Chi Squared Distribution

Chi Squared Distribution

A random variable X has a χ^2 distribution with k degrees of freedom iff:

$$X = Z_1^2 + \cdots + Z_k^2$$

where $Z_i \sim \mathcal{N}(0, 1)$

Properties:

- $\mathbb{E}\left[X\right] = k$
- $\mathrm{Var}\left[X\right] = 2k$

Notation:

- Denoted as χ_k^2
- α th quantile:
    - $P(X > \chi_{k,\alpha}^2) = \alpha$

Goodness of Fit

Goodness of Fit Test

Suppose we have a discrete random variable X with pmf f that takes k values $\{1, 2, \ldots, k\}$. A **goodness of fit test** is a test to deremine whether $H_0 : f = f_0$ or $f_1 : f \neq f_0$

Suppose we observe $X_1, \ldots, X_n \sim f$

- Let O_j denote the number of j's we saw in the data
- If the null hypothesis is correct, we expect $O_j \approx E_j := nf_0(j)$
- We compute the following test statistic that aggregates how far all O_j are from their E_j:

$$T := \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}$$

Limit of Test Statistic T

$T \rightsquigarrow \chi_{k-1}^2$ as $n \to \infty$

The reason why we have $k - 1$ instead of k degrees of freedom, the O_j$'s sum to $n, which removes a degree

Using this statistic is known as **Pearson's Chi Squared Test for Goodness of Fit**