1. **Warm-Up (Markov's Inequality)**

Let X, Y, Z be non-negative random variables such that $Z = X + 10Y$.

(a) Using Markov's inequality, upper bound the chance that X takes on a value higher than $k\mu$, where $\mu = \mathbb{E}(X)$.
$\frac{1}{k}$.

(b) If $\mathbb{E}(X) = 10$ and $\mathbb{E}(Y) = 4$, upper bound $\Pr[Z \geq 100]$.
$\frac{1}{2}$.

(c) If $\mathbb{E}(Z) = 100$ and $\mathbb{E}(Y) = 4$, upper bound $\Pr[X \geq 10]$.
By Markov's inequality, 6; by rules of probability, 1.

(d) What does this tell you about Markov's inequality?
Reasonable bound above mean, poor bound below mean.

(e) Suppose $Y = (X - \mu)^2$, where $\mu = \mathbb{E}(X)$. Using Markov's inequality, upper bound the chance that Y takes on a value higher than α^2.
$\frac{Var(X)}{\alpha^2}$. This is the proof of Chebyshev's inequality.

2. **Getting Warmer (Chebyshev)**

Suppose we want to use a low-cost thermometer for measuring the weather. The thermometer has a standard deviation of 2 degrees from the actual temperature.

(a) Upper bound the chance that the thermometer is more than 5 degrees from the actual temperature.
$\frac{4}{25}$

(b) If we had two identical thermometers, bound the chance that at least one thermometer stays within 5 degrees of the actual temperature.
$\geq 1 - \frac{16}{625}$.

(c) If we had n identical thermometers, each with standard deviation σ degrees from the actual temperature, lower bound the chance that at least one thermometer stays within k degrees of the actual temperature.
$\geq 1 - (\frac{\sigma}{k})^{2n}$.

3. **Summertime**

You're a DJ for a jazz club and you have r songs, each of which lasts 5 minutes. You can't attend work tonight, so your assistant randomly selects songs (with replacement) before the live band arrives.

(a) The live band arrives 2 hours after the club opens. What's the chance that at least one song is repeated?
There are $\frac{120}{5} = 24$ songs played. There are r^{24} possible arrangements of songs, but $\frac{r!}{(r-24)!}$ ways to order 24 songs without repeats (ordering with no replacement). Thus, there is a $1 - \frac{r!}{r^{24}(r-24)!}$ chance of having at least one repeat song. (Of course, if $r < 24$, the chance is 100%.)

(b) What's the expected number of repeats? A repeat is when you hear a song on two different occasions. (Hint: think about collisions)

Note: A repeat is defined pairwise. For example, if a song is played four times, there are 6 repeats.

Consider 24 balls (time slots), r bins (unique songs). Let X_{ij} be indicator for a song played at both time slot i, j. (X_{ij} counts if there is a repeat at time slots i and j.) Then, $\Pr[X_{ij} = 1] = \mathbb{E}(X_{ij}) = \frac{1}{r}$. There are $\binom{24}{2}$ of these variables (this is the number of ways to choose two time slots i, j). The number of repeats is $X = \sum_{1 \le i \le j \le 24} X_{ij}$. By linearity of expectation, there are $\binom{24}{2} \times \frac{1}{r}$ expected repeats.

(c) Let's define X to be the number of repeats. At what threshhold of repeats c can you ensure that the probability of having more than c repeats is less than 10%? Assume you have the answer to part (b).

By Markov's inequality, $\Pr[X \ge c] \le \frac{\mathbb{E}(X)}{c}$. Thus, $\frac{\mathbb{E}(X)}{c} = .1$, so $c = \frac{\mathbb{E}(X)}{.1} = \frac{\binom{24}{2}}{.1r} = \frac{2760}{r}$. If you had 2760 songs, your chance of 1 repeat would be less than 10%. If you had 276 songs, your chance of 10 repeats would be less than 10%.

4. Hard Summer

We asked 500 people to independently rate this year's Hard Summer on a scale from 0 to 10. The poll average was an 8.5. Upper bound the variance of the poll. What's the probability that the poll is accurate to within an error of 0.4?

Chebyshev, $\delta = \Pr[|\hat{p} - p| \ge \varepsilon] \le \frac{Var(\hat{p})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$. Since σ^2 maximized when 50% chance of 0 and 50% chance of 10, $\sigma^2 \le 25$, and the poll variance is less than $25/500 = 0.05$. So, $\delta \le \frac{0.05}{0.4^2} = \frac{1}{3.2} \approx 32\%$. We are at least 68% sure that the poll is accurate to within 0.4 (that the average is between an 8.1 and 8.9).