1. Midterm

Solve all of the problems on the midterm again (including the ones you got correct).

Solution:
See midterm solutions.

2. Confidence Interval Comparisons

In order to estimate the probability of a head in a coin flip, \( p \), you flip a coin \( n \) times, where \( n \) is a positive integer, and count the number of heads, \( S_n \). You use the estimator \( \hat{p} = S_n/n \).

(a) You choose the sample size \( n \) to have a guarantee

\[
P(|\hat{p} - p| \geq \epsilon) \leq \delta.
\]

Using Chebyshev Inequality, determine \( n \) with the following parameters:

(i) Compare the value of \( n \) when \( \epsilon = 0.05, \delta = 0.1 \) to the value of \( n \) when \( \epsilon = 0.1, \delta = 0.1 \).

(ii) Compare the value of \( n \) when \( \epsilon = 0.1, \delta = 0.05 \) to the value of \( n \) when \( \epsilon = 0.1, \delta = 0.1 \).

(b) Now, we change the scenario slightly. You know that \( p \in (0.4, 0.6) \) and would now like to determine the smallest \( n \) such that

\[
P\left(\frac{|\hat{p} - p|}{p} \leq 0.05\right) \geq 0.95.
\]

Use the CLT to find the value of \( n \) that you should use.

Solution:
(a) Chebyshev Inequality implies that:
\[ P\left( \left| \frac{S_n}{n} - p \right| \geq \epsilon \right) \leq \frac{\text{var}(S_n/n - p)}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \]
Thus, we set \( \delta = p(1-p)/(n\epsilon^2) \) or \( n = p(1-p)/(\delta\epsilon^2) \). Thus, when \( \epsilon \) is reduced to half of its original value, \( n \) is changed to 4 times its original value, and when \( \delta \) is reduced to half of its original value, \( n \) will be twice its original value. In order to be more concrete, we may maximize \( p(1-p)/(\delta\epsilon^2) \) by letting \( p = 1/2 \). Thus, when \( \epsilon = 0.1, \delta = 0.1, n = 250 \). Letting \( \delta = 0.05 \) results in \( n = 500 \), while letting \( \epsilon = 0.05 \) results in \( n = 1000 \).

(b) Note that by the CLT:
\[ \sqrt{n} \frac{\hat{p} - p}{\sqrt{p(1-p)}} \sim N(0,1) \]
We are interested in the following:
\[ P\left( \left| \frac{\hat{p} - p}{\sqrt{p(1-p)}} \right| \leq 0.05 \right) \approx P\left( |N(0,1)| \leq 0.05 \frac{\sqrt{np}}{\sqrt{1-p}} \right) \]
Now, we use the condition that we want:
\[ P\left( |N(0,1)| \leq 0.05 \frac{\sqrt{np}}{\sqrt{1-p}} \right) \geq 0.95 \]
This implies that \( 0.05\sqrt{np/(1-p)} \geq 2 \) (note we use 2 here for simplicity, if you used 1.96, this is completely correct), or \( n \geq 1600(1-p)/p \). We now use the fact that we know \( p \in [0.4, 0.6] \). Since \( p \in [0.4, 0.6] \), we can see that the value \( (1-p)/p \) is maximized when \( p = 0.4 \). Thus, we note that \( n \geq 1600(1-p)/p \) for all values of \( p \), so the minimum value of \( n \) must be the maximum valid value of \( 1600(1-p)/p = 2400 \).

3. Convergence in Probability
Let \( (X_n)_{n=1}^{\infty} \), be a sequence of i.i.d. random variables distributed uniformly in \([-1, 1]\]. Show that the following sequences \( (Y_n)_{n=1}^{\infty} \) converge in probability to some limit.

(a) \( Y_n = (X_n)^n \).
(b) \( Y_n = \prod_{i=1}^{n} X_i \).
(c) \( Y_n = \max\{X_1, X_2, \ldots, X_n\} \).
(d) \( Y_n = (X_1^2 + \cdots + X_n^2)/n \).

Solution:
(a) For any \( \epsilon > 0 \), \( P(|Y_n| > \epsilon) = P(|X_n| > \epsilon^{1/n}) = 1 - \epsilon^{1/n} \to 0 \) as \( n \to \infty \). Thus, the sequence converges to 0 in probability.
(b) By independence of the random variables,
\[
\mathbb{E}[Y_n] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_n] = 0,
\]
\[
\text{var } Y_n = \mathbb{E}[Y_n^2] = (\text{var } X_1)^n = \left(\frac{1}{3}\right)^n.
\]
Now since \(\text{var } Y_n \to 0\) as \(n \to \infty\), by Chebyshev’s Inequality the sequence converges to its mean, that is, 0, in probability.

(c) Consider \(\epsilon \in [0, 1]\). We see that:
\[
P(|Y_n - 1| \geq \epsilon) = P(\max\{X_1, \ldots, X_n\} \leq 1 - \epsilon)
\]
\[
= P(X_1 \leq 1 - \epsilon, \ldots, X_n \leq 1 - \epsilon)
\]
\[
= P(X_1 \leq 1 - \epsilon)^n = \left(1 - \frac{\epsilon}{2}\right)^n
\]
Thus, \(P(|Y_n - 1| \geq \epsilon) \to 0\) as \(n \to \infty\) and we are done.

(d) The expectation is
\[
\mathbb{E}[Y_n] = \frac{1}{n} \cdot n \mathbb{E}[X_1^2] = \frac{1}{3}.
\]
Then, we bound the variance.
\[
\text{var } Y_n = \frac{1}{n} \text{ var } X_1^2 \leq \frac{1}{n} \to 0 \quad \text{as } n \to \infty,
\]
since \(X_1^2 \leq 1\). Hence, we see that \(Y_n \to 1/3\) in probability as \(n \to \infty\).

**Remark:** We now provide an interpretation for the previous result. The sample space for \(Y_n\) is \(\Omega_n = [-1, 1]^n\), which is an \(n\)-dimensional cube. The result we have just proved shows that, for any \(\epsilon > 0\), the set
\[
B_n = \left\{ x \in \mathbb{R}^n : \frac{1}{3}(1 - \epsilon) \leq \frac{x_1^2 + \cdots + x_n^2}{n} \leq \frac{1}{3}(1 + \epsilon) \right\}
\]
makes up “most” of the volume of \(\Omega_n\), in the sense that
\[
\frac{\text{volume}(B_n \cap [-1, 1]^n)}{2^n} \to 1 \quad \text{as } n \to \infty.
\]
Since \(B_n\) is close to the boundary of a ball of radius \(\sqrt{n/3}\), the result can be stated facetiously as “nearly all of the volume of a high-dimensional cube is contained in the boundary of a ball”. Although this may seem like a meaningless comment, in fact various phenomena such as these contribute to the so-called “curse of dimensionality” in machine learning, which concerns the sparsity of data in high-dimensional statistics.

4. Almost Sure Convergence

In this question, we will explore almost sure convergence and compare it to convergence in probability. Recall that a sequence of random variables \((X_n)_{n \in \mathbb{N}}\) converges **almost surely** (abbreviated a.s.) to \(X\) if \(P(\lim_{n \to \infty} X_n = X) = 1\).
(a) Suppose that, with probability 1, the sequence \((X_n)_{n \in \mathbb{N}}\) oscillates between two values \(a \neq b\) infinitely often. Is this enough to prove that \((X_n)_{n \in \mathbb{N}}\) does not converge almost surely? Justify your answer.

(b) Suppose that \(Y\) is uniform on \([-1, 1]\), and \(X_n\) has distribution
\[
\mathbb{P}(X_n = (y + n^{-1})^{-1} \mid Y = y) = 1.
\]

Does \((X_n)_{n=1}^{\infty}\) converge a.s.?

(c) Define random variables \((X_n)_{n \in \mathbb{N}}\) in the following way: first, set each \(X_n\) to 0. Then, for each \(k \in \mathbb{N}\), pick \(j\) uniformly randomly in \(\{2^k, \ldots, 2^{k+1} - 1\}\) and set \(X_j = 2^k\). Does the sequence \((X_n)_{n \in \mathbb{N}}\) converge a.s.?

(d) Does the sequence \((X_n)_{n \in \mathbb{N}}\) from the previous part converge in probability to some \(X\)? If so, is it true that \(\mathbb{E}[X_n] \to \mathbb{E}[X]\) as \(n \to \infty\)?

**Solution:**

(a) Yes. If a sequence oscillates between two values infinitely often, then it does not converge. Here, we have a sequence that oscillates between two values infinitely often (with probability 1), which means that the sequence does not converge (with probability 1). (Perhaps we could name this “almost surely not converging”!)

The above paragraph was very cumbersome to read, which is why we often abbreviate “with probability 1” with a.s. With this abbreviation, here is how the above justification reads: \((X_n)_{n \in \mathbb{N}}\) oscillates between two values infinitely often a.s., so \((X_n)_{n \in \mathbb{N}}\) does not converge a.s.

(b) Yes. Observe that when \(Y = y \neq 0\), \((X_n)_{n \in \mathbb{N}}\) will converge to \(y^{-1}\). When \(Y = 0\), \((X_n)_{n \in \mathbb{N}}\) does not converge; however, \(\mathbb{P}(Y = 0) = 0\) since \(Y\) is a continuous random variable. In other words,
\[
\mathbb{P}(X_n \text{ does not converge as } n \to \infty) = \mathbb{P}(Y = 0) = 0,
\]
\[
\mathbb{P}(X_n \text{ converges as } n \to \infty) = \mathbb{P}(Y \neq 0) = 1,
\]
so \((X_n)_{n \in \mathbb{N}}\) converges a.s.

(c) No. The sequence \((X_n)_{n \in \mathbb{N}}\) oscillates between 0 and successively higher powers of two infinitely often a.s., so it does not converge a.s.

(d) Yes. Fix \(\varepsilon > 0\). For \(n \in \mathbb{Z}_+\), one has
\[
\mathbb{P}(|X_n| > \varepsilon) = \frac{1}{2^k},
\]
where \(k = \lfloor \log_2 n \rfloor\). As \(n \to \infty\), the above probability goes to 0, so \(X_n \to 0\) in probability. Intuitively, \((X_n)_{n \in \mathbb{N}}\) has infinitely many oscillations, so it cannot converge a.s. However, the probability of each oscillation shrinks to 0, so \((X_n)_{n \in \mathbb{N}}\) converges in probability.

The expectations do not converge. For all \(n\), one has \(\mathbb{E}[X_n] = 1\), so it is not the case that \(\mathbb{E}[X_n] \to 0\) as \(n \to \infty\). Hence, convergence in probability is not sufficient to imply that the expectations converge (in fact, almost sure convergence is not sufficient either).
5. Compression of a Random Source

Let \( (X_i)_{i=1}^\infty \overset{i.i.d.}{\sim} p(\cdot) \), where \( p \) is a discrete PMF on a finite set \( \mathcal{X} \). Additionally define the entropy of a random variable \( X \) as \( H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) \). That is, we define

\[
H(X) = \mathbb{E}\left[ \log_2 \frac{1}{p(X)} \right].
\]

(We could also write this as \( H(p) \), since the entropy is really a property of the distribution of \( X \).) In this problem, we will show that a random source whose symbols are drawn according to the distribution \( p \) can be compressed to \( H(X) \) bits per symbol. In the lab, you will implement this coding and compare it to Huffman coding.

(a) Show that

\[
-\frac{1}{n} \log_2 p(X_1, \ldots, X_n) \xrightarrow{n \to \infty} H(X_1) \quad \text{almost surely.}
\]

(Here, we are extending the notation \( p(\cdot) \) to denote the joint PMF of \( (X_1, \ldots, X_n) : p(x_1, \ldots, x_n) := p(x_1) \cdots p(x_n) \).)

(b) Fix \( \epsilon > 0 \) and define \( A^{(n)}_\epsilon \) as the set of all sequences \( (x_1, \ldots, x_n) \in \mathcal{X}^n \) such that:

\[
2^{-n(H(X_1)+\epsilon)} \leq p(x_1, \ldots, x_n) \leq 2^{-n(H(X_1)-\epsilon)}.
\]

Show that \( \mathbb{P}(\{(X_1, \ldots, X_n) \in A^{(n)}_\epsilon\}) > 1 - \epsilon \) for all \( n \) sufficiently large. Consequently, \( A^{(n)}_\epsilon \) is called the typical set because the observed sequences lie within \( A^{(n)}_\epsilon \) with high probability.

(c) Show that \( (1 - \epsilon)2^{n(H(X_1)-\epsilon)} \leq |A^{(n)}_\epsilon| \leq 2^{n(H(X_1)+\epsilon)} \), for \( n \) sufficiently large.

Parts (b) and (c) are called the asymptotic equipartition property (AEP) because they say that there are \( \approx 2^{nH(X_1)} \) observed sequences which each have probability \( \approx 2^{-nH(X_1)} \). Thus, by discarding the sequences outside of \( A^{(n)}_\epsilon \), we need only keep track of \( 2^{nH(X_1)} \) sequences, which means that a length-\( n \) sequence can be compressed into \( \approx nH(X_1) \) bits, requiring \( H(X_1) \) bits per symbol.

(d) Now show that for any \( \delta > 0 \) and any positive integer \( n \), if \( B_n \subseteq \mathcal{X}^n \) is a set with \( |B_n| \leq 2^{n(H(X_1)-\delta)} \), then \( \mathbb{P}(\{(X_1, \ldots, X_n) \in B_n\}) \to 0 \) as \( n \to \infty \).

This says that we cannot compress the observed sequences of length \( n \) into any set smaller than size \( 2^{nH(X_1)} \).

[Hint: Consider the intersection of \( B_n \) and \( A^{(n)}_\epsilon \).]

(e) Next we turn towards using the AEP for compression. Recall that in order to encode a set of size \( n \) in binary, it requires \( \lfloor \log_2 n \rfloor \) bits. Therefore, a naïve encoding requires \( \lfloor \log_2 |\mathcal{X}| \rfloor \) bits per symbol.

From (b) and (d), if we use \( \log_2 |A^{(n)}_\epsilon| \approx nH(X_1) \) bits to encode the sequences in \( A^{(n)}_\epsilon \), ignoring all other sequences, then the probability of error
with this encoding will tend to 0 as \( n \to \infty \), and thus an asymptotically error-free encoding can be achieved using \( H(X_1) \) bits per symbol.

Alternatively, we can create an error-free code by using \( 1 + \lceil \log_2 |A^{(n)}_\epsilon| \rceil \) bits to encode the sequences in \( A^{(n)}_\epsilon \) and \( 1 + n \lfloor \log_2 |X| \rfloor \) bits to encode other sequences, where the first bit is used to indicate whether the sequence belongs in \( A^{(n)}_\epsilon \) or not. Let \( L_n \) be the length of the encoding of \( X_1, \ldots, X_n \) using this code; show that \( \lim_{n \to \infty} E[L_n]/n \leq H(X_1) + \epsilon \). In other words, asymptotically, we can compress the sequence so that the number of bits per symbol is arbitrary close to the entropy.

**Solution:**

(a) Since \( (X_i)_{i=1}^\infty \) is an i.i.d. sequence, so is \( (\log_2 p(X_i))_{i=1}^\infty \). Thus:

\[
-\frac{1}{n} \log_2 p(X_1, \ldots, X_n) = -\frac{1}{n} \sum_{i=1}^n \log_2 p(X_i) \xrightarrow{n \to \infty \text{ a.s.}} -E[\log_2 p(X_1)] = H(X_1)
\]

by the Strong Law of Large Numbers.

(b) As a consequence of (a), \( n^{-1} \log_2 p(X_1, \ldots, X_n) \to H(X_1) \) in probability as \( n \to \infty \), so

\[
P\left( \left| -\frac{1}{n} \log_2 p(X_1, \ldots, X_n) - H(X_1) \right| < \epsilon \right) \to 1 \quad \text{as} \quad n \to \infty.
\]

For \( n \) sufficiently large, the LHS is \( > 1 - \epsilon \).

(c) We have:

\[
1 = \sum_{x \in X^n} p(x) \geq \sum_{x \in A^{(n)}_\epsilon} p(x) \geq \sum_{x \in A^{(n)}_\epsilon} 2^{-n(H(X_1)+\epsilon)} = |A^{(n)}_\epsilon|2^{-n(H(X_1)+\epsilon)}
\]

This shows that \( |A^{(n)}_\epsilon| \leq 2^{n(H(X_1)+\epsilon)} \). Now, we have, for \( n \) sufficiently large:

\[
1 - \epsilon < P((X_1, \ldots, X_n) \in A^{(n)}_\epsilon) \leq \sum_{x \in A^{(n)}_\epsilon} 2^{-n(H(X_1)-\epsilon)} = 2^{-n(H(X_1)-\epsilon)}|A^{(n)}_\epsilon|
\]

Thus, \( |A^{(n)}_\epsilon| \geq (1 - \epsilon)2^{n(H(X_1)-\epsilon)} \).

(d) Pick \( \epsilon \in (0, \delta) \). We can write

\[
P((X_1, \ldots, X_n) \in B_n)
\leq P((X_1, \ldots, X_n) \in A^{(n)}_\epsilon \cap B_n) + P((X_1, \ldots, X_n) \notin A^{(n)}_\epsilon)
\leq \sum_{x \in A^{(n)}_\epsilon \cap B_n} p(x) + P((X_1, \ldots, X_n) \notin A^{(n)}_\epsilon)
\leq |B_n|2^{-n(H(X_1)-\epsilon)} + P((X_1, \ldots, X_n) \notin A^{(n)}_\epsilon)
\leq 2^{-n(\delta-\epsilon)} + P((X_1, \ldots, X_n) \notin A^{(n)}_\epsilon) \to 0
\]

since \( \delta > \epsilon \) and by (b).
(e) Separating out the sequences in the typical set from the sequences which are not in the typical set,

\[
\frac{\mathbb{E}[L_n]}{n} = 1 + \left\lceil \log_2 |A(n)| \right\rceil \mathbb{P}(X_1, \ldots, X_n) \in A(n) \nonumber \\
+ \frac{1 + n \log_2 |\mathcal{X}|}{n} \mathbb{P}(X_1, \ldots, X_n) \notin A(n) \\
\leq 1 + \left\lceil n[H(X_1) + \epsilon] \right\rceil + (1 + \left\lceil \log_2 |\mathcal{X}| \right\rceil) \mathbb{P}(X_1, \ldots, X_n) \notin A(n).
\]

Since \(\mathbb{P}(X_1, \ldots, X_n) \in A(n) \rightarrow 1\) and \(\mathbb{P}(X_1, \ldots, X_n) \notin A(n) \rightarrow 0\), then the second term \(\rightarrow 0\). Asymptotically, only the first term matters, and by taking \(n \rightarrow \infty\) we get \(\lim_{n \rightarrow \infty} \mathbb{E}[L_n]/n \leq H(X_1) + \epsilon\).

6. [Bonus] Balls and Bins: Poisson Convergence

The bonus question is just for fun. You are not required to submit the bonus question, but do give it a try and write down your progress.

Consider throwing \(m\) balls into \(n\) bins uniformly at random. In this question, we will show that the number of empty bins converges to a Poisson limit, under the condition that \(n \exp(-m/n) \rightarrow \lambda \in (0, \infty)\).

(a) Let \(p_k(m, n)\) denote the probability that exactly \(k\) bins are empty after throwing \(m\) balls into \(n\) bins uniformly at random. Show that

\[
p_0(m, n) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \left(1 - \frac{j}{n}\right)^m.
\]

(Hint: Use the Inclusion-Exclusion Principle.)

(b) Show that

\[
p_k(m, n) = \binom{n}{k} \left(1 - \frac{k}{n}\right)^m p_0(m, n - k).
\]

(c) Show that

\[
\binom{n}{k} \left(1 - \frac{k}{n}\right)^m \leq \frac{\lambda^k}{k!}
\]

as \(m, n \rightarrow \infty\) (such that \(n \exp(-m/n) \rightarrow \lambda\)).

(d) Show that

\[
\binom{n}{k} \left(1 - \frac{k}{n}\right)^m \geq \frac{\lambda^k}{k!}
\]

as \(m, n \rightarrow \infty\) (such that \(n \exp(-m/n) \rightarrow \lambda\)). This is the hard part of the proof. To help you out, we have outlined a plan of attack:

i. Show that

\[
\binom{n}{k} \left(1 - \frac{k}{n}\right)^m \geq \left(1 - \frac{k}{n}\right)^{k+m} \frac{n^k}{k!}.
\]
ii. Show that
\[ \ln\left\{ n^k \left(1 - \frac{k}{n}\right)^m \right\} \to k \ln \lambda \]
as \( m, n \to \infty \) (such that \( n \exp(-m/n) \to \lambda \)). You may use the inequality \( \ln(1-x) \geq -x - x^2 \) for \( 0 \leq x \leq 1/2 \).

iii. Show that
\[ \left(1 - \frac{k}{n}\right)^k \to 1 \]
as \( m, n \to \infty \) (such that \( n \exp(-m/n) \to \lambda \)). Conclude that (3) holds.

(e) Now, show that
\[ p_0(m, n) \to \exp(-\lambda). \]
(Try using the results you have already proven.) Conclude that
\[ p_k(m, n) \to \frac{\lambda^k \exp(-\lambda)}{k!}. \]

Solution:

(a) The probability that there are no empty bins is, by the Inclusion-Exclusion Principle,
\[ p_0(m, n) = \sum_{j=0}^{n} (-1)^j \mathbb{P} \text{(some } j \text{ bins are empty)} \]
\[ = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \mathbb{P} \text{(a specific set of } j \text{ bins are empty)} \]
\[ = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \left(1 - \frac{j}{n}\right)^m. \]
The last equality is justified by the following reasoning: if a specific set of \( j \) bins are empty, then each of the \( m \) balls must land in one of the \( n - j \) bins, which occurs with probability \( (1 - j/n)^m \).

(b) If there are exactly \( k \) empty boxes, then there are \( \binom{n}{k} \) ways to choose which boxes are empty; \((1-k/n)^m\) is the probability that these boxes are empty; and \( p_0(m, n-k) \) is the probability that none of the other \( n-k \) boxes are empty.

(c) One has
\[ \binom{n}{k} \left(1 - \frac{k}{n}\right)^m \leq \frac{n!}{k!(n-k)!} \exp\left(-\frac{km}{n}\right) \leq \frac{n \exp(-m/n)^k}{k!} \to \frac{\lambda^k}{k!}. \]

(d) i. Observe that
\[ \binom{n}{k} = \frac{n!}{k!(n-k)!} \geq \frac{(n-k)^k}{k!} = \frac{n^k}{k!} \left(1 - \frac{k}{n}\right)^k, \]
which implies the desired result.
ii. Although we did not ask you to prove \( \ln(1-x) \geq -x - x^2 \), we include a proof for completeness. Assume \( 0 \leq x \leq 1/2 \). Using the power series expansion for \( \ln(1-x) \),

\[
\ln(1-x) = -\sum_{i=1}^{\infty} \frac{x^i}{i} \geq -x - \frac{x^2}{2} \sum_{i=2}^{\infty} \frac{2x^{i-2}}{i}.
\]

Since \( x^i \leq 2^{-i} \), \( \sum_{i=2}^{\infty} i^{-2}x^{i-2} \leq \sum_{i=2}^{\infty} i^{-1}2^{-i-1} \leq \sum_{i=2}^{\infty} 2^{-i-2} = 2 \). Hence (the negative sign reverses the direction of the inequality) \( \ln(1-x) \geq -x - x^2 \).

Applying the inequality, we have

\[
\ln\{n^k(1-k/n)^m\} \geq k \ln n - \frac{km}{n} - \frac{k^2m}{n^2}.
\]

Now, it’s time to estimate the order of \( m \). Since \( n \exp(-m/n) \to \lambda \),

\[
m = n \ln n - n \ln \lambda + o(n),
\]

where \( o(n) \) is a term such that \( o(n)/n \to 0 \) as \( n \to \infty \). With this estimate, we see that \( k^2m/n^2 \to 0 \) as \( n \to \infty \). Therefore, we ignore this term and obtain

\[
\ln\{n^k(1-k/n)^m\} \to k \ln n - k \ln n + k \ln \lambda = k \ln \lambda,
\]

as desired.

iii. \( k \) is a constant, so clearly

\[
\left(1 - \frac{k}{n}\right)^k \to 1 \quad \text{as} \quad n \to \infty.
\]

Putting together all of the various parts together, we have

\[
\binom{n}{k}(1-k/n)^m \geq \left(1 - \frac{k}{n}\right)^m \left(1 - \frac{k}{n}\right)^{k \ln \lambda} \to \left(1 - \frac{k}{n}\right)^{k \ln \lambda} = \frac{\lambda^k}{k!},
\]

as desired.

(e) One has

\[
p_0(m, n) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} (1 - j/n)^m \to \sum_{j=1}^{\infty} (-1)^j \frac{\lambda^j}{j!} = \exp(-\lambda).
\]

Therefore, for any fixed \( k \), \( p_0(m, n - k) \to \exp(-\lambda) \). Hence, from (1), (2), (3), and \( p_0(m, n - k) \to \exp(-\lambda) \), we have our desired result:

\[
p_k(m, n) \to \frac{\lambda^k \exp(-\lambda)}{k!}.
\]