Staff Solutions to In-Class Problems Week 13, Fri.

STAFF NOTE: Deviation, Chebyshev Bound, Ch. 19-19.4

Problem 1.
A herd of cows is stricken by an outbreak of *cold cow disease*. The disease lowers the normal body temperature of a cow, and a cow will die if its temperature goes below 90 degrees F. The disease epidemic is so intense that it lowered the average temperature of the herd to 85 degrees. Body temperatures as low as 70 degrees, but no lower, were actually found in the herd.

(a) Prove that at most 3/4 of the cows could have survived.

*Hint:* Let $T$ be the temperature of a random cow. Make use of Markov’s bound.

**Solution.** Notice that the result to be proved is a purely arithmetic fact about averages, not about probabilities. But we settle it by turning it into a question about random variables.

Namely, let $T$ be the temperature of a random cow. More precisely, the sample space for $T$ is the set of cows in the herd, the probabilities are defined to be uniform—the probability of each cow is $1/n$ where $n$ is the size of the herd—and $T(c)$ is the temperature of cow $c$. Since the probabilities are uniform, it follows immediately that

- the fraction of cows that survive is the probability that $T \geq 90$, and
- the average temperature of the herd equals $\mathbb{E}[T]$.

Applying Markov’s Bound to $T$:

$$\Pr[T \geq 90] \leq \frac{\mathbb{E}[T]}{90} = \frac{85}{90} = \frac{17}{18}.$$  

But $17/18 > 3/4$, so this bound is not good enough.

Instead, we apply Markov’s Bound to $T - 70$:

$$\Pr[T \geq 90] = \Pr[T - 70 \geq 20] \leq \frac{\mathbb{E}[T - 70]}{20} = \frac{85 - 70}{20} = \frac{3}{4}.$$

(b) Suppose there are 400 cows in the herd. Show that the bound of part (a) is the best possible by giving an example set of temperatures for the cows so that the average herd temperature is 85, and with probability 3/4, a randomly chosen cow will have a high enough temperature to survive.

**Solution.** Let 100 cows have temperature 70 degrees and 300 have 90 degrees. So the probability that a random cow has a high enough temperature to survive is exactly 3/4. Also, the mean temperature is

$$(1/4)70 + (3/4)90 = 85.$$  

So this distribution of temperatures satisfies the conditions under which the Markov bound implies that the probability of having a high enough temperature to survive cannot be larger than 3/4.
Problem 2.
A gambler plays 120 hands of draw poker, 60 hands of black jack, and 20 hands of stud poker per day. He wins a hand of draw poker with probability 1/6, a hand of black jack with probability 1/2, and a hand of stud poker with probability 1/5.

(a) What is the expected number of hands the gambler wins in a day?

Solution. \[ 120 \times \frac{1}{6} + 60 \times \frac{1}{2} + 20 \times \frac{1}{5} = 54. \]

(b) What would the Markov bound be on the probability that the gambler will win at least 108 hands on a given day?

Solution. The expected number of games won is 54, so by Markov, \( \Pr[R \geq 108] \leq \frac{54}{108} = 1/2. \)

(c) Assume the outcomes of the card games are pairwise, but possibly not mutually, independent. What is the variance in the number of hands won per day? You may answer with a numerical expression that is not completely evaluated.

Solution. Pairwise independence is sufficient to ensure additivity of variance. For an individual hand the variance is \( p(1-p) \) where \( p \) is the probability of winning. Therefore the variance is

\[
120(1/6)(5/6) + 60(1/2)(1/2) + 20(1/5)(4/5) = 523/15 = 34 \frac{13}{15}.
\]

(d) What would the Chebyshev bound be on the probability that the gambler will win at least 108 hands on a given day? You may answer with a numerical expression that is not completely evaluated.

Solution.

\[
\Pr[R \geq 108] = \Pr[R - 54 \geq 54] \leq \Pr[|R - 54| \geq 54] \leq \frac{\text{Var}[R]}{54^2} = \frac{523}{15(54)^2} \approx 0.01196.
\]

Problem 3.
The hat-check staff has had a long day serving at a party, and at the end of the party they simply return the \( n \) checked hats in a random way such that the probability that any particular person gets their own hat back is \( 1/n. \)

Let \( X_i \) be the indicator variable for the \( i \)th person getting their own hat back. Let \( S_n \) be the total number of people who get their own hat back.

(a) What is the expected number of people who get their own hat back?
Solution. $S_n = \sum_{i=1}^{n} X_i$, so by linearity of expectation,

$$\text{Ex}[S_n] = \sum_{i=1}^{n} \text{Ex}[X_i].$$

Since the probability a person gets their own hat back is $1/n$, therefore $\text{Pr}[X_i = 1] = 1/n$. Now, since $X_i$ is an indicator, we have $\text{Ex}[X_i] = 1/n$. By linearity of expectation,

$$\text{Ex}[S_n] = \sum_{i=1}^{n} \text{Ex}[X_i] = n \cdot \frac{1}{n} = 1.$$

(b) Write a simple formula for $\text{Ex}[X_i \cdot X_j]$ for $i \neq j$.

*Hint:* What is $\text{Pr}[X_j = 1 \mid X_i = 1]$?

Solution. We observed above that $\text{Pr}[X_i = 1] = 1/n$. Also, given that the $i$th person got their own hat, each other person has an equal chance of getting their own hat among the remaining $n-1$ hats. So

$$\text{Pr}[X_j = 1 \mid X_i = 1] = \frac{1}{n-1},$$

for $j \neq i$. Therefore,

$$\text{Pr}[X_i = 1 \text{ AND } X_j = 1] = \text{Pr}[X_j = 1 \mid X_i = 1] \cdot \text{Pr}[X_i = 1] = \frac{1}{n(n-1)}.$$

But $X_i = 1 \text{ AND } X_j = 1$ iff $X_i X_j = 1$, so

$$\text{Ex}[X_i X_j] = \text{Pr}[X_i X_j = 1] = \text{Pr}[X_i = 1 \text{ AND } X_j = 1],$$

and hence

$$\text{Ex}[X_i X_j] = \frac{1}{n(n-1)}.$$

(e) Explain why you cannot use the variance of sums formula to calculate $\text{Var}[S_n]$.

Solution. The principle of additivity of variances requires the variables be pairwise independent, but the indicator variables for people getting their hats back are not pairwise independent, since $\text{Pr}[X_j = 1 \mid X_i = 1] = 1/(n-1) \neq 1/n = \text{Pr}[X_j = 1]$ for $i \neq j$.

(d) Show that $\text{Ex}[(S_n)^2] = 2$. *Hint:* $(X_i)^2 = X_i$.
Solution.

\[
\text{Ex}[S_n^2] = \text{Ex} \left[ \sum_{i=1}^{n} (X_i)^2 + 2 \sum_{1 \leq i < j \leq n} X_i X_j \right] \quad \text{(expanding the sum for } S_n) \\
= \sum_{i=1}^{n} \text{Ex}[(X_i)^2] + 2 \sum_{1 \leq i < j \leq n} \text{Ex}[X_i X_j] \quad \text{(linearity of } \text{Ex}[\cdot]) \\
= \sum_{i=1}^{n} \text{Ex}[X_i] + 2 \sum_{1 \leq i < j \leq n} \frac{1}{n(n-1)} \quad \text{(since } (X_i)^2 = X_i) \\
= n \cdot \frac{1}{n} + 2 \left( \frac{n}{2} \right) \frac{1}{n(n-1)} \\
= 1 + \frac{1}{n-1} = 2. 
\]

(e) What is the variance of \( S_n \)?

Solution.

\[
\text{Var}[S_n] = \text{Ex}[(S_n)^2] - \text{Ex}^2[S_n] = 2 - 1^2 = 1. 
\]

(f) Show that there is at most a 1% chance that more than 10 people get their own hat back.

Solution.

\[
\text{Pr}[S_n \geq 11] = \text{Pr}[S_n - \text{Ex}[S_n] \geq 11 - \text{Ex}[S_n]] \\
= \text{Pr}[S_n - \text{Ex}[S_n] \geq 10] \\
\leq \text{Pr}[[S_n - \text{Ex}[S_n]] \geq 10] \\
\leq \frac{\text{Var}[S_n]}{10^2} = .01 
\]

Supplementary Problem

Problem 4.
For any random variable, \( R \), with mean, \( \mu \), and standard deviation, \( \sigma \), the Chebyshev Bound says that for any real number \( x > 0 \),

\[
\text{Pr}[[R - \mu] \geq x] \leq \left( \frac{\sigma}{x} \right)^2. 
\]

Show that for any real number, \( \mu \), and real numbers \( x \geq \sigma > 0 \), there is an \( R \) for which the Chebyshev Bound is tight, that is,

\[
\text{Pr}[[R] \geq x] = \left( \frac{\sigma}{x} \right)^2. \tag{1}
\]

Hint: First assume \( \mu = 0 \) and let \( R \) only take values 0, \(-x\), and \( x \).
Solution. From the hint, we aim to find an $R$ with $\text{Ex}[R] = 0$ and $\text{Var}[R] = \sigma^2$ that satisfies equation (1).

Using the further hint that $R$ takes only values $0, -x, x$, we have

$$0 = \text{Ex}[R] = x \Pr[R=x] - x \Pr[R=-x] = x (\Pr[R=x] - \Pr[R=-x])$$

so

$$\Pr[R=x] = \Pr[R=-x],$$

since $x > 0$. Also,

$$\sigma^2 = \text{Ex}[R^2] = x^2 \Pr[R=-x] + x^2 \Pr[R=x] = 2x^2 \Pr[R=x],$$

so

$$\Pr[R=x] = \frac{\sigma^2}{2x^2}.$$ 

This implies

$$\Pr[R = 0] = 1 - 2 \Pr[R = x] = 1 - \left( \frac{\sigma}{x} \right)^2,$$

which completely determines the distribution of $R$. Moreover,

$$\Pr[|R| \geq x] = \Pr[R = -x] + \Pr[R = x] = 2 \Pr[R = x] = \left( \frac{\sigma}{x} \right)^2$$

which confirms (1).

Finally, given $\mu, x$, and $\sigma$, define $R' := R + \mu$. Since $\text{Var}[R'] = \text{Var}[R]$, the random variable $R'$ will be the desired random variable with mean $\mu$ and standard deviation $\sigma$ for which the Chebyshev Bound is tight. \hfill \blacksquare