CLEMSON UNIVERSITY

Lecture 32

Statistical Inference for Multinomial Parameters

STAT 8010 Statistical Methods I
November 11, 2019

Whitney Huang
Clemson University

Binomial Experiments and Inference for Binary Category Data

- Fixed number of n trials, each trial is an independent event
- Binary outcomes (Success and Failure), where the probability of success, p, for each trial is constant
- The number of successes $X \sim \text{Bin}(n, p)$

> We use a random sample X to infer p, the population proportion

Multinomial Experiments and Inference for Multi-Category Data

- Fixed number of n trials, each trial is an independent event
- K possible outcomes, each with probability $p_k, k = 1, \cdots, K$ where $\sum_{k=1}^{K} p_k = 1$
- $(X_1, X_2, \cdots, X_K) \sim \text{Multi}(n, p_1, p_2, \cdots, p_K)$

> We use a random sample $(X_1, X_2, \cdots, X_K)$ to infer $\{p_k\}_{k=1}^{K}$, the event probabilities

Question: How many parameters here?

Example: Multinomial Probability

Suppose that in a three-way election for a large country, candidate 1 received 20% of the votes, candidate 2 received 35% of the votes, and candidate 3 received 45% of the votes. If ten voters are **selected randomly**, what is the probability that there will be exactly two supporter for candidate 1, three supporters for candidate 2 and five supporters for candidate 3 in the sample?

$$\mathbb{P}(X_1 = 2, X_2 = 3, X_3 = 5) = \frac{10!}{2!3!5!}(0.2)^2(0.35)^3(0.35)^5 \approx 0.08$$

Example: Inference for Multinomial Parameters

If we **randomly select** ten voters, two supporter for candidate 1, three supporters for candidate 2 and five supporters for candidate 3 in the sample. What would our best guess for the population proportion each candidate would received?

Pearson's χ^2 Test

- The Hypotheses:
  $H_0 : p_1 = p_{1,0}; p_2 = p_{2,0}; \cdots, p_K = p_{K,0}$
  H_a : At least one is different

- The Test Statistic:

$$\chi^2_* = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k},$$

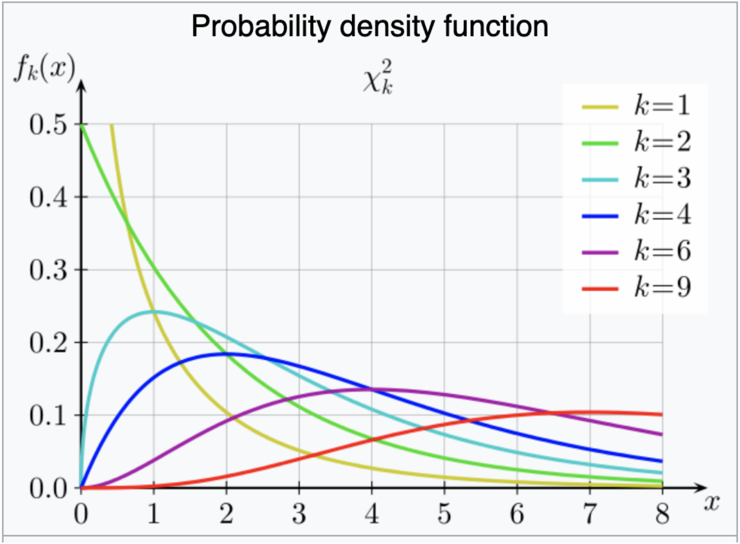
  where O_k is the observed frequency for the k_{th} event and E_k is the expected frequency under H_0

- The Null Distribution: $\chi^2_* \sim \chi^2_{df=K-1}$

- Assumption: $np_k > 5, k = 1, \cdots, K$

χ^2-Distribution

chi-square



Example: Testing Mendel's Theories (pp 22–23, "Categorical Data Analysis" 2_{nd} Ed by Alan Agresti)

"Among its many applications, Pearson's test was used in genetics to test Mendel's theories of natural inheritance. Mendel crossed pea plants of pure yellow strain (dominant strain) plants of pure green strain. He predicted that second generation hybrid seeds would be 75% yellow and 25% green. One experiment produced $n = 8023$ seeds, of which $X_1 = 6022$ were yellow and $X_2 = 2001$ were green."

Use Pearson's χ^2 test to assess Mendel's hypothesis.

Example

In Child Psychology, color preference by young children is used as an indicator of emotional state. In a study of 112 children, each was asked to choose "favorite" color from the 7 colors indicated below. Test if there is evidence of a preference at the 5% level.

Color	Blue	Red	Green	White	Purple	Black	Other
Frequency	13	14	8	17	25	15	20