Math 340 HW 4

Asa Royal (ajr74) [collaborators: none]

February 29, 2024

1. Meester 2.3.28

   Prove that Markov's inequality follows from theorem 2.3.5

   Proof. Theorem 2.3.25 states that for a positive-valued r.v. Y and $b > 0$,

   $$\mathbb{P}(Y \geq b) \leq \frac{1}{b}\mathbb{E}[Y] \tag{1}$$

   Assume $Y = |X|^k$ for a positive-valued r.v. X and $b = a^k$ □

   Then

   $$\mathbb{P}(|X|^k \geq a^k) \leq \frac{1}{a^k}\mathbb{E}[|X|^k]$$

   And since $|X|^k \geq a^k \Leftrightarrow |X| \geq a$,

   $$\mathbb{P}(|X| \geq a) = \frac{1}{a^k}\mathbb{E}[|X|^k] \tag{2}$$

   Prove that Chebyshev's inequality follows from theorem 2.3.5

   Proof. Theorem 2.3.25 states that for a positive-valued r.v. Y and $b > 0$,

   $$\mathbb{P}(Y \geq b) \leq \frac{1}{b}\mathbb{E}[Y] \tag{3}$$

   Assume $Y = \mathrm{Var}(X)$ for a positive-valued r.v. X and $b = a^2$ Then

   $$\mathbb{P}(\mathrm{Var}(X) \geq a^2) \leq \frac{1}{a^2}\mathrm{Var}(X) \tag{4}$$

   Integrating the definition of Var(X) and noting that $\forall m, m^2 = |m|^2$, we find

   $$\mathbb{P}((X - \mathbb{E}[X])^2 \geq a^2) = \mathbb{P}(|X - \mathbb{E}[X]|^2 \geq a^2) \leq \frac{1}{a^2}\mathrm{Var}(X) \tag{5}$$

   And once again, since for any event A, $\mathbb{P}(A)^2 \geq q^2 \Leftrightarrow \mathbb{P}(A) \geq q$

   $$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{1}{a^2}\mathrm{Var}(X) \tag{6}$$

   □

2. Meester 2.7.15

3. Suppose X is a discrete random variable.

   (i) Prove that $\forall x f(x), \geq g(x) \Rightarrow \mathbb{E}[f(X)] \geq \mathbb{E}[g(X)]$, assuming these are well-defined.

   Proof. $\mathbb{E}[f(X)] = \sum_{x \in R(X)} f(x)\mathbb{P}(X = x)$ and $\mathbb{E}[g(X)] = \sum_{x \in R(X)} g(x)\mathbb{P}(X = x)$. So the following are equivalent.

   $$\mathbb{E}[f(X)] \stackrel{?}{=} \mathbb{E}[g(X)] \tag{1}$$

   $$\sum_{x \in R(X)} f(x)\mathbb{P}(X = x) \stackrel{?}{=} \sum_{x \in R(X)} g(x)\mathbb{P}(X = x) \tag{2}$$

   $$\sum_{x \in R(X)} f(x) \stackrel{?}{=} \sum_{x \in R(X)} g(x) \tag{3}$$

   We know that $\forall x, f(x) \geq g(x)$, so the operator in (1), (2), and (3) must be $\geq$. □

(ii) Suppose that $f(x) : \mathbb{R} \mapsto \mathbb{R}$ is differentiable. Supose $\mathbb{E}[X] =; u$. Let

(iii) ...

Proof. $\mathbb{E}[f(X)]$ is given by the tangent line approximation to f, l; $\mathbb{E}[X] = \mu$, so $f(\mathbb{E}[X]) = f(\mu)$. Thus, the following are equivalent

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]) \tag{1}$$

$$\ell(x) \geq f(\mu) \tag{2}$$

$$f(\mu) + f'(\mu)(x - \mu) \geq f(\mu) \tag{3}$$

$$f'(\mu)(x - \mu) \geq 0 \tag{4}$$

□

We assumed that $f(x) > \ell(x)$ except at $;u$, which means the tangent line must be downward sloping at μ. Equivalent, $f'(\mu) < 0$.

4. In a box there are n identical marbles, labeled $1, \ldots, n$. There are n people who take turns drawing a marble from the box, with replacement. Let X_n be the nubmer of marbles that were not drawn by anyone.

(i) Compute $\mathbb{E}[\frac{1}{n}X_n$, the expected fraction of marbles not chosen.

Let χ_i represent an indictaor function for the event that marble i was not drawn by anyone. Then by linearity and the method of indicators,

$$\mathbb{E}\left[\frac{1}{n}X_n\right] = \frac{1}{n}\mathbb{E}\left[X_n\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\chi_i\right] = \left(\frac{1}{n}\right)(n)\left(\mathbb{E}\left[\chi_i\right]\right) = \mathbb{E}[\chi_i]$$

The expected value of an indicator fuction is the probability of its underlying event. Since each marble is equally likely to be drawn and rwas are independent, the probability that any given marble was not drawn is $(n - 1/n)^n$. So

$$\mathbb{E}\left[\frac{1}{n}X_n\right] = \left(\frac{n-1}{n}\right)^n$$

(ii) What is $\lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n}X_n\right]$?

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n}X_n\right] = \lim_{n\to\infty}\left(\frac{n-1}{n}\right)^n = \lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n = e^{-1}$$

(iii) What is $\mathrm{Var}\left(\frac{1}{n}X_n\right)$? The χ_i are not independent, because if we know that marble j was not chosen, that reduces the likelihood that marble k was not chosen. Thus,w e need to take into account covariance when calculating the variance of the sum. How do I think about calculating covariance between the indicators, though?

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{n}X_n\right) &= \left(\frac{1}{n}\right)^2 \mathrm{Var}(\chi_1 + \ldots \chi_n) && \text{linearity, inner product}\\ &= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{Var}(\chi_i) + 2\sum_{\substack{i<j\\ j<n}}\mathrm{Cov}(\chi_i, \chi_j)\right) && \text{var of sum}\end{aligned}$$

5. .. **Using Chebychev's inequality:**

Chebychev's inequality states that

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{1}{a^2}\mathrm{Var}(X)$$

Let Y be a random variable denoting the number of heads we toss in 10,000 trials. We wish to bound the probability that $Y \geq 5000$. $\mathbb{E}[Y] = 5{,}000$, so we can express $\mathbb{P}(Y \geq 5000)$ as $\mathbb{P}(Y - \mathbb{E}[Y] \geq 500$. Per Chebychev's inequality,

$$\mathbb{P}(|Y - \mathbb{E}[Y]| \geq 500 \leq \frac{1}{500^2}\mathrm{Var}(Y) \tag{5}$$

Y can be represented as the sum of 10,000 indicator functions for the event of each coin flip. Thus, by linearity $\text{Var}(Y) = np(1-p)$. For a fair coin with 10000 flips, $\text{Var}(Y) = (0.5)(0.5)(10000) = 2500$. Plugging this into (5), we see

$$\begin{aligned} \mathbb{P}(Y \geq 5000) &= \mathbb{P}(|Y - \mathbb{E}[Y] \geq 500) & (1) \\ &\leq \left(\frac{1}{500^2}\right) 2500 = 0.01 & (2) \end{aligned}$$

Using the law of large numbers

One version of the law of large numbers states that

$$\mathbb{P}\left(\bigcup_{k \geq n\left(\frac{1}{2}+\varepsilon\right)} A_{k,n}\right) \leq e^{-\varepsilon^2 n} \quad (3)$$

In our 10,000 fair coin toss case, we use $\varepsilon = 1/10$ to find that

$$\mathbb{P}(Y \geq 5500) = \mathbb{P}\left(\bigcup_{k \geq 5500 A_{k,10000}}\right) \leq e^{-(0.1)^2 10000} = e^{-100} \quad (4)$$

The law of large numbers provides an upper probability bound of $\mathbb{P}(Y \geq 500) \leq e^{-100}$. Chebychev's inequality provides a probability bound of $\mathbb{P}(Y \geq 500) \leq 0.01$. Clearly, the law of large numbers provides a tighter bound.

6. Suppose that every timme you shop at a certain store, there is a small randomly selected prize that comes with you rpurchase. Suppose there are n different prizes that you could win, all equally likely. It is possible that you get the same prize multiple times. Let X_n be the number of visits you make until you have won all n distinct prizes. Calculate $\mathbb{E}[X_n]$ by

   (i) How many visits N_1 are needed to win one prize?
   1.

   (ii) Let N_2 be the number of add'l visits until you get a second unique prize. What is the distribution of N_2?

   $$N_2 \sim \text{Geo}\left(\frac{n-1}{n}\right)$$

   (iii) What is the distribution of N_{k+1}? k prizes have already been picked, so the probability of "success" on any given visit to the shop is $(n-k)/(n)$, since there are $n-k$ unique prizes we still need to collect. Thus

   $$N_{k+1} \sim \text{Geo}\left(\frac{n-k}{n}\right)$$

   (iv) How is X_n related to the random variables N_k?

   $$X_n = \sum_{k=1}^{n} N_k$$

Calculating $\mathbb{E}[X_n]$:

Per part iv,

$$
\begin{aligned}
\mathbb{E}[X_n] &= \mathbb{E}\left[\sum_{k=1}^{n} N_k\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[N_k\right] && \text{by linearity} \\
&= \mathbb{E}[N_1] + \sum_{k=2}^{n} \mathbb{E}\left[N_k\right] && \text{split up sum} \\
&= \mathbb{E}[N_1] + \sum_{k=1}^{n-1} \mathbb{E}\left[N_{k+1}\right] && \text{adjust summation bounds} \\
&= 1 + \sum_{k=1}^{n-1} \frac{1}{\left(\frac{n-k}{n}\right)} && \mathbb{E}\text{ of geo r.v.} \\
&= 1 + \sum_{k=1}^{n-1} \frac{n}{n-k}
\end{aligned}
$$