1. Answer the following questions based on the sessions you carried out in the lab.

(1 mark) (a) The likelihood function for a certain parameter θ is $(1/10)\sin(\theta)$ for θ in the range $[0, \pi]$. What is the maximum likelihood estimate for the parameter?

Solution: The function $(1/10)\sin(\theta)$ (on the interval $[0, \pi]$) takes its maximum at $\theta = \pi/2$.

(1 mark) (b) In the above situation, someone claims that $\theta = \pi/6$ is a theoretical estimate of the parameter. Can this claim be rejected on the basis of likelihood?

Solution: The likelihood ratio between the two parameter values $\theta = \pi/2$ and $\theta = \pi/6$ is $\sin(\pi/2)/\sin(\pi/6) = 2$. This is the likelihood ratio of predicting a single coin toss! So it is not enough to reject $\theta = \pi/6$.

(1 mark) (c) Under the null hypothesis, the distribution of an estimator $\hat{a}$ is normal with mean 0 and variance 1. The value of a calculated from the experiment is 0.5. Which of the following can be asserted with good confidence?

1. The null hypothesis can be accepted.
2. The null hypothesis can be rejected.
3. The null hypothesis cannot be accepted.
4. The null hypothesis cannot be rejected.

Solution: Since the point 0.5 lies well within the "bell" of the normal distribution, the null hypothesis cannot be rejected. The other options either cannot be determined by this experiment (since acceptance of a hypothesis is not the realm of statistics) or lack evidence.

(1 mark) (d) In summing an alternating series $\sum_{n=0}^{\infty}(-1)^n a_n$ (with $a_n \geq 0$), the partial sums obtained from the first six terms are as follows

$$1, 1/2, 3/4, 2/3, 5/7, 7/10$$

What is the smallest interval $[a, b]$ which contains the sum?

Solution: The series is alternating, so it is alternately above and below the limiting value. Thus the interval is $[7/10, 5/7]$.

(1 mark) (e) For which of the following two cases will the Poisson distribution give a better approximation of the answer?

1. The probability of getting a '6' exactly 2 times when a fair 6-sided die is thrown 10 times.

2. The probability of getting a pair of '6' exactly 2 times when two fair 6-sided dice are thrown simultaneously 60 times.

Solution: The expectation in both cases is $10/6 = 60/36$. In such a case, the Binomial distribution with larger number of Bernoulli trials will be closer to the Poisson distribution. Hence, it is the second one.

2. In each of the following questions write *only the formula* for the answer. Do not calculate.

(2 marks) (a) A fair tetrahedral die (4 sides numbered 1 to 4) is tossed repeatedly until we see a '1' appear exactly 7 times (not necessarily in sequence). What is the probability that the die is tossed exactly 19 times?

Solution: This is a case of the Negative Binomial distribution. However, we are counting the *number* of trials instead of failures. Hence, the probability is

$$\binom{19-1}{7-1}\frac{3^{19-7}}{4^{19}}$$

1 mark for using Negative Binomial distribution and 1 mark for the correct formula.

(2 marks) (b) What is the expected number of tosses and the variance in the number of tosses in the above experiment?

Solution: The number of expected tosses is the number of "successes" in the Bernoulli trial plus the total number of allowed "failures" (which is 7).

$$7 + \frac{7 \cdot (3/4)}{1/4} = 28$$

Alternatively, one can just write the formula

$$\sum_{r=7}^{\infty} r\binom{r-1}{7-1}\frac{3^{r-7}}{4^r}$$

The variance for the Negative Binomial distribution is:

$$\frac{7 \cdot (3/4)}{(1/4)^2} = 84$$

Alternatively, one can just write the formula

$$\sum_{r=7}^{\infty} r^2\binom{r-1}{7-1}\frac{3^{r-7}}{4^r} - \left(\sum_{r=7}^{\infty} r\binom{r-1}{7-1}\frac{3^{r-7}}{4^r}\right)^2$$

1 mark for mean and 1 mark for variance.

(2 marks) (c) A fair icosahedral die (20 sides numbered 1 to 4) is tossed 67 times. What is the probability that we see a '1' *at most* 3 times.

Solution: This is a case of the Binomial distribution.

$$\sum_{r=0}^{3} \binom{67}{r} \frac{19^{67-r}}{20^{67}}$$

1 mark for using Binomial distribution and 1 mark for the correct formula.

(2 marks) (d) What is the expected number of '1' seen and the variance in this number in the above experiment?

Solution: The expected number for the Binomial distribution is $67 \cdot (1/20)$ and the variance is $67 \cdot (1/20) \cdot (19/20)$. Alternatively, one can just write the formula for expectation

$$\sum_{r=0}^{\infty} r \binom{67}{r} \frac{19^r}{20^{67}}$$

And, one can just write the formula for variance

$$\sum_{r=0}^{\infty} r^2 \binom{67}{r} \frac{19^r}{20^{67}} - \left(\sum_{r=0}^{\infty} r \binom{67}{r} \frac{19^r}{20^{67}} \right)^2$$

1 mark for mean and 1 mark for variance.

(2 marks) (e) A Gieger counter clicks once every 27 seconds *on average*. What is the probability that we will not hear a click for 2 minutes?

Solution: This is a case of waiting time where the frequency $\lambda = 1/27$. Thus, the probability is:

$$\int_{120}^{\infty} \frac{\exp(-t/27)}{27} dt$$

or perhaps directly

$$\exp(-120/27)$$

1 mark for using Waiting time (exponential) distribution and 1 mark for the correct formula.

(2 marks) (f) What is the expected waiting time for a single click and the variance in this expected waiting time?

Solution: The expected waiting time is $1/\lambda$ which is 27 seconds, and the variance is $1/\lambda^2$ which is $(27)^2$. Alternatively, one can write the formula for expectation

$$\int_0^\infty (t/27)\exp(-t/27)dt$$

And, one can write the formula for variance

$$\int_0^\infty (t^2/27)\exp(-t/27)dt - \left(\int_0^\infty (t/27)\exp(-t/27)dt\right)^2$$

1 mark for mean and 1 mark for variance.

(2 marks) (g) An assembly line produces 1 defective product out of every 500 produced. What is the (approximate) probability that there are no defective products in a batch of 80?

Solution: This is a case where we can apply the Poisson distribution (as an approximation to Binomial)

$$\exp(-(80/500))$$

or directly apply the Binomial distribution.

$$\left(1 - \frac{1}{500}\right)^{80}$$

1 mark for using Poisson distribution (not for Binomial!) and 1 mark for the formula (either with Binomial or Poisson).

(2 marks) (h) What is the expected number of defective products and the variance in this number?

Solution: The expected number (from Poisson) is 80/500 and the variance is also 80/500. Alternatively, one can write the formula for the expectation

$$\sum_{r=0}^{\infty} r\frac{(80/500)^r}{r!}\exp(-80/500)$$

And, one can write the formula for the variance

$$\sum_{r=0}^{\infty} r^2\frac{(80/500)^r}{r!}\exp(-80/500) - \left(\sum_{r=0}^{\infty} r\frac{(80/500)^r}{r!}\exp(-80/500)\right)^2$$

Or, the expected number (from Binomial) is $80 \cdot (1/500)$ and the variance is $80 \cdot (1/500) \cdot (499/500)$. Alternatively, one can write the formula for the expectation

$$\sum_{r=0}^{80} r \binom{80}{r} \frac{499^{80-r}}{500^{80}}$$

And, one can write the formula for the variance

$$\sum_{r=0}^{80} r^2 \binom{80}{r} \frac{499^{80-r}}{500^{80}} - \left(\sum_{r=0}^{80} r \binom{80}{r} \frac{499^{80-r}}{500^{80}} \right)^2$$

1 mark for mean and 1 mark for variance.

(2 marks) (i) Let X be a normally distributed random variable with mean $1/2$ with a standard deviation of $1/2$. What is the probability that X lies in the interval $[3/2, 2]$?

Solution: The variable $Y = (X - 1/2)/(1/2)$ follows the standard normal distribution. When X varies over $[3/2, 2]$, the variable Y varies over $[2, 3]$. Thus, the required probability is

$$\frac{1}{\sqrt{2\pi}} \int_2^3 \exp(-t^2/2)dt$$

1 mark for calculating the correct range for the normal integral. 1 mark for the correct formula.

(2 marks) (j) Let $Z = h(X)$ where $h(x) = x$ if $x \geq 0$ and $h(x) = 0$ if $x < 0$. What is the expectation of Z?

Solution: We note that if $E(X) = \int tp(t)dt$, then $E(h(X)) = \int h(t)p(t)dt$. We have $X = (Y + 1)/2$. The condition $X \geq 0$ on Y becomes $Y \geq -1/2$. Thus we have the expectation of Z as

$$\frac{1}{\sqrt{2\pi}} \int_{-1/2}^{\infty} (t + 1)/2 \cdot \exp(-t^2/2)dt$$

1 mark for noting expectation of Z in terms of integral of $h(t)$. 1 mark for the correct formula.

3. *Calculate* the mathematical expectation for each of the following random variables.

(1 mark) (a) The random variable X which counts the number of tosses of '1' obtained in 37 tosses of a fair octahedral die (8 sides numbered 1 to 8).

Solution: The expectation of a Binomial random variable gives the answer 37/8.

(1 mark) (b) The random variable X^2 where X is as above.

Solution: The formula for the variance of the Binomial random variable gives $(37/8) \cdot (7/8)$. We need to add $(37/8)^2$ to this to get expectation of X^2. So the answer is $37 \cdot (7 + 37)/64$ which is $407/16$.

(1 mark) (c) The random variable Y that counts the number of (independent) Bernoulli trials needed to obtain 5 success, when the probability of success in a single trial is 1/7.

Solution: The expectation of the Negative Binomial random variable that counts the number of failures in order to obtain 5 successes is $5 \cdot (6/7)/(1/7) = 30$. Thus, the total number of trials is 35.

(1 mark) (d) The random variable W^2 where W is a normally distributed random variable with mean 1 and variance 2.

Solution: The formula $E(W^2) = \sigma^2(W) + E(W)^2$ gives $2 + 1^2 = 3$.

(1 mark) (e) The random variable Z^2 where Z is a Poisson random variable with expectation 2.

Solution: The formula $E(Z^2) = \sigma^2(Z) + E(Z)^2$ gives $2 + 2^2 = 6$. (Note that variance for the Poisson random variable is equal to the expectation.)

4. Answer the following questions about the laws of probability.

(1 mark) (a) If A and B are events, which is bigger $P(A)$ or $P(A \cap B)$?

Solution: Since $A \cap B \subset A$ we have $P(A \cap B) \leq P(A)$.

(1 mark) (b) If A and B are mutually exclusive events, what can you say about the probability $P(A \cap B)$.

Solution: Since A and B are mutually exclusive $P(A \cap B) = 0$.

(1 mark) (c) Give a condition on $P(A \cup B)$ which will ensure that A and B are exhaustive events.

Solution: Events A and B are exhaustive if and only if $P(A \cup B) = 1$.

(3 marks) (d) If $A_1, A_2, \ldots$ is a sequence of events such that $P(A_n) = 1/3^n$, which if the following statements is *definitely* true?

1. The probability $P(\cap_n A_n) = 0$.
2. The probability $P(\cup_n A_n) = 1/2$.
3. The probability $P(\cap_n A_n^c) \geq 1/2$.

Solution: The solution is (1) and (3). 1 mark for each and 1 mark for *omitting* (2).
Clearly, if $B_N = \cap_{n=1}^N A_n$, then we have $B_N \subset A_N$. Hence $P(B_N) \leq 1/3^N$. On the other hand B_N is a *decreasing* sequence of events and so by the law of probability $P(\cap_N B_N) = \inf P(B_N)$, which is 0. On the other hand $\cap_N B_N = \cap_n A_n$.
Now, *if* A_n are mutually exclusive then by the law of probability

$$P(\cup_n A_n) = \sum_n P(A_n) = (1/3)/(1 - 1/3) = 1/2$$

However, we don't know that A_n's are mutually exclusive. So we only get

$$P(\cup_n A_n) \leq \sum_n P(A_n) = (1/3)/(1 - 1/3) = 1/2$$

Finally,

$$\cap_n A_n^c = (\cup_n A_n)^c$$

So the last inequality follows from

$$P\left((\cup_n A_n)^c\right) = 1 - P\left(\cup_n A_n\right) \geq 1 - 1/2 = 1/2$$

(1 mark) (e) Given events A and B (with $P(B) \neq 0$), what is the relation between $P(A|B)$ and $P(A^c|B)$?

Solution: We have $P(A|B)P(B) + P(A^c|B)P(B) = P(A \cap B) + P(A^c \cap B) = P(B)$. Since $P(B) \neq 0$, we can divide by it to get $PA(|B) + P(A^c|B) = 1$.

(1 mark) (f) Given events A and B which is the relation between $P(A)$, $P(B)$, $P(A|B)$ and $P(B|A)$?

Solution: Since $P(A \cap B) = P(A|B)P(B)$, we get

$$P(A|B)P(B) = P(B|A)P(A)$$

(2 marks) (g) If X and Y are independent discrete random variables with positive integer values, write an expression for $P(XY = 6)$ in terms of the probabilities $P(X = a)$ and $P(Y = b)$ for various values of a and b.

Solution: We see that $XY = 6$ means (X, Y) lies in one of the following pairs $(1, 6)$, $(2, 3)$, $(3, 2)$, $(6, 1)$. Since X and Y are independent we have $P((X, Y) = (a, b)) = P(X = a)P(Y = b)$. So we get

$$P(XY = 6) = \\ P(X = 1)P(Y = 6) + P(X = 2)P(Y = 3) \\ + P(X = 3)P(Y = 2) + P(X = 6)P(Y = 1)$$

1 mark for using independence correctly. 1 mark for correct formula.

5. We obtain n independent samples X_i for $i = 1, \ldots, n$ from a continuous probability distribution with mean 1 and variance 1. Let Y denote the sample mean.

(2 marks) (a) When $n = 30$ what is the smallest interval $[a, b]$ so that $P(Y \in [a, b]) \geq 1/3$?

Solution: We note that $E(Y) = E(X_i) = 1$ and that $\sigma^2(Y) = \sigma^2(X_i)/n$. By Chebychyev's inequality, we have

$$P(|Y - 1| \geq c) \cdot c^2 \leq \sigma^2(Y) \text{ or } P(|Y - 1| \leq c) \geq 1 - \frac{\sigma^2(Y)}{c^2}$$

So we want $1 - 1/(30 \cdot c^2) \leq 1/3$. This gives $20c^2 \geq 1$. So $c \geq 1/\sqrt{20}$. Thus, the smallest interval

$$\left[1 - \frac{1}{\sqrt{20}}, 1 + \frac{1}{\sqrt{20}}\right]$$

1 mark for using Chebychyev's inequality. 1 mark for correct calculation.

(1 mark) (b) If we replace 30 by 60 by what factor does the length of the interval shrink?

Solution: When we replace n by $2n$, we see that $\sigma^2(Y)$ gets replaced by $\sigma^2(Y)/2$. By the above calculation we see that c gets replaced by $c/\sqrt{2}$.

(2 marks) (c) Assuming that the Central Limit Theorem is applicable for $n = 100$, write a formula for the probability that Y lies in the interval $[-1, 1]$.

Solution: By the Central limit theorem, Y can be assumed to follow the Normal distribution with mean 1 and variance $1/\sqrt{100} = 1/10$. Thus, the probability that it lies in $[-1, 1]$ is the same as the probability that $W = (Y - 1)/(1/10)$

lies in $[-20, 0]$

$$\frac{1}{\sqrt{2\pi}} \int_{-20}^{0} \exp(-t^2/2)dt$$

(Note that this is close to 1/2 since 20 is very large!).

1 mark for correction variance. 1 mark for correct formula (upto change of variables).

6. A fair dice (6 sides numbered 1 to 6) is tossed repeatedly.

(2 marks) (a) You have 5 rupees and have to give 1 rupee each time a number different from '6' shows. What is the expected number of times you have to throw the dice before your pocket is empty?

Solution: Let W be the number of non-sixes you throw on the dice before you through a '6' at least 5 times (not in sequence). Then W follows the Negative Binomial distribution

$$P(W = k) = \binom{k+5-1}{5-1} \frac{5^5}{6^{5+k}}$$

The total number of throws is $W + 5$. It has expectation

$$E(W + 5) = \frac{5 \cdot (1/6)}{5/6} + 5 = 6$$

1 mark for using Negative Binomial. 1 mark for correct expression for expectation.

(2 marks) (b) In addition to the above rule, you are *given* 5 rupees each time a '6' shows. What is the expected amount of money in your pocket after 20 throws? (Assume that you are allowed to keep a negative balance.)

Solution: Let X denote the number of '6' faces thrown in 20 throws. This is a Binomially distributed random variable

$$P(X = r) = \binom{20}{r} \frac{5^{20-r}}{6^{20}}$$

Since we are allowed to keep a negative balance, the money in your pocket is $5(X + 1) - (20 - X) = 6X - 15$ (which can be negative). Its expectation is

$$6 \cdot 20 \cdot (1/6) - 15 = 5$$

1 mark for using Binomial. 1 mark for correct expression for expectation.

(1 mark) (c) Now, assume that you are *not* allowed to keep a negative balance. (In other words, when your pocket is empty you have to quit.) Will the expected amount in the previous question increase or decrease?

Solution: Let Y be the random variable that counts the number of '6' throws during this new game.

There are two factors at play here:

1. In order to have a positive amount of money in the pocket, the first '6' must occur *on or before* the 5th throw, the second '6' must occur *on or before* the 11th throw and the third '6' must occur *on or before* the 17th throw. It follows that (with X as in the previous game)

$$P(Y = r) < P(X = r) \text{ for } r \geq 3$$

   This *decreases* the contribution to expectation from terms with $r \geq 3$.

2. If $Y \leq 2$, then the pocket will contain 0 instead of $6Y - 15$ (which is less than 0). This *increases* the contribution to expectation from terms with $r \leq 2$.

Thus the amount could increase or decrease. However, it actually stays the *same* as seen below! (I got this solution only *after* the examination!)

Let us assume that the dice are thrown whether or not the player is in the game. All that the condition says is that if the player has no money in the pocket then the throw of dice no longer leads to win or loss for the player! We can formulate this as follows.

Let W_i be the random variable indicating the amount of money won (negative indicating money lost) by the player on the i-th throw. The amount of money in the player's pocket *after* the i-th throw is $X_i = 5 + W_1 + \cdots + W_i$; as a result of the condition of game (3) we have $X_i \geq 0$. The condition of the game (3) can be stated as $W_{i+1} = 0$ given $X_i = 0$. Stated in terms of probability $P(W_{i+1} = 0|X_i = 0) = 1$. It follows that $E(W_{i+1}|X_i = 0) = 0$. On the other hand, if $X_i > 0$, then the $i + 1$-th dice throw *does* result in win or loss for the player. However, the expected win amount is also 0 in that case! In other words $E(W_{i+1}|X_i = r) = 0$ for $r > 0$ as well as for $r = 0$! In other words $E(W_i) = 0$ without any condition! It follows easily that $E(X_{20}) = 5$.

A fancy way of saying this is to say that X_i is a Martingale.

1 mark will be given if any of the reasons above is stated.