ASSIGNMENT 3

MATH 356

QUESTION 1

Part (a): Using a fair die, the probability that we roll a 4 is $\frac{1}{6}$. Thus, define $S_n := F_1 + F_2 + ... + F_n$, where each F_i is a Bernoulli variable signifying a 4 rolled, with $p = 1/6$. S_n, then, is $\text{Bin}(n, \frac{1}{6})$.

We want to consider the limit of the ratio of S_n to all rolls, where $\frac{S_n}{n} \geq 17\%$ is the event desired. To help clutter, notate $\frac{S_n}{n}$ as E_n.

Let $\varepsilon = .17 - 1/6$. Then

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(E_n \geq .17) &= \lim_{n\to\infty} \mathbb{P}(E_n \geq 1/6 + \varepsilon) \\ &= \lim_{n\to\infty} [1 - \mathbb{P}(E_n < 1/6 + \varepsilon)] \\ &= 1 - \lim_{n\to\infty} \mathbb{P}(E_n - 1/6 < \varepsilon)\end{aligned}$$

But the law of large numbers tells us that $\lim \mathbb{P}(|E_n - 1/6| < \varepsilon) = 1$, which thus implies that $\lim \mathbb{P}(E_n - 1/6 < \varepsilon) \geq 1$, or just 1.

$$\implies \lim_{n\to\infty} \mathbb{P}\left(\frac{S_n}{n} \geq .17\right) = 1 - \lim_{n\to\infty} \mathbb{P}\left(\frac{S_n}{n} - 1/6 < \varepsilon\right) = 1 - 1 = 0$$

Part (b): Let S_{n_i} denote the number of rolls of $i \in [1,6]$ out of n attempts, and let A_{n_i} be the event that $.16 \leq \frac{S_{n_i}}{n} \leq .17$. A_n, then, is $A_{n_1} \cap ... \cap A_{n_6}$.

Since the A_{n_i}'s are not necessarily independent, we need instead to consider

$$A_n^c = A_{n_1}^c \cup ... \cup A_{n_6}^c$$

where $\mathbb{P}(A_n) = 1 - \mathbb{P}(A_n^c)$ and $A_{n_i}^c$ is the event that $\frac{S_{n_i}}{n}$ is *not* bound by $[.16, .17]$, or $\left\{\frac{S_{n_i}}{n} > .17\right\} \cup \left\{\frac{S_{n_i}}{n} < .16\right\}$.

From the previous part, adapting notation, we have that $\lim \mathbb{P}\left(E_{n_i} \geq .17\right) = 0$, and $\mathbb{P}\left(E_{n_i} \geq .17\right) \geq \mathbb{P}\left(E_{n_i} > .17\right)$, so $\lim \mathbb{P}\left(E_{n_i} > 0.17\right) \leq 0 = 0$. However, we still need to show that $\mathbb{P}\left(E_{n_i} < .16\right) \to 0$ as well. For this, we replicate the proof above, with $\varepsilon = 1/6 - .16$

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(E_{n_i} < .16) &= \lim_{n\to\infty} \mathbb{P}(E_{n_i} < 1/6 - \varepsilon) \\
&= \lim_{n\to\infty} \left[1 - \mathbb{P}(E_{n_i} \geq 1/6 - \varepsilon)\right] \\
&= 1 - \lim_{n\to\infty} \mathbb{P}(-\varepsilon \leq E_{n_i} - 1/6)
\end{aligned}$$

The LOBN implies that $\lim \mathbb{P}(-\varepsilon < E_{n_i} - 1/6) = 1$, and thus $\lim \mathbb{P}(-\varepsilon \leq E_{n_i} - 1/6) \geq 1$, or just 1.

$$\implies \lim_{n\to\infty} \mathbb{P}(E_{n_i} < .16) = 1 - 1 = 0$$

$$\mathbb{P}(A_{n_i}^c) = \mathbb{P}\left(\frac{S_{n_i}}{n} > .17 \cup \frac{S_{n_i}}{n} < .16\right) \leq \mathbb{P}\left(\frac{S_{n_i}}{n} > .17\right) + \mathbb{P}\left(\frac{S_{n_i}}{n} < .16\right) \to 0$$

$$\text{Thus} \quad \mathbb{P}(A_n^c) = \mathbb{P}(A_{n_1}^c \cup ... \cup A_{n_6}^c) \leq \sum_{i=1}^{6} \mathbb{P}(A_{n_i}^c) \to 0$$

$$\text{And finally} \quad \mathbb{P}(A_n) = 1 - \mathbb{P}(A_n^c) \to 1 - 0 = 1$$

Now that $\lim \mathbb{P}(A_n) = 1$ has been established, we know for certain that there exists some n such that this probability is greater than 0.999. Which might've been easier to show directly, but oh well.

QUESTION 2

Consider the PMF of a $\text{Bin}(n, p)$:

$$\rho(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

We want to find $\frac{\rho(k+1)}{\rho(k)}$ for certain values of k, which will tell us whether or not ρ is increasing or not at those values.

$$\begin{aligned}
\rho(k+1) &= \binom{n}{k+1} p^{k+1}(1-p)^{n-k-1} \\
&= \frac{n!(n-k)}{k!(n-k)!(k+1)} p^k (1-p)^{n-k} \left(\frac{p}{1-p}\right) \\
&= \binom{n}{k} p^k (1-p)^{n-k} \left(\frac{n-k}{k+1}\right)\left(\frac{p}{1-p}\right) \\
&= \rho(k)\left(\frac{n-k}{k+1}\right)\left(\frac{p}{1-p}\right) \quad \implies \boxed{\frac{\rho(k+1)}{\rho(k)} = \left(\frac{n-k}{k+1}\right)\left(\frac{p}{1-p}\right)}
\end{aligned}$$

ρ is increasing when $\left(\frac{n-k}{k+1}\right)\left(\frac{p}{1-p}\right) > 1$ and decreasing when $\left(\frac{n-k}{k+1}\right)\left(\frac{p}{1-p}\right) < 1$:

$$\begin{aligned}
&\left(\tfrac{n-k}{k+1}\right)\left(\tfrac{p}{1-p}\right) > 1 \\
&\implies pn > k - p + 1 \\
&\implies k < pn + p - 1 \\
&\text{holds for} \quad k \le \lfloor pn + p - 1 \rfloor
\end{aligned}
\qquad \text{and} \qquad
\begin{aligned}
&\left(\tfrac{n-k}{k+1}\right)\left(\tfrac{p}{1-p}\right) < 1 \\
&\implies pn < k - p + 1 \\
&\implies k > pn + p - 1 \\
&\text{holds for} \quad k \ge \lfloor pn + p \rfloor
\end{aligned}$$

Thus, we've found a range of k for which the PMF is increasing (you can see this happens first), and a range for which it is decreasing.

If p and n are such that $\exists\ k = pn + p - 1$ exactly, then this is the maximum. If not, choose $k \in \{\lfloor pn + p - 1 \rfloor, \lfloor pn + p \rfloor\}$ such that $|k - A| = \min\{A - \lfloor A \rfloor, \lfloor A + 1 \rfloor - A\}$, where $A := pn + p - 1$, per the equations above (we are singling out the nearest integer value to the theoretic maximum, $pn + p - 1$). If $\lfloor pn + p \rfloor$ and $\lfloor pn + p - 1 \rfloor$ are equidistant from A, then *both* of these are maximums.

As justification for this last paragraph (and that of the next page), note that a maximum is attained when

$$\frac{\rho(k+1)}{\rho(k)} = 1$$

The closer this expression is to 1, the nearer k is to the "continuous" maximum.

Now we consider the PMF of a Poi(λ), $\rho(k) = \frac{e^{-\lambda}\lambda^k}{k!}$.

We have $\rho(k+1) = \frac{e^{-\lambda}\lambda^{k+1}}{(k+1)!} = \frac{e^{-\lambda}\lambda^k}{k!}\left(\frac{\lambda}{k+1}\right) = \rho(k)\left(\frac{\lambda}{k+1}\right)$. Thus, $\boxed{\frac{\rho(k+1)}{\rho(k)} = \frac{\lambda}{k+1}}$

As before, we consider the k's for which $\frac{\lambda}{k+1}$ is > or < than 1.

$\frac{\lambda}{k+1} > 1 \implies \lambda > k+1 \implies k < \lambda - 1 \implies k \leq \lfloor \lambda - 1 \rfloor$, since k is an integer, as done previously.

Similarly, $\frac{\lambda}{k+1} < 1 \implies k > \lambda - 1 \implies k \geq \lfloor \lambda \rfloor$. Thus, we've found ranges of k for which $\rho(k)$ is strictly increasing and strictly decreasing, one before the other, since $\lfloor \lambda - 1 \rfloor \leq \lfloor \lambda \rfloor$.

We can further deduce that, if $\exists\ k = \lambda - 1$ exactly, this is the maximum. If not, take $k \in \{\lfloor \lambda - 1 \rfloor, \lfloor \lambda \rfloor\}$ such that $|k - (\lambda - 1)| = \min\{\lambda - 1 - \lfloor \lambda - 1 \rfloor, \lfloor \lambda \rfloor - (\lambda - 1)\}$ (the nearest integer value to the proposed maximum), and this will be the discrete maximum. And finally, if $\exists$ integers k_1, k_2 which are equidistant from $\lambda - 1$, they are both maximums.

QUESTION 3

Part (a): Consider $\mathbb{E}[X(X-1)...(X-n+1)]$, where $X \sim \text{Poi}(\lambda)$. We have

$$\begin{aligned}
\mathbb{E}[X(X-1)...(X-n+1)] &= \sum_{k=0}^{\infty} k(k-1)...(k-n+1)\frac{e^{-\lambda}\lambda^k}{k!} \\
&= \sum_{k=n}^{\infty} k(k-1)...(k-n+1)\frac{e^{-\lambda}\lambda^k}{k!} \quad \text{since the first } n-1 \text{ terms vanish} \\
&= \sum_{k=n}^{\infty} \frac{k!}{(k-n)!}\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=n}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-n)!} \\
\text{Letting } k \to k+n \ &= \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^{k+n}}{k!} = \lambda^n \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \boxed{= \lambda^n}
\end{aligned}$$

Part (b): Consider $\mathbb{E}[X^3]$. This is the same as $\mathbb{E}[X(X-1)(X-2)] + 3\mathbb{E}[X^2] - 2\mathbb{E}X$, since $\mathbb{E}[X(X-1)(X-2)] = \mathbb{E}[X^3 - 3X^2 + 2X]$, and applying linearity.

The first term, $\mathbb{E}[X(X-1)(X-2)]$, is the 3^{rd} factorial moment of X, and from above equals λ^3.

$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}X = \mathbb{E}[X(X-1)] + \lambda = \lambda^2 + \lambda$, substituting for the 2^{nd} factorial moment.

Finally, we have $\mathbb{E}[X^3] = \lambda^3 + 3(\lambda^2 + \lambda) - 2\lambda \boxed{= \lambda^3 + 3\lambda^2 + \lambda}$

Question 4

The probability that, out of n randomly chosen attendees, not including myself, *no one* shares my birthday is $\left(\frac{364}{365}\right)^n$, since they are allowed to share birthdays between themselves. (We've simply "removed" the unallowed date.)

Thus, the compliment of this event is the event that at least *one* person shares my birthday, which is what we want.

$\implies \mathbb{P}\{\geq 1 \text{ shared birthday}\} = 1 - \left(\frac{364}{365}\right)^n$, which we need $\geq \frac{2}{3}$

Thus $\quad \left(\frac{364}{365}\right)^n \leq \frac{1}{3} \implies n = \log_{\frac{364}{365}}(\frac{1}{3}) = n \approx 400.44$

We need at least 441 people, then, to ensure with $\mathbb{P} = 2/3$ that at least one person shares my birthday.

QUESTION 5

Let T be the event that, on any given day, you flip 5 tails in a row. $\mathbb{P}(T) = \left(\frac{1}{2}\right)^5$. We can extrapolate a random variable based on this, $T_i \sim \text{Ber}\left(\frac{1}{2}\right)^5$, which takes on values $\{0, 1\}$: 0 for a failure, 1 for a success on the day $i \in [1, 30]$.

Define the event $\{X = k\} := \{T_1 + T_2 + ... + T_{30} = k\}$. This is $\text{Bin}\left[30, \left(\frac{1}{2}\right)^5\right]$. To approximate $\mathbb{P}(X = 2)$, we are counting rare occurrences, i.e. that one flips 5 tails with probability ≈ 0.031, which are each completely independent of eachother. We can thus assume that a Poisson approximation will be OK. The error bound using this is at most

$$np^2 \approx 0.029$$

By contrast, a simple error bound on a normal approximation is

$$\frac{3}{\sqrt{npq}} = \frac{3}{\sqrt{30 \cdot \frac{1}{2^5}\left(1 - \frac{1}{2^5}\right)}} \approx 3.15$$

which is no good at all.

Let $\lambda = np$. Then we can approximate $\mathbb{P}(X = 2)$ as

$$\frac{e^{-\frac{30}{2^5}}\left(\frac{30}{2^5}\right)^2}{2!} \boxed{\approx .172}$$

QUESTION 6

Let $F_n(t)$ be the cumulative distribution function of X_n, with $F_n(t) = \mathbb{P}(X_n \leq \lfloor t \rfloor)$. We then have

$$\mathbb{P}(X_n \leq \lfloor t \rfloor) = \sum_{i=1}^{\lfloor t \rfloor} \rho(t) = \sum_{i=1}^{\lfloor t \rfloor} \frac{1}{n} = \frac{\lfloor t \rfloor}{n}$$

with $\rho(t)$ being the PMF of X_n. Note that X_n takes on values $S_n = \{1, ..., n\}$, and the set of $i \in S_n$ with $i \leq t$ is just $\{1, ..., \lfloor t \rfloor\}$.

As $n \to \infty$, $F_n(t) = \frac{\lfloor t \rfloor}{n} \to 0$. The distribution of X_n, $\mathbb{P}(X = k) = \frac{1}{n}$, similarly tends to 0. We can then conclude

$$F_n(t) = \lfloor t \rfloor \mathbb{P}(X = k) \quad \text{and} \quad F_n(t) \sim \mathbb{P}(X = k), \text{ the distribution of } X_n$$