Proving Markov's Inequality

Markov's inequality: for any nonnegative random variable X, and for any $t > 0$,

$$\mathbf{Pr}[X \geq t] \leq \frac{\mathbf{E}[X]}{t}.$$

Let's say that X can take values $x_1 < x_2 < \ldots < x_j = t < \ldots < x_n$. First, prove the inequality.

Proof:

$$\mathbf{E}[X] = \sum_{i=1}^{n} x_i * \mathbf{Pr}[X = x_i] \geq \sum_{i=j}^{n} x_i * \mathbf{Pr}[X = x_i] \geq \sum_{i=j}^{n} t * \mathbf{Pr}[X = x_i]$$

Then, show the equivalence to the sometimes more useful form, for $s > 0$.

$$\mathbf{Pr}[X \geq s \cdot \mathbf{E}[X]\,] \leq \frac{1}{s}.$$

Proof: let $t = s\mathbf{E}[X]$.

Finally, invent a random variable and a distribution such that,

$$\mathbf{Pr}[X \geq 10 \cdot \mathbf{E}[X]\,] = \frac{1}{10}.$$

Answer: Consider Bernoulli(1, 1/10). So, getting 1 w.p 1/10 and 0 w.p 9/10. This importantly shows that Markov's inequality is tight, because we could replace 10 with t and use Bernoulli(1, 1/t), at least with $t \geq 1$.

Proving the Chebyshev Inequality.

1. For any random variable X and scalars $t, a \in \mathbb{R}$ with $t > 0$, convince yourself that

$$\mathbf{Pr}[\,|X - a| \geq t] = \mathbf{Pr}[\,(X - a)^2 \geq t^2]$$

2. Use the second form of Markov's inequality and (1) to prove *Chebyshev's Inequality*: for any random variable X with $\mathbf{E}[X] = \mu$ and var$(X) = c^2$, and any scalar $t > 0$,

$$\mathbf{Pr}[\,|X - \mu| \geq tc] \leq \frac{1}{t^2}.$$

(*Hint:* To use Markov, don't forget the definition of variance... var$(X) = \mathbf{E}[(X - \mu)^2]$.)

Proving the Weak Law of Large Numbers.

Muse about the following (no calculations for stuff in italics). *If you flip an unbiased coin n times how many heads and tails do you expect to get? What if you flip it even more times, so n gets really large? What if you take the average of n rolls of a fair six-sided die?*

Consider a sequence $X_1, X_2, \ldots, X_n$ of independent, identically distributed random variables (think coins/dice), each with mean μ and variance σ^2. Define the **sample mean** by

$$M_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

1. Show that $\mathbf{E}[M_n] = \mu$ and $\text{var}(M_n) = \dfrac{\sigma^2}{n}$ (where these μ and σ^2 are mean/variance of the X_i).
   Proof: The mean follows from linearity of expectation. The variance follows because $var(aX) = a^2 var(x)$ always, and for independent RVs $var(X+Y) = var(X) + var(Y)$.

2. Prove the *Weak Law of Large Numbers:* for any deviation parameter $\epsilon > 0$,

   $$\mathbf{Pr}[\,|M_n - \mu| \geq \epsilon] \to 0, \qquad \text{as } n \to \infty.$$

   (Hint: Use Chebyshev's inequality.)
   Proof: We need to use the freedom that comes with the inequality holding for any positive t. In particular, set $t = \frac{\epsilon}{\sqrt{var(M_n)}} = \frac{\epsilon\sqrt{n}}{\sigma}$. Then we can apply Chebyshev to show

   $$\mathbf{Pr}[\,|M_n - \mu| \geq \epsilon] \leq \frac{\sigma^2}{n\epsilon^2},$$

   which indeed goes to 0 as n grows large!

   Interpret this statement in terms of real life experiments, and explain how this relates to the questions at the beginning of this problem.
   The idea is that no matter how small you set ϵ, if you gather enough samples, you will still get closer and closer to the mean. This is a really nice example of how intuition and quantitative bounds coincide in probability theory.

Using Chebyshev and the Weak Law of Large Numbers.

1. Let X be uniformly distributed in the interval $[0, 4]$ and note that $\mathbf{E}[X] = 2$. Use Markov's inequality to compute upper bounds on

   $$\mathbf{Pr}[X \geq 2] \leq \qquad \mathbf{Pr}[X \geq 3] \leq \qquad \mathbf{Pr}[X \geq 4] \leq$$

   Now, compute the probabilities directly, and compare them to the upper bounds.

   $$\mathbf{Pr}[X \geq 2] = \qquad \mathbf{Pr}[X \geq 3] = \qquad Pr[X \geq 4] =$$

   (The point is that sometimes Markov's gives exact bounds (above), but other times they are loose.)
   The answers are the in BT book.

2. A fair coin is tossed 100 times.

   (a) Calculate the expected number of heads and calculate the standard deviation for the number of heads.

   (b) What does Chebyshev's Inequality tell you about the probability that the number of heads that turn up deviates from the expected number 50 by three or more standard deviations (*i.e.*, by at least $3\sigma = 15$)?

   This is a good practice question.. let me know if you have questions.

3. A \$1 bet on craps has an expected winning of $\$ - .0141$. What does the Weak Law of Large Numbers say about your winnings if you make a large number of \$1 bets at the craps table? Does it assure you that your losses will be small? Does it assure you that if n is very large you will lose?
   This is a good practice question.. let me know if you have questions.

4. Page 270 of the BT book has another nice, real-life example about polling voters. In particular, they give bounds on the number of voters you need to poll to get an "accurate" estimate of the overall community's opinions.