EP219: Data Analysis and Interpretation Tutorial Sheet 2, Fall 2018

Problem 1 Consider two correlated Gaussian random variables with 0 mean, x and y with covariance matrix given by,

$$C = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} 9 & -2 \\ -2 & 6 \end{pmatrix}.$$

a) Sketch the contours of the probability density function of x and y.
b) Write down the weight matrix V for these random variables.
c) Find linear combinations of x and y, let us call them x' and y', that are *uncorrelated*. Also, write down the covariance matrix and weight matrix of x' and y'.
d) Show that the random variable z defined as,

$$z \equiv \frac{1}{50}\left(6x^2 + 4xy + 9y^2\right)$$

is χ_f^2 distributed with $f = 2$.

Problem 2 Mass of the Higgs boson. Consider two measurements of the Higgs boson mass as reported by two independent experiments called ATLAS and CMS,

$$m_{\mathrm{H}}^{\mathrm{ATLAS}} = 125.36 \pm 0.37\ \mathrm{GeV/c^2},$$
$$m_{\mathrm{H}}^{\mathrm{CMS}} = 125.02 \pm 0.26\ \mathrm{GeV/c^2}.$$

Assuming the errors are purely statistical and the experiments are independent, combine the two measurements to report the averaged measurement of the Higgs mass and the error.

Problem 3 Consider an attempt to measure the acceleration due to gravity (g) by dropping a ball from a tower and measuring how long it takes to hit the ground. Assume that you have been given the height of the tower as 20.15 m and it has an uncertainty of 0.01 m. Let us assume that our timing measurement is taken by a digital stopwatch which rounds numbers to the nearest hundredth of a second. The reading on the clock when the ball hits the ground is 2.03 s.

a) What is the error on the measured time?
b) What is the value of g and error that is inferred?
c) If we perform N repeated trials, write down an expression for the error on the inferred value of g. What is the minimum error on g that can be obtained by repeated trials?

Problem 4 Consider two experiments (A, B) that attempt to measure the energies of the 1st excited state of a Hydrogen atom. Both experiments measure emission lines corresponding to the $2s^2 \to 1s^2$ transition energy ΔE_{21} and measure different values $\Delta E_A, \Delta E_B$. Let us assume that experiments A and B have different precisions and therefore report different errors δ_A, δ_B for measuring the transition energy. These results are then used by each experiment to reconstruct the energy of the $2s^2$ states by using $E_2 = \Delta E + E_1$. Both experiments use the *same* value of E_1, which they take from a data book which quotes an error δ_1. This leads to a correlation in their errors. Combine the results of the two measurement to get the best estimate of E_2 and its error.

Problem 5 Suppose that X is a discrete random variable which takes on values from $0, 1, 2, 3$, with probability distribution function which depends upon a parameter θ, where $0 \le \theta \le 1$. The pdf of X is given below:

X	0	1	2	3
$P(X)$	$\theta/3$	$2(1-\theta)/3$	$2\theta/3$	$(1-\theta)/3$

After drawing variables X from this distribution you get the values 1, 0, 3, 3, 1, 1, 3, 1, 1, 3, 1, 2, 1, 0, 1, 0, 3, 2, 1, 3.

a) Find the Likelihood function and log-Likelihood function of θ.
b) Find the Maximum Likelihood Estimate (MLE) of θ.
c) Find the 1-sigma interval of θ values around the maximum likelihood estimate. (Hint: for a large enough data set, you may assume the Likelihood function is nearly Gaussian near the maximum. Perform a Taylor expansion of the log-Likelihood function about the MLE to find the width of the Gaussian.)

Problem 6 Deal or no deal. Consider a game-show called "deal or no deal", where you have to choose between three suitcases labelled (A, B, C), one of which contains Rs. 1 crore and the others are empty. On this show contestants pick a suitcase (say suitcase A) and keep it temporarily without opening it. The host (who already knows what's in the suitcases) then opens up one of the other two cases (say suitcase B) which is empty. The contestant is now offered the option to keep the suitcase (A) or switch it with the remaining unopened suitcase (C). Suppose that you are on this show, would you switch or keep your suitcase to get the prize? First make a guess.

Now let us analyze this using likelihoods. Let H_i denote the hypothesis that the prize is in suitcase i where $i = A, B, C$.

a) What are the a-priori probabilities of H_i before you pick a case?

b) Calculate the likelihood of each hypothesis H_i given that the host opened suitcase B after you picked suitcase A.

c) Compute the a-posteriori probabilities of each H_i after the host has opened case B and shown that it was empty.

d) To get the prize should you (i) keep A, (ii) switch to C, or (iii) does it not make a difference? Justify.

Problem 7 Your friend has two decks of cards, one is complete and one of them is missing all the Aces (there are 4 Aces in a 52 card deck). He gives you one deck and you deal yourself 10 cards and you do not find an Ace. What is the likelihood that the deck you are holding does not have Aces? What is the likelihood that it is a full deck? Can you infer anything about the probability that you have a full deck of cards? (Be explicit about your reasoning.)