Probability in course so far
- Next part of course: Statistics
- Today: A Bridge

Features of Bridge

- Chi-squared random variable
- ---- central role in ----
- Chi-Square statistical test

The Chi-Squared Random Variable

- Defined as the sum of the squares of n independent standard normal random variables.
- Derive the distribution for n=2 today.
- For general n,

$$f_{\chi^2_{(n)}}(z) = \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} z^{\frac{n-2}{2}} e^{-\frac{z}{2}} U(z)$$

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \qquad \Gamma(n) = (n-1)!$$

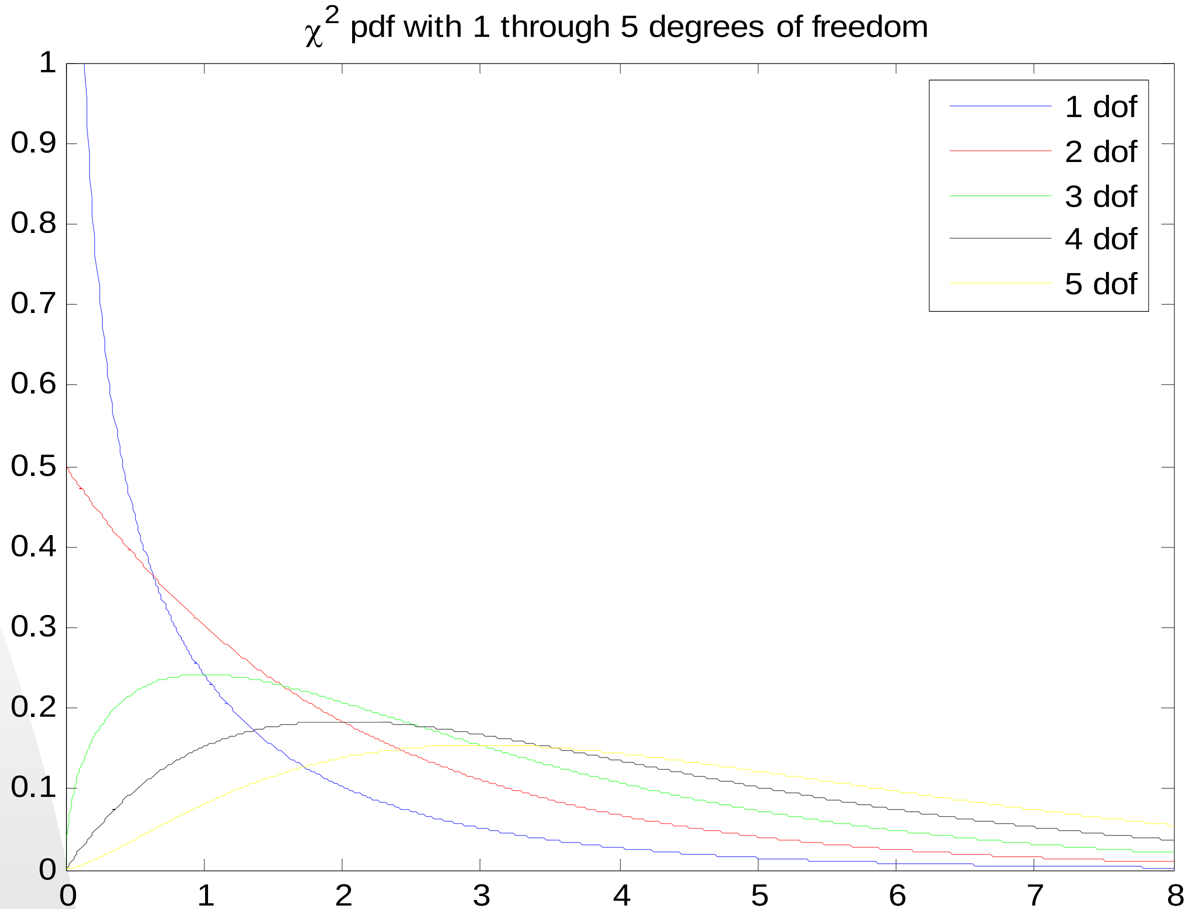

Figure 1: Probability Density Function of the Chi-Squared Distribution with 1 through 5 degrees of freedom.

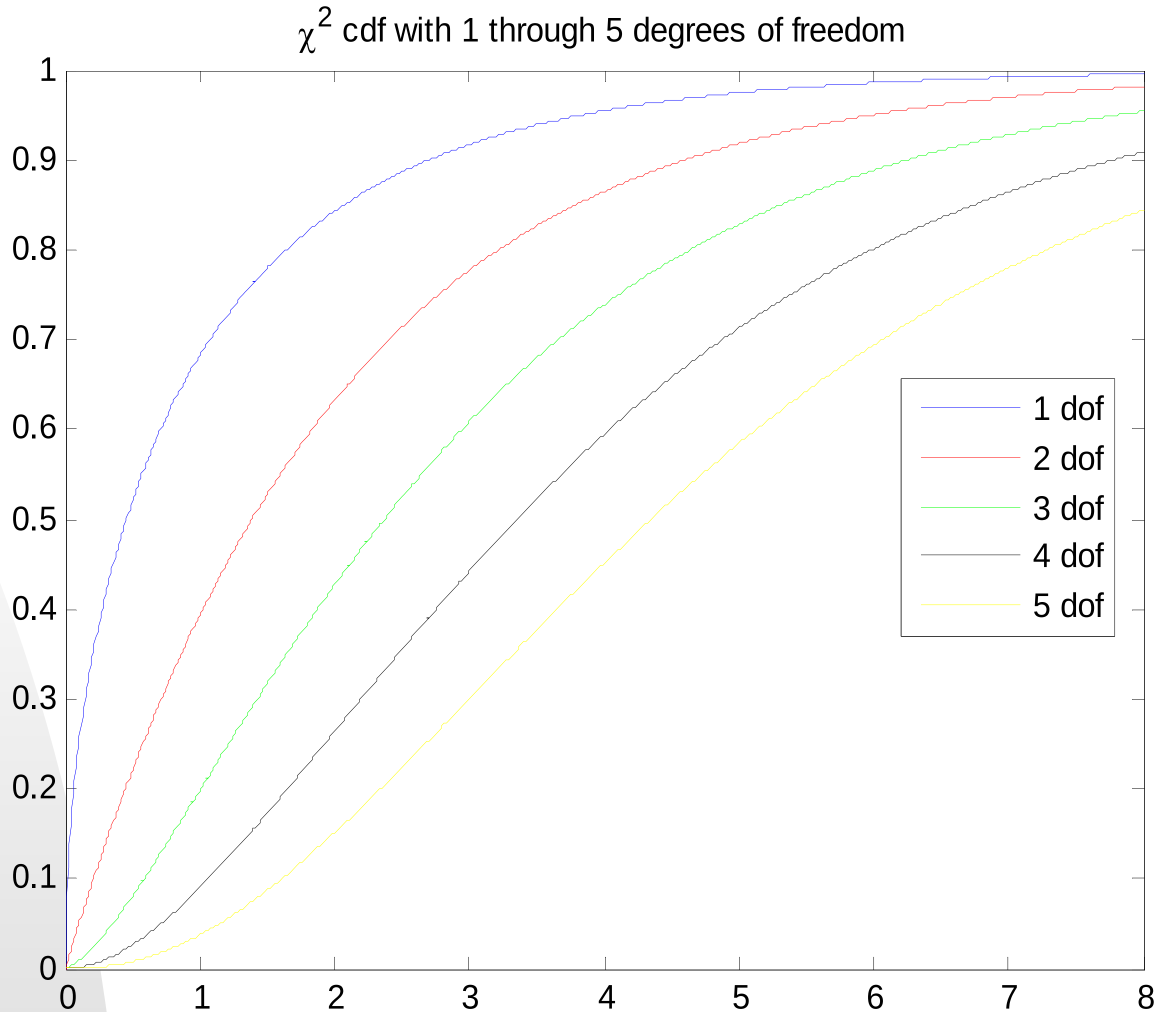

Figure 2: Cumulative Distribution Function of the Chi-Squared Distribution with 1 through 5 degrees of freedom.

Chi-Squared Test

- One of the most widely used statistical tests
- Derived in 1900 by Karl Pearson
- An Illustration, used by Pearson himself, serves well to elucidate.

Pearson’s Illustration

- 12 dice are thrown
- Number of dice that show up with a 5 or 6 is counted.
- This experiment is repeated a total of 26,306 times
- Motivation: Determine fairness of dice.
- Fair die has equal probability of landing on any one of its 6 faces.

The Data

No. of Dice with 5 or 6 points	Observed
0	185
1	1149
2	3265
3	5475
4	6114
5	5194
6	3067
7	1331
8	403
9	105
10	14
11	4
12	0
Total:	26306

Illustration (contd.)

- Under the fairness hypothesis, compute probabilities
- Pr(No die with 5 or 6 points in a throw of 12 dice) = (2/3)^12
- Pr(k dice with 5 or 6 points in a throw of 12 dice) =

$$C_k^{12} \times \left(\frac{1}{3}\right)^k \times \left(\frac{2}{3}\right)^{12-k}$$

Illustration (contd.)

- Expected number of trials yielding 0 dice with 5 or 6 points = 26306 * (2/3)^12 = 202.7495, etc.

Data

No. of Dice with 5 or 6 points	Observed	Expected	Deviation
0	185	203	-18
1	1149	1217	-68
2	3265	3345	-80
3	5475	5576	-101
4	6114	6273	-159
5	5194	5018	+176
6	3067	2927	+140
7	1331	1254	+77
8	403	392	+11
9	105	87	+18
10	14	13	+1
11	4	1	+3
12	0	0	0
Total:	26306	26306	

Illustration (contd.)

- Under fairness hypothesis, each observed value is a multinomial r.v.
- By Central Limit Theorem, since n=26306 is large, this can be thought of as a Normal random variable.
- Subtracting the expected value, squaring and dividing by the expected value gives a standard normal random variable

$$\chi^2 = \sum_k \frac{(observed - expected)^2}{expected} = 43.87241$$

Chi-test

- Thus Chi-statistic has Chi-squared distribution of order 12.
- Using knowledge of Chi-squared distribution, compute the Probability that a Chi-squared order 12 r.v. takes on a value greater than the observed value.
- If this Probability is ‘large’, it implies that the observed value of Chi-stat is typical
- Since the statistic measures the deviations between observed data and values expected under fairness hypothesis, this implies that the fairness hypothesis is not unwarranted.