Homework #10 Solutions

Submission is not required for this problem set.

1. Chernoff bound.
   a. Show that the inequality
      \[ P(X \geq a) \leq e^{-sa}M(s) \]
      holds for every \( a \) and every \( s \geq 0 \), where \( M(s) = E[e^{sX}] \) is the moment generating function of \( X \).
      Hint: \( P(X \geq a) = P(e^{sX} \geq e^{sa}) \).
   b. Let \( U_1, U_2, \ldots, U_n \) be i.i.d. such that \( U_i \sim \text{Bern}(p) \) and \( X_n = \sum_{i=1}^{n} U_i \). Show that
      \[ P(X_n \geq (1 + \epsilon)np) \leq e^{-np(1 + \epsilon)} \]
      Hint:
      • Use the result of part (a) to bound \( P(X_n > (1 + \epsilon)np) \).
      • Bound the moment generating function of \( U \) by \( M_U(s) \leq e^{p(e^s - 1)} \).
      • Find \( s \) that gives the most tight bound (i.e., minimize the bound term).
      • You can use the following inequality.
      \[ \frac{e^\epsilon}{(1 + \epsilon)^{1+\epsilon}} \leq e^{-\frac{\epsilon^2}{3}} \]

Solution

a. We have

\[
P(X \geq a) = P(e^{sX} \geq e^{sa}) \\
\leq E[e^{sX}] \\
e^{-sa}E[e^{sX}].
\]

b. For all \( s > 0 \),

\[
P(X_n \geq (1 + \epsilon)np) = P(e^{sX_n} \geq e^{s(1+\epsilon)np}) \\
\leq E[e^{sX_n}]e^{-s(1+\epsilon)np} \\
= \left(E[e^{sU}]e^{-s(1+\epsilon)p}\right)^n \\
= \left((1 - p + pe^s)e^{-s(1+\epsilon)p}\right)^n \\
\leq \left(e^{p(e^s - 1)}e^{-s(1+\epsilon)p}\right)^n \\
= \exp(np(e^s - 1 - s(1 + \epsilon))) .
\]

Note that the minimum of \( e^s - 1 - s(1 + \epsilon) \) is achieved at \( s = \log(1 + \epsilon) \). If we use this value,
we have
\[
P \left( X_n \geq (1 + \epsilon)np \right) \leq \exp \left( np (\epsilon^4 - 1 - s(1 + \epsilon)) \right)
\leq \exp \left( np (\epsilon - (1 + \epsilon) \log(1 + \epsilon)) \right)
\leq \left( \frac{e^\epsilon}{(1 + \epsilon)^{1+\epsilon}} \right)^{np}
\leq e^{-\frac{np^2}{2}}.
\]

2. Confidence intervals. A population of $10^8$ voters choose between two candidates A and B. A fraction $p = 0.55$ of them plan to vote for candidate A and the rest for candidate B. A fair poll with sample size $n$ is performed, i.e., the $n$ samples are i.i.d. and done with replacement (same person may be polled more than once). We want to find a value of $n$ that will guarantee that the majority of those sampled vote for candidate A, with probability at least 0.99. More precisely and equivalently, let $U_1, U_2, \ldots, U_n$ be i.i.d. $\sim$ Bernoulli(0.55). The fraction of 1s in this sequence is thus $X_n = \frac{1}{n} \sum_{i=1}^{n} U_i$ and we want $n$ sufficiently large to guarantee that $P\{X_n > 0.5\} > 0.99$.

a. Use the Chebyshev inequality to find such an $n$.

b. Use the Chernoff inequality to find such an $n$.

Hint: Similar to part (b) of Problem 1, we can show that
\[
P \left( \frac{1}{n} \sum_{i=1}^{n} U_i \leq (1 - \epsilon)p \right) \leq e^{-\frac{np^2}{2}}
\]

c. Use the central limit theorem to approximate such an $n$.

Solution

a. By Chebyshev’s inequality,
\[
P\{X_n - p < -0.05\} \leq P\{|X_n - p| > 0.05\} \leq \frac{\sigma_x^2}{(0.05^2)n}
\]

since $X_i \sim$ Bern(0.1), $\sigma_x^2 = 0.2475$. We need $\frac{\sigma_x^2}{(0.05^2)n} < 0.01$, which implies $n > 9900$.

b. We have
\[
P \left( X_n \leq (1 - \epsilon)p \right) \leq e^{-\frac{np^2}{2}}.
\]

Let $\epsilon = \frac{0.05}{0.55}$ and $p = 0.55$. Then,
\[
P \left( X_n \leq 0.5 \right) \leq e^{-\frac{np^2}{2}}.
\]

We want $\exp \left( -\frac{0.05^2}{2} \right) < 0.01$, which implies $n > 2026$. 


c. We have
\[ \sum_{i=1}^{n} \frac{U_i - p}{\sqrt{np(1-p)}} \sim \mathcal{N}(0,1) \]

Therefore,
\[
P(X_n \leq 0.5) = P\left( \sqrt{\frac{n}{p(1-p)}} (X_n - p) \leq \sqrt{\frac{n}{p(1-p)}} (0.5 - 0.55) \right)
\]
\[
= P\left( \sum_{i=1}^{n} \frac{U_i - p}{\sqrt{np(1-p)}} \leq -\sqrt{\frac{n}{p(1-p)}} \times 0.05 \right)
\]
\[
\approx Q\left( \sqrt{\frac{n}{p(1-p)}} \times 0.05 \right).
\]

We want \( Q\left( \sqrt{\frac{n}{p(1-p)}} \times 0.05 \right) \leq 0.01 \), which implies
\[
n \geq \left( Q^{-1}(0.01)/0.05 \right)^2 \times p(1-p) \approx 535.78.
\]

3. Sharpness of Chebyshev inequality. Let \( X \) be an \( \text{Exp}(1) \) random variable. For \( \alpha > 1 \),
   a. Use Chebyshev inequality to derive a bound on
   \[ P(|X - 1| > \alpha) \]
   b. Compute directly \( P(|X - 1| > \alpha) \).

**Solution**

a. Since
   \[ E[X] = 1, \ Var(X) = 1, \]
   \[ P(|X - 1| > \alpha) \leq \frac{1}{\alpha^2} Var(X) = \frac{1}{\alpha^2}. \]
   b. \( P(|X - 1| > \alpha) = 1 - F_X(\alpha + 1) = e^{-1-\alpha}. \)

4. Gambling. Let \( X_n \) be the amount you win on the \( n \)th round of a game of chance. Assume that \( X_1, X_2, \ldots, X_n \) are i.i.d. with finite mean \( E(X) \) and variance \( \sigma^2 \). Make the realistic assumption that \( E[X] < 0 \). Show that \( P\{((X_1 + X_2 + \ldots + X_n)/n < E[X]/2) \to 1 \). What is the moral of this result?

**Solution**

We know \( \frac{1}{n} \sum_{i=1}^{n} X_i \to E[X] \) by the weak law of large numbers, which means \( P(|S_n - E[X]| > \epsilon) \to 0 \) as \( n \to \infty \). The limiting value of \( P(S_n < \frac{E[X]}{2}) \) depends on \( E[X] \). When \( E[X] < 0 \),
$P(S_n < \frac{E[X]}{2})$ approaches 1. This is because $P(|S_n - E[X]| > \epsilon) \to 0$ as $n \to \infty$ for all finite $\epsilon$. But this means $P(|S_n - E[X]| < \epsilon) \to 1$ as $n \to \infty$. Since this requires $E[X]$ negative and $S_n \to E[X]$, then we know $P(S_n < \frac{E[X]}{2}) \to 1$. 