Recall Test 2 Wednesday

Discrete Random Variables + Generalities on Continuous Random Variables

+ Uniform Distribution

Review Monday

Statistics vs. Probability

Testing Hypotheses

Calculating Probabilities and Generalizations vs. Strategy
A TYPICAL EXAMPLE OF STATISTICS:

A COMPANY MAKES A NEW MEDICINE.

THIS WILL BE TESTED ON A SAMPLE OF PATIENTS.

EVALUATION OF OUTCOMES, VS. DISTRIBUTION

OF OUTCOMES WITHOUT MEDICATION.

NULL HYPOTHESIS

SUPPOSE WE CAN TEST THE PROBABILITY P

OF THE MEASURED OUTCOME UNDER THE NULL HYPOTHESIS.
Very general rule: statistically significant means $p < 0.05$.

If this happens, we say the drug had an effect. (Reject the null hypothesis).

If $p \geq 0.05$, we should not say anything.

Why? Suppose we follow this scientific method.
Bayesian Analysis:

Null Hypothesis

\[ p \]

Correct

\[ 1 - p \]

False

Probability of being wrong is

\[ (1 - p) < 0.05 < 0.05 \leq 1 \]
THE CHI-SQUARED TEST
\[ \chi^2 \]

A very simple test.

PEARSON TEST: DRAW N BALLS (WITH RETURN)

k COLORS, PROBABILITY OF

\[ \text{TH COLOR IS } p_i. \]
\[ p_1 + p_2 + \cdots + p_k = 1. \]

NULL HYPOTHESIS - NUMBERS OF INDIVIDUAL BALLS

FOLLOW THE DISTRIBUTION

FROM EQUAL OUTCOMES. (APPROXIMATELY NORMAL)
The $\chi^2$ Test: Suppose I got $a_i$: Balls of color $i$.

$$
\chi^2 = \frac{(a_1 - p_1 N)^2}{p_1 N} + \frac{(a_2 - p_2 N)^2}{p_2 N} + \cdots + \frac{(a_k - p_k N)^2}{p_k N}
$$

degrees of freedom: $k-1$.

Look at a $\chi^2$ table with $k-1$ degrees of freedom, how likely is it $\chi^2$? Compare with 0.05 ($\chi^2$ with $l$ degrees of freedom: sum of $l$ independent standard Gaussian variables).
Example: Suppose I throw a pair of coins 20 times.

The number of times I got:

- 2 heads: 2
- 1 head: 16
- 0 heads: 2

\[ \chi^2 = \frac{(2-5)^2}{5} + \frac{(16-10)^2}{10} + \frac{(2-5)^2}{5} = \frac{9}{5} + \frac{36}{10} + \frac{9}{5} \]

\[ = \frac{72}{10} = 7.2 \]
2 degrees of freedom:

<table>
<thead>
<tr>
<th>p</th>
<th>0.5</th>
<th>0.1</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^2$</td>
<td>1.386</td>
<td>4.605</td>
<td>5.991</td>
</tr>
</tbody>
</table>

We can disprove this is random

(it is statistically significant).

In the case of 1 degree of freedom, one can understand why $X^2$ is the square of a Gaussian:

$p = p_1, \quad p_1 = 1 - p, \quad \alpha = \alpha_1,$
\[ \chi^2 = \frac{(a - pN)^2}{pN} + \frac{(N - a - (1-p)N)^2}{(1-p)N} = \]

\[ = \frac{(a - pN)^2}{pN} + \frac{(1-p)^2}{pN} \]

\[ = \frac{(a - pN)^2}{Np(1-p)} \sim \text{the square of a standard Gaussian!} \]

NO HW TODAY, REVIEW MONDAY