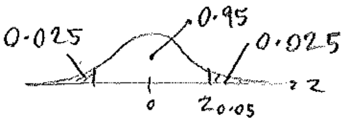
Barnett
5/23/13

MATH 56 WORKSHEET : statistical tests in expt. math

$z := \dfrac{S_N - \mu_N}{\sqrt{\mathrm{var}_N}}$



$z_{0.05} \approx 1.96$

A) 10^3 digits of a binary number have 530 ones & 470 zeros.
Do you reject the null hypothesis H_0 that the frequency of 1 and 0 are equal?

B) 10^6 digits have 530000 ones & 470000 zeros. (same ratio as above)
Do you reject H_0?

What p-value can you claim? Interpret it.

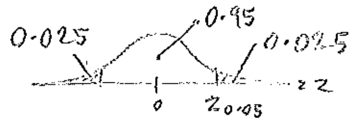
C) How could you generalize this idea of counting the frequencies of occurrence of some pattern to test whether the binary digits are <u>independent</u>? (ie each digit uncorrelated with the previous one(s))

~ SOLUTIONS ~

Barnett 5/23/13

MATH 56 WORKSHEET : statistical tests in expt. math

$$Z := \frac{S_N - \mu_N}{\sqrt{\mathrm{var}_N}}$$



$z_{0.05} \approx 1.96$

A) 10^3 digits of a binary number have 530 ones & 470 zeros.

Do you reject the null hypothesis H_0 that the frequency of 1 and 0 are equal?

$$Z = \frac{S_N - \frac{1}{2}N}{\sqrt{\frac{1}{4}N}} = \frac{530 - 500}{\sqrt{250}} = 1.89 < 1.96 \text{ so}$$

do not reject H_0; can't tell if biased.

B) 10^6 digits have 530000 ones & 470000 zeros. (Same ratio as above)

Do you reject H_0?

$$Z = \frac{530000 - 500000}{\sqrt{250000}} \approx \sqrt{10^3} \cdot 1.89 \approx 60 > 1.96 \text{ so,}$$

yes, it completely rules out H_0.

What p-value can you claim? Interpret it. $p_{z=60} = \mathrm{erfc}(z/\sqrt{2}) =$ zero in Matlab. ($< 10^{-308}$!)

C) How could you generalize this idea of counting the frequencies of occurrence of some pattern to test whether the binary digits are <u>independent</u>? (ie each digit uncorrelated with the previous one(s))

Eg. 10101010010101010101101010101101

has similar occurrence of 1 & 0 but is <u>not</u> uncorrelated, rather 0 follows 1 most often & 1 follows 0.

Test # occurrences of <u>pairs</u>:

$$\begin{cases} 00 & \leftarrow \text{infrequent} \\ 01 & \\ 10 & \text{seem to occur way more than 25\% of time.} \\ 11 & \leftarrow \text{infrequent} \end{cases}$$

Set $S_N = \sum_{j=1}^{N} s_j$ $\quad s_j = \begin{cases} 1 & \text{if "00"} \\ 0 & \text{otherwise} \end{cases}$

ie $S_N =$ # times "00" occurs

then $\mu = 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} = \frac{1}{4}$.

$\mathrm{var} = E[s_j^2] - \mu^2 = \frac{1}{4} - \frac{1}{16} = \frac{3}{16}$.

so $$Z = \frac{S_N - \frac{1}{4}N}{\sqrt{\frac{3}{16}N}}$$ is z-statistic, test as before.

Note this is equivalent to testing equidistribution in base 4 over $\{0,1,2,3\}$.