Parametric tests

- Z- test, t test, F tests assumes that sample data comes from a normally distributed population
- They has fixed set of parameters(mean, standard deviation)
- Parametric tests infer about a population
- Parametric tests need parameters and knowlege about population distribution

Non - parametric tests

- **Non-parametric tests** makes no assumptions about a parametric distribution when modeling the data
- *Distribution-free* tests
- Non-parametric tests do not estimate population parameters.
- Tell the charateristic of entire distribution
- **non-parametric tests** statistic is defined to be a function on a sample. no dependency on a parameter insted we need entire distribution of data

Non - parametric tests

- The data that are not normally distributed, can be analyzed with non-parametric tests.
- Categorical data do not normally distributed so when we are dealing with categorical data we can use non parametric tests
- With categorical variables, you can’t calculate a mean or standard deviation. But measured as frequencies.
- Chi-squared test is non parametric test used to determine the relationship between categorical variables.

Chi-square Distribution

- If z_1, z_2,....,z_k are independent,standard nornal random variables, then the sum of their squares

$$Q = \sum_{i}^{k} Z_i^2$$

  is distributed according to the chi-square distribution with *k* degrees of freedom.

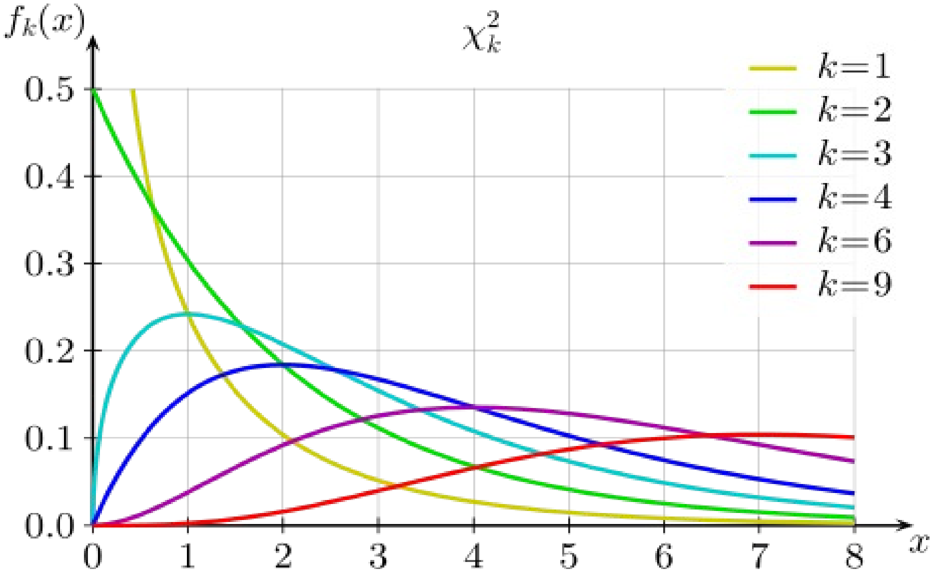
- Denoted as

$$Q \sim \chi^2(k)$$

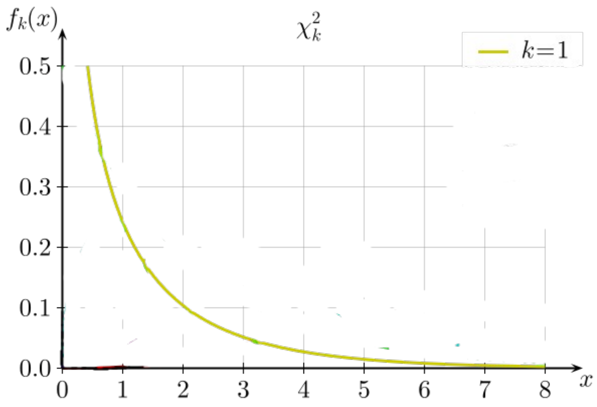
$$Q \sim \chi_k^2$$

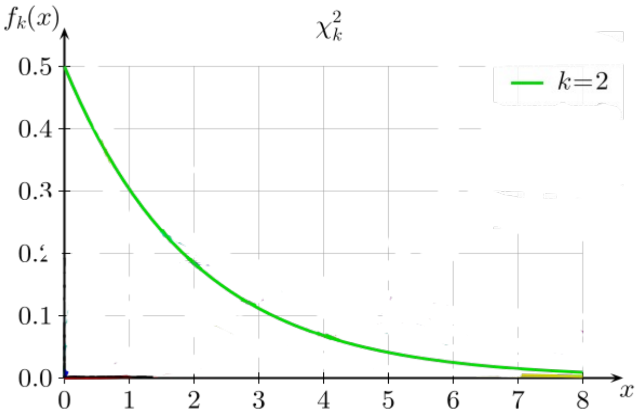
Chi-square Distribution

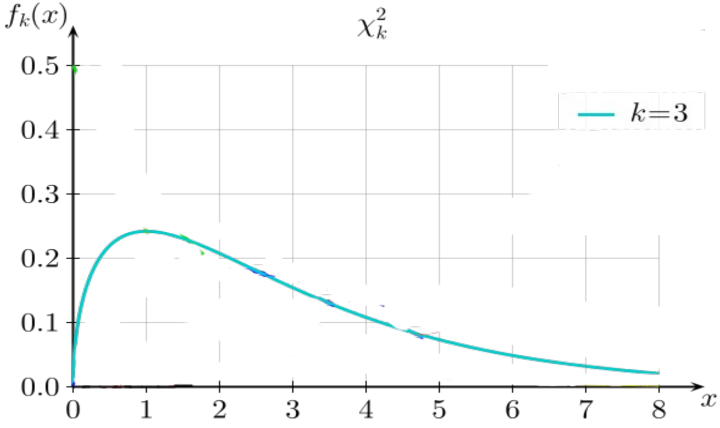
- Suppose that Z is a random variable sampled from the standard normal distribution
- Now consider the random variable $Q = Z^2$
- The distribution of the random variable Q is an example of a chi-square distribution

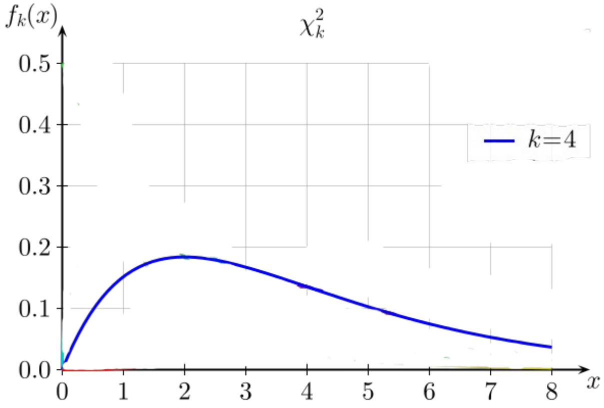


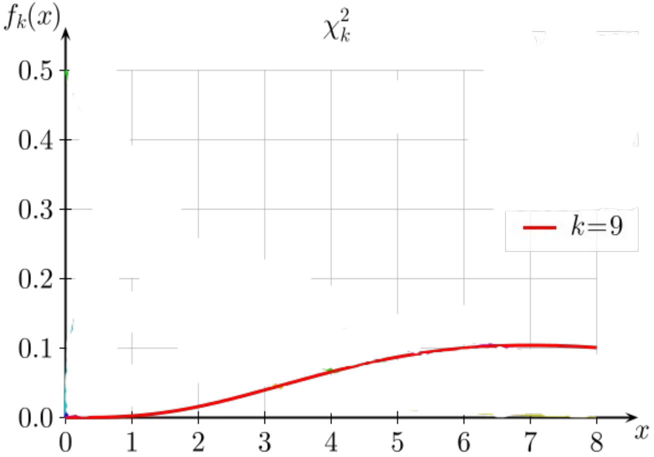
Chi-square Distribution

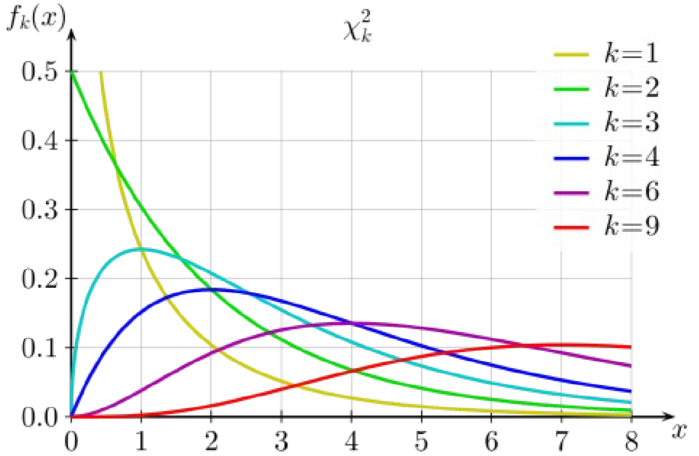




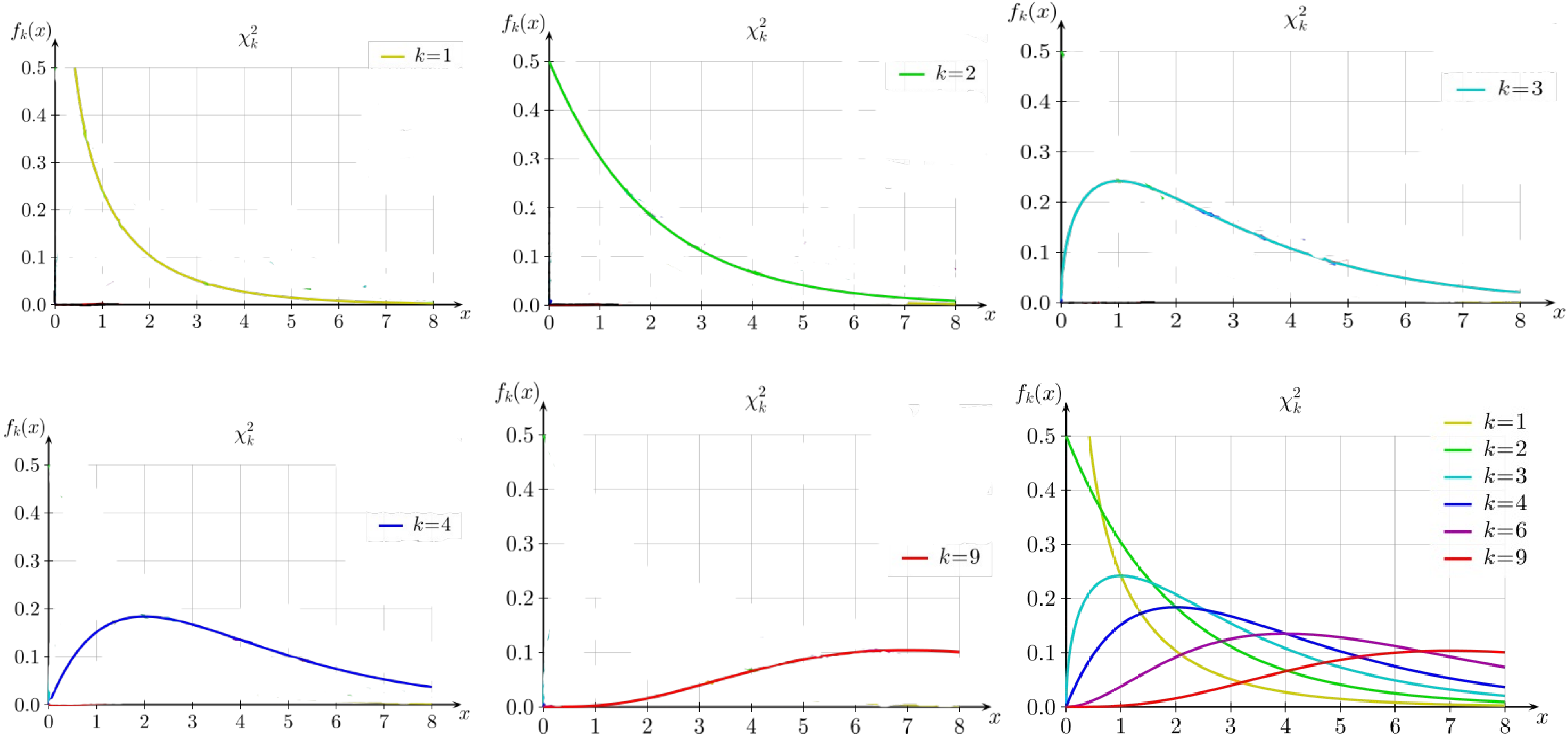




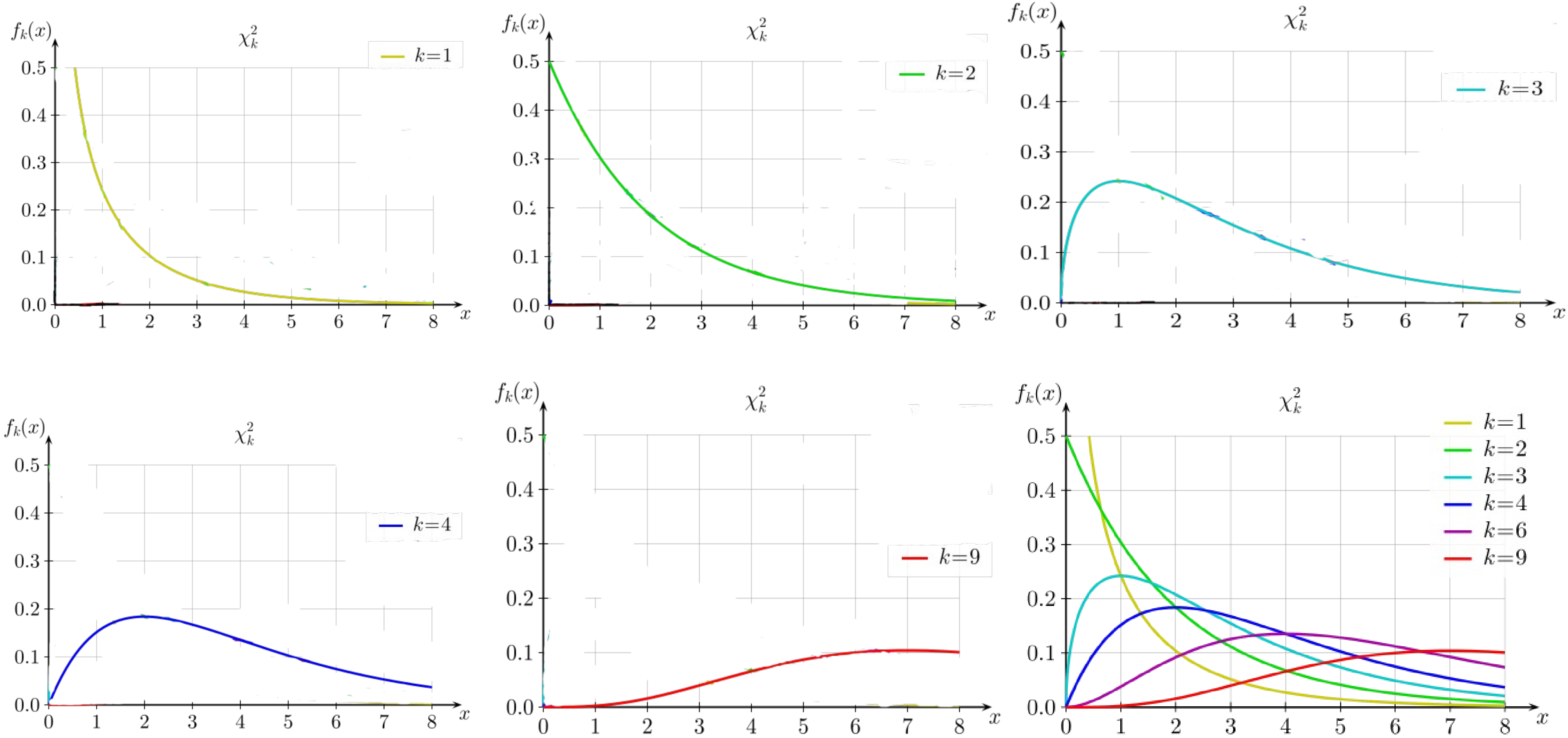




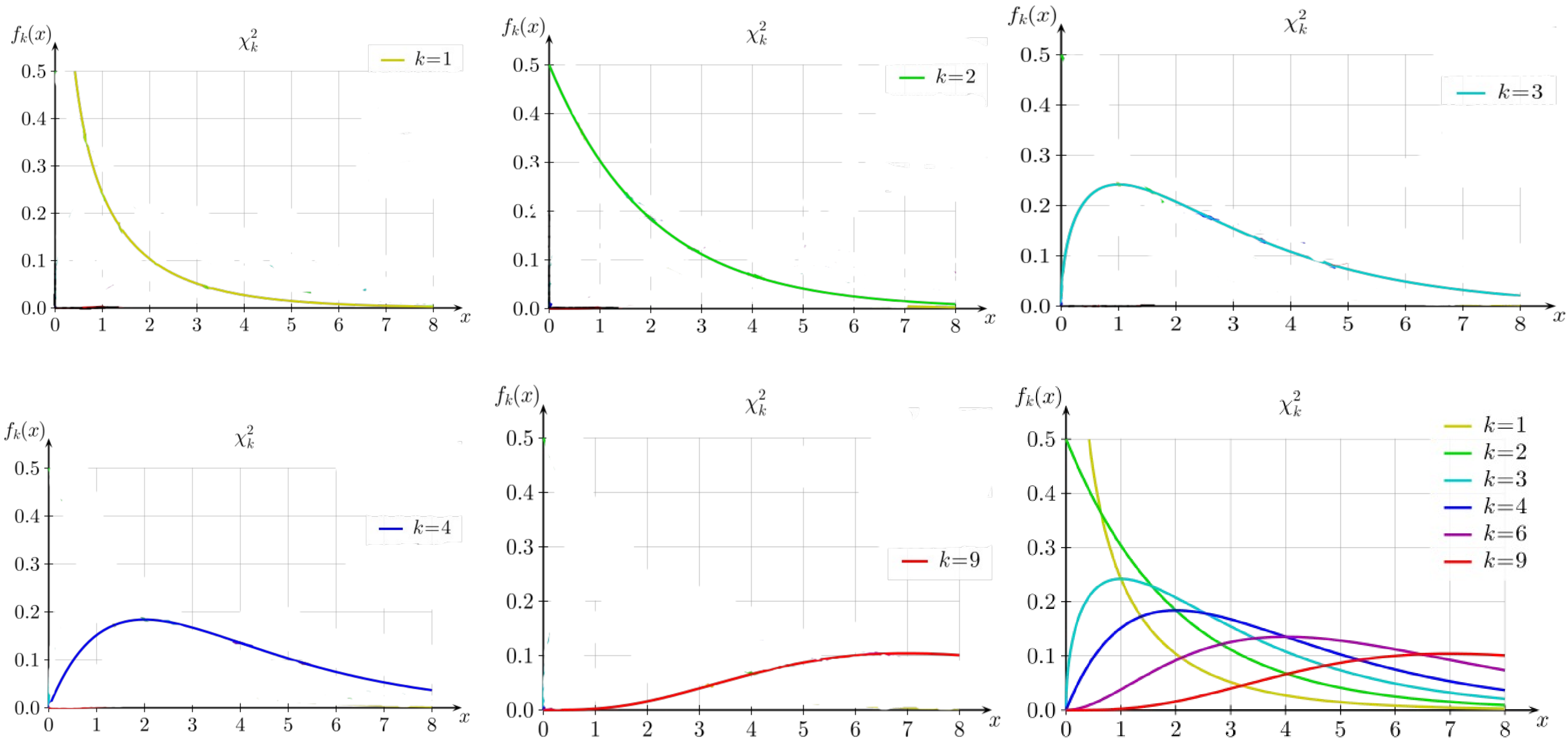
Chi-square Distribution



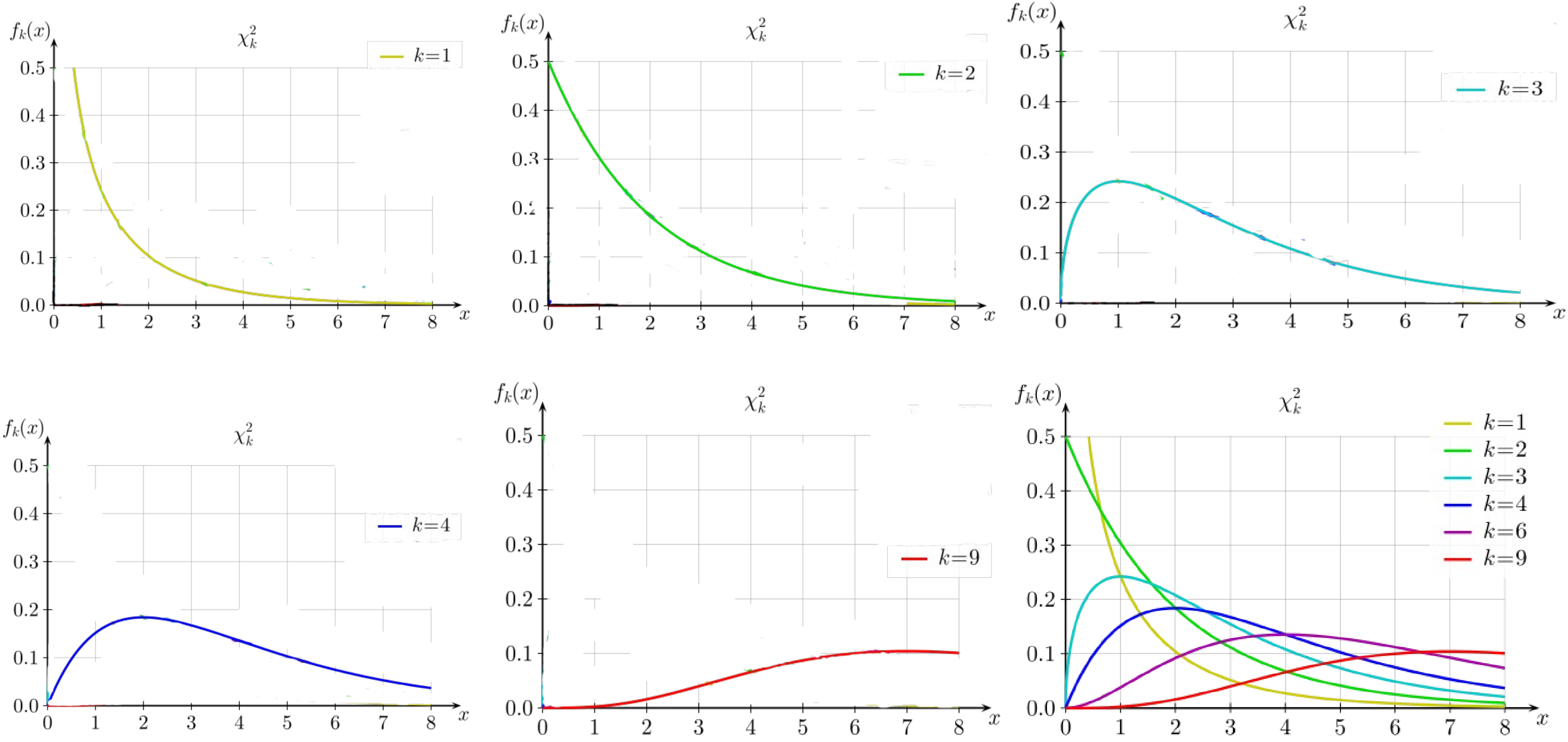
Chi-square Distribution



Chi-square Distribution



Chi-square Distribution



Chi-square Test

- It is used to determine the relationship between categorical variables
  - Chi-square Goodness fit test
  - Chi-squared test of independence
- Computation Procedure
  1. Calculate the chi-squared test statistic(χ^2)
  2. Determine the degrees of freedom(df), of that statistic
  3. Select a desired level of confidence
  4. Compare statistic(χ^2) to the critical value from the chi-squared distribution
  5. Sustain or reject the null hypothesis that the observed frequency distribution is the same as the theoretical distribution based on whether the test statistic exceeds the critical value

Goodness of fit test

- It is used when you want to determine whether the data follow a particular distribution.
- Goodness of fit test is a hypothesis test. It use chi square statistic to measured how the observed data is significantly different from the theoritical distribution
- The chi-square statistic is a measure of the deviation between observed and expected values
  - Very large deviations are unlikely. It is possibly not due to chance alone.
  - Small chi-square values (small deviations) are very likely and so they can be attributed to randomness.
- But how much deviation can be considered as random chance it can be decided based on critical values.

Chi square statistic

- It it is the sum of differences between observed and expected outcome frequencies (that is, counts of observations), each squared and divided by the expectation

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

- Where:

O_i = an observed count for bin i

E_i = an expected count for bin i

Example

- Test the hypothesis that a random sample of 100 people has been drawn from a population in which men and women are equal in frequency, the observed number of men and women would be compared to the theoretical frequencies of 50 men and 50 women. If there were 44 men in the sample and 56 women

Example

- Given

	Men	Women
Observed	44	56
Expected	50	50

- H_0 = men and women are chosen with equal probability in the sample
- H_1 = men and women are not chosen with equal probability in the sample

Example

- Chi squared statistic

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \frac{(44-50)^2}{50} + \frac{(56-50)^2}{50} = \frac{36}{50} + \frac{36}{50} = 0.72 + 0.72 = 1.44$$

- Df = 2-1 = 1
- $\alpha = 0.05$
- chi square critical value for df =1 and alpha 0.05 is 3.841
- chi square statistic 1.44 < 3.841(Critical value)
- so we can’t not reject the null hypothesis that the number of men in the population is the same as the number of women

Critical values of the Chi-square distribution with *d* degrees of freedom

	Probability of exceeding the critical value							
d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

***INTRODUCTION TO POPULATION GENETICS*, Table D.1**

Example

- A 6-sided die is thrown 60 times. The number of times it lands with 1, 2, 3, 4, 5 and 6 face up is 5, 8, 9, 8, 10 and 20, respectively. Is the die biased, according to the Pearson's chi-squared test at a significance level of 95% and/or 99%?

Example 2

- Given

	1	2	3	4	5	6
Observed	5	8	9	8	10	20
Expected	10	10	10	10	10	10

- H_0 = the die is unbiased, (i.e. each number is expected to occur the same number of times)
- H_1 = the die is biased

Example 2

- Chi squared statistic

i	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	5	10	−5	25	2.5
2	8	10	−2	4	0.4
3	9	10	−1	1	0.1
4	8	10	−2	4	0.4
5	10	10	0	0	0
6	20	10	10	100	10
				Sum	13.4

Example 2

- The number of degrees of freedom(df) = $n - 1 = 5$
- $\alpha = 0.05$
- chi square critical value for df =5 and alpha 0.05 is 11.070
- chi square statistic 13.4 > 11.070(Critical value)
- so we reject the null hypothesis and conclude that the die is biased at 95% confidence level
- At 99% confidence level, the critical value is 15.086.
- As the chi-squared statistic does not exceed it, we fail to reject the null hypothesis and thus conclude that there is insufficient evidence to show that the die is biased at 99% confidence level.