Problem 1 (10 pts): You have two opponents with whom you alternate play. Whenever you play A, you win with probability p_A; when ever you play B, you win with probability p_B, where $p_B > p_A$. If your objective is to minimize the number of games you need to play in order to win two in a row, should you start with A or B?

Hint: Let $\mathbf{E}N_i$ denote the mean number of games needed if you initially play i. Derive an expression for $\mathbf{E}N_A$ that involves $\mathbf{E}N_B$; write down the equivalent expression for $\mathbf{E}N_B$ and then subtract.

Solution: Consider the following 'states': $\{P_i\}$ where P is the player you are currently playing, either A or B, and i is the number of consecutive wins you have. For example A_1 would mean that you are playing A and have just beat B. Following the rules of the game, we can apply the law of total expectation a couple times to get,

$$\begin{aligned}\mathbf{E}N_A &= \mathbf{E}[N_A|A_0 \to B_1]\mathbf{P}(A_0 \to B_1) + \mathbf{E}[N_A|A_0 \to B_0]\mathbf{P}(A_0 \to B_0)\\ &= \left(\mathbf{E}[N_A|A_0 \to B_1 \to A_2]\mathbf{P}(B_1 \to A_2) + \mathbf{E}[N_A|A_0 \to B_1 \to A_0]\mathbf{P}(B_1 \to A_0)\right)\mathbf{P}(A_0 \to B_1)\\ &\quad + \mathbf{E}[N_A|A_0 \to B_0]\mathbf{P}(A_0 \to B_0).\end{aligned}$$

Note that $\mathbf{E}[N_A|A_0 \to B_0] = 1 + \mathbf{E}N_B$, $\mathbf{E}[N_A|A_0 \to B_1 \to A_2] = 2$, and $\mathbf{E}[N_A|A_0 \to B_1 \to A_0] = 2 + \mathbf{E}N_A$. Using this, the above reduces to,

$$\mathbf{E}N_A = \left(2p_B + (2 + \mathbf{E}N_A)(1 - p_B)\right)p_A + (1 + \mathbf{E}N_B)(1 - p_A).$$

Simplification yields,

$$\mathbf{E}N_A = 1 + p_A + p_A(1 - P_B)\mathbf{E}N_A + (1 - p_A)\mathbf{E}N_B.$$

Similarly,

$$\mathbf{E}N_B = 1 + p_B + p_B(1 - P_A)\mathbf{E}N_B + (1 - p_B)\mathbf{E}N_A.$$

Finally,

$$\mathbf{E}[N_B - N_A] = p_B - p_A + \mathbf{E}N_A\left((1 - p_B) - p_A(1 - p_B)\right) + \mathbf{E}N_B\left(p_B(1 - p_A) - (1 - p_A)\right)$$

which, with rearrangement, gives,

$$\mathbf{E}[N_B - N_A] = \frac{p_B - p_A}{1 + (1 - p_A)(1 - P_B)} > 0.$$

Thus, we have $\mathbf{E}N_B > \mathbf{E}N_A$. This tells us that the expected number of games played starting with player A is smaller than for starting with B. Therefore we should start with player A.

Problem 2 (10 pts): A set of n dice is thrown. All those that land on six are put aside, and the others are again thrown. This is repeated until all the dice have landed on a six. Let N_n denote the number of throws needed. (For instance, suppose that $n = 3$ and that on the initial throw exactly two of the dice land on six. Then the other die will be thrown, and if it lands on a six, the $N_3 = 2$.) Let $m_n = \mathbf{E}N_n$.

1. Derive a recursive formula for m_n and use it to calculate m_i, $i = 2, 3, 4$ and to show that $m_5 \approx 13.024$.

   Solution: We will form the solution inductively. We start by calculating the expected number of throws given one die. This can be thought of as the expected number of times one die needs to be thrown to get a six. This probability is geometric.

   $$\mathbf{P}\{\text{first six at toss k}\} = p_k = (1 - p)^{k-1}p$$

   where $p = \frac{1}{6}$. The mean of a random variable X with this distribution is well known:

   $$\mathbf{E}X = \sum_k kp_k = \sum_k k(1 - p)^{k-1}p = p^{-1}$$

It is easy to see that $m_1 = 6$. We build inductively on this. Define:

$$P_i^m = \binom{m}{i} p^i (1-p)^{m-i}$$

as the probability that, given m dice, i of them are sixes. Given m_1 we can construct m_2 using a conditioning argument:

$$m_2 = P_2^2 + P_1^2(m_1 + 1) + P_0^2(m_2 + 1) \quad \text{or} \quad m_2 = \frac{P_2^2 + P_1^2(m_1+1) + P_0^2}{1 - P_0^2}$$

and

$$\begin{aligned} m_3 &= P_3^3 + P_2^3(m_1+1) + P_1^3(m_2+1) + P_0^3(1+m_3) \\ \Rightarrow m_3 &= \frac{P_3^3 + P_2^3(m_1+1) + P_1^3(m_2+1) + P_0^3}{1 - P_0^3} \end{aligned}$$

Following this model we arrive at:

$$m_n = \frac{P_n^n + \sum_{i=1}^{n-1} P_{n-i}^n(1+m_i) + P_0^n}{1 - P_0^n}$$

The following MATLAB program was used to verify the above equation:

```
function [expected] = stochasticsP1(n)
 expected = zeros(n,1);
 expected(1) = 6;
 temp = 0;
 for i = 2:n
  temp = 0;
   for j = i:-1:0
    temp = temp + prob(i,j);
     if(j ~= 0 && j ~= i)
       temp = temp + prob(i,j) * expected(i-j);
     end
    end
  expected(i) = temp / (1 - prob(i,0));
end

function p = prob(dice, sixes)
 p = nchoosek(dice, sixes) * (5/6)^(dice - sixes) * (1/6)^sixes;
```

$$m_1 = 6 \quad m_2 = 8.7273 \quad m_3 = 10.5554 \quad m_4 = 11.9267 \quad m_5 = 13.024$$

2. Let X_i denote the number of dice rolled on the i^{th} throw. Find $\mathbf{E}\sum_{i=1}^{N_n} X_i$.

Solution: Rather than answer this question directly, as presented in the problem, it is easier to think about what they are asking for. We are asked to find the expected number of dice needed to be thrown to get n sixes. We assume that the dice are all independent and their outcomes are identically distributed. If we let Y_i be the number times die i needs to be thrown to get a six, this gives us:

$$\mathbf{E}\left[\sum_{i=1}^{N_n} X_i\right] = \mathbf{E}\left[\sum_{i}^{n} Y_i\right] = n\mathbf{E}[Y_i] = m_1 n = 6n$$

using the result from the first section.

Problem 3 (10 pts): Consider the quadratic equation $x^2 + Bx + C = 0$ where B and C are independent and have uniform distributions on $[-n, n]$. Find the probability that the equations has real roots. What happens as $n \to \infty$?

Solution: For the quadratic equation above to have real root we must have

$$\frac{B^2}{4} \geq C.$$

Now note that the pair (B, C) are drawn uniformily (and independently) from the square $[-n, n] \times [-n, n]$ and thus the probability of real roots is proportional to the area under the curve $y(x) = x^2/4$. Thus

$$\mathbf{P}\left\{\text{real roots}\right\} = \begin{cases} \frac{1}{2} + \frac{n}{24} & n \leq 4 \\ 1 - \frac{2}{3\sqrt{n}} & n > 4 \end{cases}.$$

Taking the limit as $n \to \infty$ we see that $\mathbf{P}\left\{\text{real roots}\right\} = 1$.

Problem 4 (10 pts): Let $X_1, X_2, \cdots$ be a sequence of independent identically distributed continuous random variables. We say that record at time n occurs if $X_n > \max\{X_1, \cdots, X_{n-1}\}$. That is, X_n is a record if it is larger than each of the previous X_i's.

Let

$$N = \min\{n : n > 1 \text{ and a record occurs at time } n\}.$$

Show $\mathbf{E}N = \infty$.

Solution: Let

$$\mathbf{P}(X_i \leq x) = \int_{\infty}^{x} f(x)dx = F(x)$$

First we note that from how a record is defined, we have from the i.i.d. assumption,

$$\mathbf{P}(N = n | X_1 = x) = \mathbf{P}(X_2 \leq x, \ldots, X_{n-1} \leq x, X_n > x | X_1 = x) = F(x)^{n-2}(1 - F(x)).$$

Applying this and a conditioning argument, we have,

$$\begin{aligned}
\mathbf{P}(N = n) &= \int_{-\infty}^{\infty} \mathbf{P}(N = n | X_1 = x) f(x)dx \\
&= \int_{-\infty}^{\infty} f(x)F^{n-2}(x)(1 - F(x))dx \\
&= \frac{1}{n-1}\int_{-\infty}^{\infty} (n-1)f(x)F^{n-2}(x)dx - \frac{1}{n}\int_{-\infty}^{\infty} nf(x)F^{n-1}(x)dx \\
&= \frac{1}{n-1}\int \frac{d}{dx}F^{n-1}(x)dx - \frac{1}{n}\int \frac{d}{dx}F^{n}(x)dx \\
&= \left.\frac{F^{n-1}(x)}{n-1}\right|_{-\infty}^{\infty} - \left.\frac{F^{n}(x)}{n}\right|_{-\infty}^{\infty} \\
&= \frac{1}{n(n-1)}.
\end{aligned}$$

Finally, we have,

$$\mathbf{E}N = \sum_{n=2}^{\infty} n\mathbf{P}(N = n) = \sum_{n=2}^{\infty} \frac{n}{n(n-1)} = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

Problem 5 (10 pts): Compute the maximum likelihood estimators for a random sample of Beta (α^*, β^*) population.

Solution: We start with a set of samples drawn from a Beta distribution: X_i, $i = 1, \ldots, n$. The likelihood of these observations is given by

$$L(X_1, \ldots, X_n; \alpha, \beta) = \prod_{i=1}^{n} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} X_i^{\alpha-1}(1-X_i)^{\beta-1} = \frac{\Gamma^n(\alpha+\beta)}{\Gamma^n(\alpha)\Gamma^n(\beta)} \prod_{i=1}^{n} X_i^{\alpha-1}(1-X_i)^{\beta-1}.$$

Thus we want

$$(\alpha^*, \beta^*) = \arg\max_{(\alpha,\beta)} L(X_1, \ldots, X_n; \alpha, \beta) \quad \text{s.t. } \alpha > 0, \beta > 0.$$

Problem 6 (10 pts): Suppose we observe n independent samples of a random variable X, which has mean μ. We call X_i the i-th independent sample and denote the sample mean with $\hat{\mu}$.

Compute the mean of the following two quantities. Which is an unbiased estimator?

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$
$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

Solution:

$$\begin{aligned}
\mathbf{E}[\hat{\sigma}^2] &= \mathbf{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - n^{-1}\sum_{j=1}^{n} X_j)^2\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \mathbf{E}\left[X_i^2 - \frac{2}{n} X_i \sum_{j=1}^{n} X_j + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} X_j X_k\right] \\
&= \frac{1}{n-1}\frac{1}{n}\left(n^2\mathbf{E}X^2 - 2n(\mathbf{E}X^2 + (n-1)\mathbf{E}[X]^2) + n\mathbf{E}X^2 + (n^2-n)\mathbf{E}[X]^2\right) \\
&= \frac{1}{n(n-1)}\left((n^2 - 2n + n)\mathbf{E}X^2 + (-2n^2 + 2n + n^2 - n)\mathbf{E}[X]^2\right) \\
&= \mathbf{E}X^2 - \mathbf{E}[X]^2 = \sigma^2
\end{aligned}$$

Thus it is an unbiased estimator. For $\mathbf{E}[s^2]$, everything is the same except the leading coefficient. Therefore, $\mathbf{E}[s^2] = \frac{n-1}{n}\sigma^2$ which means s^2 is a biased estimator.

Problem 7 (10 pts): Suppose that f is a strictly positive continuous joint density of a random vector (X, Y) with $\mathbf{E}X^2 < \infty$.

1. Compute $\mathbf{E}(X|Y \in [y, y+h])$

   Solution: Recall the marginal probability density function for Y:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx \tag{1}$$

   Consider the probability distribution function for

$$\mathbf{P}(X \le x | Y \in [y, y+h]),$$

   which is given by,

$$F_{X|Y}(x; h) = \mathbf{P}(X \le x | Y \in [y, y+h]) = \frac{\mathbf{P}(X \le x \cap |Y \in [y, y+h])}{\mathbf{P}(Y \in [y, y+h])} \tag{2}$$

Using the joint probability distribution we have,

$$\mathbf{P}(X \leq x \cap y \in [y, y+h]) = \int_{-\infty}^{x} \int_{y}^{y+h} f(\eta, \xi) d\xi d\eta, \tag{3}$$

and using the marginal density for Y we have,

$$\mathbf{P}(Y \in [y, y+h]) = \int_{y}^{y+h} f_Y(\xi) d\xi. \tag{4}$$

Plugging (3) and (4) into (2) we have,

$$\mathbf{P}(X \leq x | Y \in [y, y+h]) = \frac{\int_{-\infty}^{x} \int_{y}^{y+h} f(\eta, \xi) d\xi d\eta}{\int_{y}^{y+h} f_Y(\xi) d\xi} \tag{5}$$

We now differentiate (5) to get the density,

$$\begin{aligned}
f_{X|Y}(x; h) = {F_{X|Y}}'(x; h) &= \frac{d}{dx} \left(\frac{\int_{-\infty}^{x} \int_{y}^{y+h} f(\eta, \xi) d\xi d\eta}{\int_{y}^{y+h} f_Y(\xi) d\xi} \right) \\
&= \frac{\frac{d}{dx} \int_{-\infty}^{x} \int_{y}^{y+h} f(\eta, \xi) d\xi d\eta}{\int_{y}^{y+h} f_Y(\xi) d\xi} \\
&= \frac{\int_{y}^{y+h} f(x, \xi) d\xi}{\int_{y}^{y+h} f_Y(\xi) d\xi}
\end{aligned}$$

Thus we can define the expectation of X as:

$$\mathbf{E}\left[X | Y \in [y, y+h]\right] = \int_{-\infty}^{\infty} x f_{X|Y}(x; h) dx = \int_{-\infty}^{\infty} x \left(\frac{\int_{y}^{y+h} f(x, \xi) d\xi}{\int_{y}^{y+h} f_Y(\xi) d\xi} \right) dx \tag{6}$$

2. Compute $\phi(y) = \lim_{h \to 0} \mathbf{E}\left(X | Y \in [y, y+h]\right)$

   Solution: First note that the condition $\mathbf{E}X^2 < \infty$ allows for interchange of the limit and integral from the dominated convergence theorem. This is a technical point that is stated here just to make clear that interchanges like this cannot be performed on a whim. We then can apply L'Hospital's Rule as we take the limit in h and obtain the result as follows:

$$\begin{aligned}
\lim_{h \to 0} \mathbf{E}\left[X | Y \in [y, y+h]\right] &= \lim_{h \to 0} \int_{-\infty}^{\infty} x \left(\frac{\int_{y}^{y+h} f(x, \xi) d\xi}{\int_{y}^{y+h} f_Y(\xi) d\xi} \right) dx \\
&= \int_{-\infty}^{\infty} x \lim_{h \to 0} \left(\frac{\frac{d}{dh} \int_{y}^{y+h} f(x, \xi) d\xi}{\frac{d}{dh} \int_{y}^{y+h} f_Y(\xi) d\xi} \right) dx \\
&= \int_{-\infty}^{\infty} x \lim_{h \to 0} \left(\frac{f(x, y+h)}{f_Y(y+h)} \right) dx \\
&= \int_{-\infty}^{\infty} x \frac{f(x, y)}{f_Y(y)} dx
\end{aligned}$$