Phys 515 Assignment #1 (2019)

Due: October 3, 5:00 pm

Consider two observables modelled as correlated random variables as follows:

$$X = \mu_X + \sigma_A A$$
$$Y = \mu_Y + \sigma_A A + \sigma_B B$$

where μ_X, μ_Y, σ_A, and σ_B are constants, and A and B are independent random variables distributed according to the normal distribution (mean = 0 and standard deviation = 1).

1. Find the expectation values, variances, covariance, and correlation coefficient for X and Y in terms of the constants.
2. Write a method, `getSample(n)`, that returns a sample of `n` pairs of random numbers that follow this model, by using a library method that returns random numbers distributed according to the normal distribution.
3. Call the method to produce a sample of $n = 20$ "events", each event consisting of one pair of random numbers, x_i, y_i, with the following constants: $\mu_X = 5$, $\mu_Y = 4$, $\sigma_A = 3$, and $\sigma_B = 2$. Make a scatter plot showing the event sample. With the sample, calculate the following descriptive statistics: the sample means, the sample variances, the sample covariance, and the sample correlation coefficient. Compare these with expectations given in your answers to question 1.
4. Consider 1000 repititions of the experiment with 20 events in each sample. Show the distribution of the sample correlation coefficient (call this r) in a histogram from 0 to 1, with bin size 0.02. Find the mean and standard deviation (descriptive statistics) for these 1000 values of r.
5. Repeat the procedure outlined in question 4 to find the standard deviation of the r distribution when the sample size is 200.
6. The distribution for small sample sizes is not Gaussian. Consider the descriptive statistic
   $$z = \tanh^{-1}(r)$$
   where r is the sample correlation coefficient. It can be shown that the distrbution for that statistic is approximately Gaussian even with small sample sizes, with
   $$E[Z] = \tanh^{-1}(\rho) + \frac{\rho}{2(n-1)}$$
   $$V[Z] = \frac{1}{n-3}$$
   where ρ is the true correlation coefficient and n is the sample size. Show a histogram (with bins of size 0.1 from 0 to 2) of the distribution of z for 1000 samples of size 20, and overlay the Gaussian pdf with mean and variance as given by the equations above. Compare the sample mean z and sample z standard deviation with their expected values.