Distributions and Data

a. Mean and Variance,
b. The Binomial Distribution,
c. The Poisson Distribution,
d. The Luria-Delbrück Experiment,
e. The Chi-Squared Test of Goodness of Fit,
f. The Bhattacharyya Coefficient, and
g. Bayes' Theorem, Conditional Probability and the Law of Multiplication.

Please attach a cover page (-10%),
and submit any Mathematica computations in a .nb file.

Problem 1 (12%):
By using only the definition of the variance as a mean of squared differences from the mean:

$$V(X) = E[(X - \bar{X})^2], \qquad \bar{X} \equiv E(X),$$

and the following three theorems of the mean of a linearly translated variable, of a sum of variables and of a multiplication of independent variables:

1) $E(aX + b) = aE(X) + b$,
2) $E(X + Y) = E(X) + E(Y)$,
3) $E(XY) = E(X)E(Y)$ if X and Y are independent variables.

derive the variances of:

a) **(4%)** the sum of two independent variables X and Y –
$V(X + Y)$, and

b) **(4%)** the difference of two independent variables X and Y –
$V(X - Y)$.

c) **(4%)** Comment on the signal-to-noise ratios of the variables *X+Y* and *X-Y*, defined here as

$$\frac{[E(X+Y)]^2}{V(X+Y)} \text{ and } \frac{[E(X-Y)]^2}{V(X-Y)}.$$

Problem 2 (16% = 4 x 4% Each):
For Table 14 in Bulmer, Dover (1979), p. 95:

a) Suggest two different approaches to derive the Poisson distribution (that describes the observed numbers of particles) from the data. Do these two approaches give the exact same result? Comment. Plot the Poisson distributions of Problem 2(a) by using Mathematica.
b) Calculate the column of expected numbers based on each of these two approaches.
c) Calculate the p-value that corresponds to a chi-squared comparison of each of the expected distributions with the observed one. What would you conclude from the p-values you calculated? (Write down the expressions, also numerically, but evaluate in Mathematica).
d) What is the Bhattacharyya coefficient between the observed and expected distributions? What is the Bhattacharyya coefficient between the two expected distributions? What would you conclude from the coefficients you calculated?

Problem 3: (36% = 6 x 6% Each):
For Table 1 in Luria and Delbrück, *Genetics* (1943), p. 503:

a) Calculate the mean and the variance of the observations in each experiment (column) separately.
b) For each experiment, what is the Poisson distribution that describes it? Explain. Plot each Poisson distribution by using Mathematica.
c) What is the biological meaning of a different Poisson distribution describing each experiment? Explain.
d) Based on the Poisson distribution that describes each experiment, write down and evaluate in Mathematica the probabilities and the p-values to observe the largest and the smallest numbers in each experiment. Comment on your results.
e) Calculate the mean and the variance of the observations from all three experiments together.
f) Is it likely that a single Poisson distribution describes the observations from all three experiments? Explain. Assuming a single Poisson distribution, write it down, plot it and calculate the p-values to observe the largest and the smallest numbers in each experiment. Comment on your results.

Problem 4 (36% = 6 x 6% Each):
For Table 2 in Luria and Delbrück, *Genetics* (1943), p. 504:

a) Calculate the mean and variance of the observations in each experiment separately.
b) For each experiment, what is the Poisson distribution that describes it? Explain.
c) Based on the Poisson distribution that describes each experiment, write down and evaluate in Mathematica the probability and the p-value to observe the largest and the smallest numbers in each experiment. Comment on your results.
d) Compare your results in Problem 4(c) to your results in Problems 3 (d) and (f).
e) What is the biological meaning of your results?
f) What is the statistical meaning of your results?

Problem 5 (36% Extra Credit): Bayes' Theorem and Conditional Probabilities

a) **(6%)** Derive Bayes' theorem –

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

from the law of multiplication.

b) **(18%)** Assume a complete set of A_i mutually exclusive events, such that –

$$\sum_i P(A_i) = 1.$$

It follows that –

$$\sum_i P(A_i \cap B) = P(B).$$

Use this to derive the relation –

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_j P(B \mid A_j)P(A_j)}.$$

You are interested in the probability of a single throw of a die to give a number that is divisible by 2 AND 3.

c) **(3%)** How would your estimate of that probability change if you were told that the number is divisible by 2, but you were not told whether it is also divisible by 3? Explain.

d) **(3%)** How would your estimate of that probability change if you were told that the number is divisible by 3, but you were not told whether it is also divisible by 2? Compare to your answer to Problem 5(c) and comment.

You are interested in the probability of a single throw of a die to give a number that is divisible by 3 AND 6.

e) **(3%)** How would your estimate of that probability change if you were told that the number is divisible by 3, but you were not told whether it is also divisible by 6? Explain.

f) **(3%)** How would your estimate of that probability change if you were told that the number is divisible by 6, but you were not told whether it is also divisible by 3? Compare to your answers to Problems 5 (b)–(d) and comment.

Problem 6 (Extra Credit 64%): Random Walk

A "random walker" takes a step (of a fixed length L) to the left or the right of their position with equal probabilities. Write down the probability distribution of the person's position starting at $m=0$ after,

a) **(16%)** one single step, two steps, three steps, four steps;
b) **(8%)** n steps, where n is an odd number;
c) **(8%)** $2n$ steps.
d) **(8%)** What is the mean of each of these distributions in Problems 6 (a) and (b)?
e) **(8%)** What is the variance of each of these distributions?
f) **(16%)** You are asked to evaluate three "random walkers." You are told that two of them took two steps and one of them took four steps. What is the (**prior**) probability that one randomly selected walker is the one who took four steps? You find the randomly selected walker at position $m=2$. What is the (**posterior**) probability that this randomly selected walker is the one who took four steps? What is the (**posterior**) probability that this randomly selected walker is the one who took four steps if you find the walker at position $m=4$? Use the relation derived in Problem 5(b) to evaluate the posterior probabilities.