Theoretical Statistics, STOR 655

Asymptotic Analysis of the T^2 and χ^2 Statistics

Andrew Nobel

February 2023

Limiting Distribution of Hotelling's T^2

Preliminaries

Fact: If $X \sim \mathcal{N}_d(\mu, \Sigma)$ with $\Sigma > 0$ then

$$W \;=\; (X - \mu)\Sigma^{-1}(X - \mu)^t \;\sim\; \chi^2_d$$

Definition: The sample covariance matrix of $X_1, \ldots, X_n \in \mathbb{R}^d$ is

$$S_n \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(X_i - \overline{X}_n)^t \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i X_i^t - (\overline{X}_n)(\overline{X}_n)^t$$

Fact: If $X_1, X_2, \ldots \in \mathbb{R}^d$ are iid with common variance matrix $\mathrm{Var}(X_i) = \Sigma$ then $S_n \to \Sigma$ wp1 as $n \to \infty$

Hotelling's T^2

Definition: Let $X_1, \ldots, X_n \in \mathbb{R}^d$ be iid, and let $\mu \in \mathbb{R}^d$ be a mean vector of interest. Hotelling's T^2 statistic is

$$T_n^2 \; = \; (n-1)(\overline{X}_n - \mu)^t \, S_n^{-1} \, (\overline{X}_n - \mu)$$

where S_n is the sample variance matrix based on $X_1, \ldots, X_n$

- Multivariate analog of the one-sample t-statistic
- Used for inference about the common mean of the X_i's

Fact: If $X_1, X_2, \ldots \in \mathbb{R}^d$ are iid with $\mathbb{E}X = \mu$ and $\mathrm{Var}(X_i) = \Sigma > 0$ then $T_n^2 \Rightarrow \chi_d^2$

Limiting Distribution of Pearson's χ^2

Review: Projection Matrices

Definition: $A \in \mathbb{R}^{d \times d}$ is a projection matrix if $A^2 = A$

Idea: Suppose $A^2 = A$, and let $V =$ span of columns of A

- Matrix A maps vector $u \in \mathbb{R}^d$ to vector $v = Au \in V$
- Matrix A leaves vectors in V unchanged: if $v \in V$ then $v = Au$ so

$$Av = A(Au) = A^2u = Au = v$$

- Matrix A projects $\mathbb{R}^d$ onto subspace V

Projections and χ^2

Fact: Let A be a projection matrix

1. All eigenvalues of A are 0 or 1
2. rank(A) = trace(A)
3. If A is symmetric then $Ax \perp (x - Ax)$ for every $x \in \mathbb{R}^d$

Fact: Let $X \sim \mathcal{N}_d(0, \Sigma)$. Then $X^t X \sim \chi^2_r$ iff Σ is a projection of rank r

Multinomial Experiment

Multinomial Experiment

- Sequence of n iid trials where each trial has one of d possible outcomes
- Let $p_k =$ probability of outcome k on any given trial
- Let $p = (p_1, \ldots, p_d)^t$ be pmf of trial outcomes
- Let $n_k =$ number of trials having outcome k. Thus $\sum_{k=1}^{d} n_k = n$

Definition: Multinomial(n, p) is the joint distribution of $(n_1, \ldots, n_d)$

$$P(x_1, \ldots, x_d) = \frac{n!}{x_1! \cdots x_d!} p_1^{x_1} \cdots p_d^{x_d}$$

Multinomial Goodness of Fit via the χ^2 Statistic

Inference problem

- Perform multinomial experiment. Observe counts $n_1, \ldots, n_d$
- Assess fit of $n_1, \ldots, n_d$ to a Multinomial(n, p) distribution, where p is a fixed pmf of interest

Note: Under the Multinomial(n, p) distribution, $\mathbb{E}(n_k) = np_k$

Definition: The χ^2 statistic is given by

$$\chi^2_n(n_1, \ldots, n_d) = \sum_{k=1}^{d} \frac{(n_k - np_k)^2}{np_k} = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Limiting Distribution of χ^2

Theorem: If $n_1, \ldots, n_d$ are obtained from the target Multinomial(n, p) distribution then $\chi^2_n \Rightarrow \chi^2_{d-1}$ as the number of trials n tends to infinity

Modified χ^2**.** Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be of the form $g(x) = (g_1(x_1), \ldots, g_d(x_d))^t$ where $g_k : \mathbb{R} \to \mathbb{R}$. Using the delta method, one can show

$$\chi^2_n(g) \; = \; n \sum_{k=1}^{d} \frac{(g_k(n_k/n) - g_k(p_k))^2}{p_k \, g_k'(p_k)^2} \; \Rightarrow \; \chi^2_{d-1}$$

Special case $g_k(x) = x^{1/2}$ for $1 \leq k \leq d$ gives Hellinger's χ^2

$$\chi^2(g) \; = \; 4n \sum_{k=1}^{d} (\sqrt{n_k/n} - \sqrt{p_k})^2$$