Assignment II.

second problem set for probability concepts

due: October 26, 2016

problem 1 PHYS 139/239

1. There are two caves, A and B. Cave A contains 7 red kangaroos and 2 blue kangaroos. Cave B contains 3 red kangaroos and 5 blue kangaroos. You pick a cave at random. One kangaroo, randomly selected, hops out of it.

1(a). What is the probability that the kangaroo that hops out is blue?

1(b). If it *is* blue, what is the probability that it came from Cave B?

problem 2 PHYS 139/239

1. You throw a pair of fair dice 10 times and, each time, you record the total number of spots. When you are done, what is the probability that exactly 5 of the 10 recorded totals are prime?

2. If you flip a fair coin one billion times, what is the probability that the number of heads is between 500010000 and 500020000, inclusive? (Give answer to 4 significant figures.)

problem 3 PHYS 139/239

(a) prove the additivity of the semi-invariant I_4 analytically and in simulation

(b) PHYS 239 only show the additivity of I_6 analytically and in simulation to reasonable accuracy for some distributions of your choosing

definition of the k_{th} centered moment M_k of a distribution:

$$M_k \equiv \left\langle (x_i - \overline{x})^k \right\rangle$$

following this definition M_2 is the variance of the distribution

Mean and variance are additive over independent random variables:

$$\overline{(x+y)} = \overline{x} + \overline{y} \qquad \mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(x)$$

note "bar" notation, equivalent to < >

Certain combinations of higher moments are also additive. These are called semi-invariants.

$$I_2 = M_2 \qquad I_3 = M_3 \qquad I_4 = M_4 - 3M_2^2$$

$$I_5 = M_5 - 10M_2M_3 \qquad I_6 = M_6 - 15M_2M_4 - 10M_3^2 + 30M_2^3$$

Skew and kurtosis are dimensionless combinations of semi-invariants

$$\mathrm{Skew}(x) = I_3/I_2^{3/2} \qquad \mathrm{Kurt}(x) = I_4/I_2^2$$

A Gaussian has all of its semi-invariants higher than I_2 equal to zero.
A Poisson distribution has all of its semi-invariants equal to its mean.

problem 4   PHYS 139/239

(a) show empirically the convergence to the central limit theorem in dice throwing simulations

(b) compare the result with your analytic expectation

problem 5 PHYS 139/239

calculate numerically the t-values and p-values in the table

Let’s dispose of the silly (all p’s = 0.25):

The test statistic: the value of the observed count under the null hypothesis that it is binomially (or equivalent normally) distributed with p=0.25.

$$\mu = 0.25\,N$$

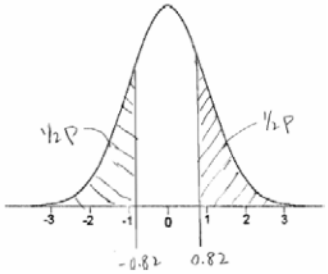
$$\sigma = \sqrt{0.25 \times 0.75\,N}$$

$$t = \frac{n-\mu}{\sigma}$$

$$p = 2[1 - P_{\text{Normal}}(|t|)]$$

t-value = number of standard deviations

p-value = tail probability (here, 2-tailed)



	t-value	p-value
A	174.965	≈ 0
C	–174.715	≈ 0
G	–170.963	≈ 0
T	170.713	≈ 0

The null hypothesis is (totally, infinitely, beyond any possibility of redemption!) ruled out.

problem 6 PHYS 239

(extra bonus for PHYS 139 solutions!)

explain and calculate numerically the two p-values of the hypothesis

The not-silly model: A and T occur with identical probabilities, as do C and G.

The test statistic: Difference between A and T (or C and G) counts under the null hypothesis that they have the same p, which we will estimate in the obvious way (which is actually an MLE).

$$\hat{p}_{AT} = \tfrac{1}{2}(n_A + n_T)/N$$
$$\hat{p}_{CG} = \tfrac{1}{2}(n_C + n_G)/N$$

$$n_A \sim \mathrm{Normal}(N\hat{p}_{AT}, \sqrt{N\hat{p}_{AT}(1-\hat{p}_{AT})})$$
$$n_T \sim \mathrm{Normal}(N\hat{p}_{AT}, \sqrt{N\hat{p}_{AT}(1-\hat{p}_{AT})})$$
$$\Rightarrow n_A - n_T \sim \mathrm{Normal}(0, \sqrt{2N\hat{p}_{AT}(1-\hat{p}_{AT})})$$

the difference of two Normals is itself Normal

the variance of the sum (or difference) is the sum of the variances

problem 7 PHYS 139/239

With p=0.3, and various values of n, how big is the largest discrepancy between the Binomial probability pdf and the approximating Normal pdf? At what value of n does this value become smaller than 10^{-15} ?