Assignment MPO624 Yueyang Lu

1. **When did Pascal and Laplace write their great works on probability? How about Bayes?**
   Pascal in 1654. Laplace issued his work *Théorie analytique des probabilités* in 1812. Bayes published *An Essay towards solving a Problem in the Doctrine of Chances* in 1763.

2. **What formula relates the probability of two events to the probability of each event, if they are independent?**
   $P(A \cap B) = P(A)P(B)$.

3. **What defines independence?**
   The occurrence of one does not affect the probability of occurrence of the other. If the equation above is satisfied, the two events are independent.

4. **In your own words, what is the frequentist interpretation of probability?**
   In this view, the probability is a limit to which the relative frequency of the occurrence of an event approaches.
   For example, in a country 55% of people are female. Suppose that we want to know the probability of the event A: 'a randomly selected person is a female'. The frequentist interpretation tells us that if we count the gender of a very large number of randomly selected person. The occurrence of A will be close to 55 out of 100. The probability of A is thus 0.55.

5. **In your own words, what is the Bayesian interpretation of probability?**
   The probability is seen as a quantification of the expectation of the phenomenon.

6. **In Bayesian statistics, what are prior and evidence and posterior? Write Bayes' theorem and define the symbols you use, perhaps with an example. How does it relate to the scientific method of a cycle from hypothesis to experiment and on to better hypothesis?**

   Prior is the initial belief of an event. Evidence is the result of experiments we did to test the initial belief. Posterior is the revised belief of 'prior' after having accounted for the evidence.

   Bayes' theorem is

   $$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)},$$

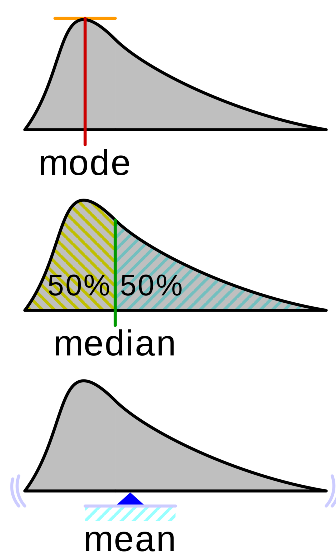
   where $P(A)$ is initial degree of belief in an event A, i.e., prior.
   $P(B)$ is the probability of the evidence B, which is related to how we perform the experiment.
   $P(A \mid B)$ is the revised probability of the proposition A after taking the evidence B into account.

   Suppose the prior probability of each hypothesis is $P(H)$. We perform experiments and obtain evidence E for or against each hypothesis. We can then calculate the posterior probability of each hypothesis given the evidence, $P(H \mid E)$. Here we used the data to alter our understanding of the probability of each of the possible hypotheses, from which we can choose a better one.

7. **What is a probability distribution? a probability density? (hint: discrete vs. continuous). What are the units of p(T), the probability density of temperature of some object at some time?**

A probability distribution is a mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment.
A probability density function is a function that quantify the probability of a continuous random variable falling within a particular infinitesimal range of values.
The unit is probability (dimensionless) per unit time, i.e., [1/s].

8. **Using the term likelihood for probability density, what is the maximum likelihood value of T if p(T) is not symmetric -- the mean, the median, or the mode? Sketch a non-symmetric distribution and indicate these 3 different measures of its central tendency.**
The maximum likelihood value of T in probability density is the *mode* because it is the value that appears most often. This can also be revealed in the schematic below that indicate the three measures in a PDF, where y-axis is pdf(T) and x-axis is T.



9. **Consider a uniform distribution over [1,2]. What is its first moment? What are its first four central moments?**
The first moment is the mean,

$$\bar{x} = \frac{1+2}{2} = 1.5.$$

The n-th central moment is

$$\mu_n = \int_1^2 (x-\bar{x})^n p(x)\, \mathrm{d}x = \int_1^2 (x-1.5)^n\, \mathrm{d}x,$$

which gives the first four central moments: 0, 0.0833, 0 and 0.0125, respectively.

10. **What fundamental mathematical operation (addition, subtraction, multiplication, division) creates the Normal distribution according to the Central Limit Theorem? Can you think of reasons it is so commonly observed? That is, can you name some natural processes that mimic that mathematical operation?**
    Addition.
    For example, the PDF of the value of velocity fluctuations in a homogeneous turbulence is approximately a normal distribution. The turbulent velocity at any location and instant is nearly random. The addition of the velocity occurs when you calculate the PDF.

11. **In your own words, what is a test statistic?**
    A test statistic is a statistic used to distinguish null and alternative hypothesis.

12. **What is the Z-test (based on the Z-statistic)? What are the one-tailed (or one-sided) p-values for Z values of 1, 2, 3? (sometimes called one-sigma, two-sigma, three-sigma events or excursions of a variable away from its mean).**
    A Z-test is any statistical test to the mean of a distribution in which we already know the population variance.
    The one-tailed p-values are 0.15866, 0.02275 and 0.00135, respectively.

13. **What is the t-test, based on the t-distribution? Find a table or Web page to answer the p-value question above, for a t-test with sample number N=10. How small is the difference from a Z-test?**
https://www.socscistatistics.com/pvalues/tdistribution.aspx
A t-test is any statistical test to the mean of a distribution in which we do NOT know the population variance.
The one-tailed p-values are 0.17172, 0.03829 and 0.00748, respectively, with the degree of freedom $df = 9$.
The relative difference from the p-value of Z-test is 8%, 70% and 400%, respectively. The difference will decrease as the degree of freedom increases.

14. **What is the chi-squared distribution? When might you use it in a statistical test? (hint: where in science do we see a sum of squares involving just two or three variables?)**
The chi-squared distribution is s the distribution of a sum of the squares of some independent normal random variables. It might be used in the study of eddy kinetic energy, $EKE = (\overline{u'^2} + \overline{v'^2})/2$.

15. **What is the F-test? Where have we seen a ratio of variances before? (hint: what was r-squared in linear regression)?**
An F-test any statistical test in which the test statistic has an F-distribution under the null hypothesis.
It may be used in the problems associated with linear regression, where the coefficient of determination is the proportion of the variance that is predictable
$$R^2 = \frac{\Sigma(f_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2}.$$

For example there are two regression models, both of which gives a fitted curve of the raw data. If we want to know whether one model gives a significantly better fit than the other, F-test is an approach.

16. **What are the joint PDF (probability density or distribution function)? marginal? conditional? Illustrate these 3 quantities with a sketch or annotation on an example like at https://en.wikipedia.org/wiki/Multivariate_normal_distribution.**
    The joint PDF for random variables X, Y,… is a function that

	A = Red	A = Blue
B = Red	4/9	2/9
B = Blue	2/9	1/9

quantifies the probability that each of X, Y,… falls in a particular infinitesimal range of values.
The marginal distribution gives the probabilities for any one of the variables X, Y,… with no reference to any specific ranges of values for the other variables.
The conditional probability distribution gives the probabilities for any subset of the variables conditional on particular values of the remaining variables.

Suppose each of two bags contains 2/3 red balls and 1/3 blue balls. Let A and B be the random variables associated with the outcomes of the draw from the first bag and second bag respectively. The joint probability distribution describes the likelihood of the four possible combinations of the two events: $AB, \bar{A}B, A\bar{B}, \bar{A}\bar{B},$ as shown in the table.

By definition, the marginal probability is 1) P(A) regardless of B, or 2) P(B) regardless of A. They can be easily derived from the joint probability: 2/3 and 1/3 respectively.