Homework 3

1. (book #3.1)
Let $X \sim \text{Unif}(\{1, \dots, n\})$. Find $\text{Var}(X)$.

Hint: Recall the identies that we used when discussing #2.1 in class.

Evaluate your answer for $n = 17$, to find $\text{Var}(X)$ for $X \sim \text{Unif}(\{1, \dots, 17\})$.

$\text{Var}(X) =$

2. (book # 3.2)
Let $Y \sim \text{Unif}(\{-k, \dots, k\})$. Find $\text{Var}(Y)$.
Compare with 3.1, when $n = 2k + 1$.

Evaluate your answer for $k = 8$, to find $\text{Var}(Y)$ for $Y \sim \text{Unif}(\{-8, \dots, 8\})$,

$\text{Var}(Y) =$

3. (book #3.3)
Suppose that we roll a standard fair die $n = 100$ times. Let X be the sum of the numbers that appear over the 100 rolls. Use Chebyshev's inequality to bound $\Pr(|X - 350| \geq 50)$.

$\Pr(|X - 350| \geq 50) \leq$

4. (book #3.4)
Prove that, for any real number c and any discrete random variable X, $\text{Var}[cX] = c^2\text{Var}(X)$.

Which of the following arguments proves the claim?

(a) $\text{Var}(cX) = E\{(cX - E(cX))^2\} = E\{(cX)^2\} - \{E(cX)\}^2 = c^2E(X^2) - \{cE(X)\}^2 = c^2E(X^2) - c^2(EX)^2 = c^2\text{Var}(X)$
(b) Variance is an expectation $\implies \text{Var}(cX) = c^2\text{Var}(X)$.
(c) Use $\text{Var}(cX) = E(cX^2) - (E(cX))^2 = cE(X^2) - \{c\,E(X)\}^2 = c^2\text{Var}(X)$.
(d) $\text{Var}(cX) = \text{Var}(c)\text{Var}(X) = c^2\text{Var}(X)$.
(e) none of these

5. (book # 3.5)
Given any two random variables X and Y, by the linearity of expectations we have $E[X - Y] = E[X] - E[Y]$. Prove that, when X and Y are independent, $\text{Var}[X - Y] = \text{Var}[X] + \text{Var}[Y]$.

Which of the following arguments proves the claim?

(a) $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$ for *any* two r.v.'s $X, Y \implies \text{Var}(X + (-1)Y) = \text{Var}(X) + \text{Var}(Y)$.
(b) $\text{Var}(X - Y) = E(X^2) - E(Y^2) = E(X^2) + E(-Y^2) = \text{Var}(X) + \text{Var}(-Y) = \text{Var}(X) + \text{Var}(Y)$
(c) $\text{Var}(X - Y) = \text{Var}(X + Y) - 2\text{Var}(Y) = \text{Var}(X) + \text{Var}(-Y) = \text{Var}(X) + \text{Var}(Y)$.
(d) By Corollary 3.4 $\text{Var}(X + (-Y)) = \text{Var}(X) + \text{Var}(-Y) = \text{Var}(X) + \text{Var}(Y)$. The last by #3.4.
(e) none of these

6. (book # 3.6)
For a coin that comes up heads independently with probability p on each flip, what is the variance in the number of flips X until the kth head appears? (counting all flips, *including* the k-th head).

Hint: Recall problems #2.14 and 2.15 from the last homework. You may use the variance of geometric r.v. $Y \sim \text{Geom}(p)$ as $\text{Var}(Y) = \frac{1-p}{p^2}$.

Evaluate your answer $\text{Var}(X)$ for $p = 0.4$ and $k = 5$,

$\boxed{\text{Var}(X) = \qquad}$

7. (book # 3.7)
A simple model of the stock market suggests that, each day, a stock with price x will increase by a factor $r > 1$ to xr with probability p and will fall to x/r with probability $q = 1 - p$. Assuming we start with a stock with price $X_0 = 1$, find a formula for the expected value $E(X_d)$ and the variance $\text{Var}(X_d)$ of the price X_d of the stock after d days. That is, X_d is the stock price after d changes (increase or decrease).

Hint: Let Y = number of days when the price increases. Then $Y \sim \text{Bin}(d, p)$ and $X_d = r^Y \cdot (1/r)^{d-Y}$. To evaluate the expression for $E(X_d)$, use the binomial theorem (see lecture #3.4). You can state the binomial theorem as

$$\sum_{\ell=0}^{k} \binom{k}{\ell} a^\ell b^{n-\ell} = (a+b)^n$$

(why?).

7a. Evaluate your solution for $E(X_d)$ for $d = 60$, $r = 1.05$ and $p = 0.6$,

$\boxed{E(X_d) = \qquad}$

7b. Evaluate your solution for $\text{Var}(X_d)$ for $d = 60$, $r = 1.05$ and $p = 0.6$,

$\boxed{\text{Var}(X_d) = \qquad}$

8. (book #3.8)
Suppose that we have an algorithm that takes as input a string of n bits. We are told that the expected running time is $O(n^2)$ if the input bits are chosen independently and uniformly at random. That is, letting X_n denote the running time with an input of size n, $E(X_n) \leq Mn^2$ for $n \geq n_0$ and some $M > 0$ and $\Pr(X_n = x_n) \geq 2^{-n}$ ($\geq$, since more than one sequence might have a given running time x_n).

What can Markov's inequality tell us about the worst-case running time $x_n^\star$ of this algorithm on inputs of size n? Nothing to turn in.

9. Let $X \sim \text{Bin}(n, p)$ be a binomial r.v. Recall the derivation of $\text{Var}(X)$ in the lecture. We used $\text{Var}(X) = E(X^2) - (EX)^2$, and used the binomial theorem to evaluate $E(X^2)$. Alternatively, one could use the representation of $X = \sum_{i=1}^n Y_i$ of X as sum of Bernoulli r.v.'s, $Y_i \sim \text{Bern}(p)$, independently. Use this representation and Corollary 3.4 from the lecture to find $\text{Var}(X)$.

Which of the following arguments proves the claim? (if multiple choices are correct, mark any of those)

(a) $E(X^2) = \sum_{i=1}^n E(Y_i^2) = np^2$ and $EX = \sum E(Y_i) = np \implies \text{Var}(X) = E(X^2) - (EX)^2 = np^2 - (np)^2 = np^2(1-n)$

(b) By Corollary 3.4 $E(X^2) = \sum_{i=1}^n E(Y_i^2) = np$, and $EX = \sum E(Y_i) = np \implies \text{Var}(X) = E(X^2) - (EX)^2 = np - (np)^2 = np^2(1-n)$

(c) $\text{Var}(X) = \sum_{i=1}^n \text{Var}(Y_i)$, by Corollary 3.4, since the Y_i are independent. $\text{Var}(Y_i) = E(Y_i^2) - (EY_i)^2 = p - p^2 = p(1-p)$, and therefore $\text{Var}(X) = np(1-p)$

(d) $\text{Var}(X) = \sum_{i=1}^n \text{Var}(Y_i) + 2\sum_{i<j} \text{Cov}(Y_i, Y_j) = np(1-p) + n(n-1)p$.

(e) none of these

10. Consider three r.v.'s X_1, X_2, X_3 with finite sample space $X_i \in \{1, \ldots, K\}$. The sequence $X_i, i = 1, 2, 3$, is called a Markov chain if

$$\begin{aligned} \Pr(X_3 = i \mid X_2 = j, X_1 = k) &= \Pr(X_3 = i \mid X_2 = j) \\ &= \Pr(X_2 = i \mid X_1 = j). \end{aligned} \quad (1)$$

That is, $\Pr(X_3 = i \mid X_2 = j) = P_{ji}$ does not depend on X_1, and it is the same as $\Pr(X_2 = i \mid X_1 = j)$. We call P_{ji} the transition probabilities.

Let $P = [P_{ji}]$ denote the $(K \times K)$ matrix of transition probabilties.

10a. Show $\Pr(X_3 = i \mid X_1 = j) = [P^2]_{ji}$ (the (j, i) element of P^2).

Hint: Use the law of total probability with $E_k = \{X_2 = k\}$. Recall that any result for probabilities, like the law of total probability, is also true for conditional probabilities, like $\Pr(\bullet \mid X_1 = j)$.

Let LHS denote $\Pr(X_3 = i \mid X_1 = j)$. Which of the following arguments shows the claim?

(a) $LHS = \sum_{k=1}^K \Pr(X_2 = k \mid X_1 = j) \Pr(X_3 = i \mid X_2 = k, X_1 = j) = \sum_k P_{jk} P_{ki} = [P^2]_{ji}$

(b) $LHS = \sum_{k=1}^K \Pr(X_3 = k \mid X_1 = j) = [P^2]_{ji}$

(c) $LHS = \Pr(X_2 = i \mid X_1 = j) \cdot \Pr(X_3 = i \mid X_2 = j) = [P^2]_{ij}$

(d) $LHS = \Pr(X_2 = i \mid X_1 = j) \cdot \Pr(X_3 = i \mid X_1 = j) = [P^2]_{ij}$

(e) none of these

Let $q_{1i} = \Pr(X_1 = i)$ denote the marginal probabilities for X_1, $\boldsymbol{q}_1 = (q_{11}, \ldots, q_{1K})$ (a $(1 \times K)$ row vector), and similarly for $\boldsymbol{q}_2$ and $\boldsymbol{q}_3$.

10b. Show $\boldsymbol{q}_2 = \boldsymbol{q}_1 P$ and $\boldsymbol{q}_3 = \boldsymbol{q}_1 P^2$

Which of the following arguments shows the claim $\boldsymbol{q}_2 = \boldsymbol{q}_1 P$?

(a) $q_{2i} = \Pr(X_2 = i) = \Pr(X_1 = j, X_2 = i)/\Pr(X_i = j) = \Pr(X_2 = i \mid X_1 = j) = P_{ji}$

(b) $q_{2i} = \sum_j p_{X_1}(j)\, p_{X_2|X_1}(i \mid X_1 = j) = \sum_j q_{1j} P_{ji}$

(c) $q_{2i} = \Pr(X_2 = i) = \Pr(X_1 = j \mid X_2 = i)/[\Pr(X_2 = i)\Pr(X_2 = i \mid X_1 = j)]$

(d) $q_{2i} = \Pr(X_2 = i) = \sum_j \Pr(X_1 = j, X_2 = i)$

(e) none of these

10c. Let $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)'$ be a probability vector (i.e., $\pi_k \geq 0$ and $\sum \pi_k = 1$) with

$$\pi_k P_{kj} = \pi_j P_{jk}$$

for any pair of states j and k. If $\boldsymbol{q}_1 = \boldsymbol{\pi}$, show that $\boldsymbol{q}_2 = \boldsymbol{q}_3 = \boldsymbol{\pi}$ as well (π is called an "equilibrium distribution").

(a) By the law of total probability $q_{2i} = \sum_j \pi_j P_{ji} = \sum_j \pi_i P_{ij} = \pi_i \sum_j P_{ij} = \pi_i$

(b) By Bayes' theorem $q_{2i} = \frac{\pi_k P_{ki}}{\sum_\ell \pi_k P_{k\ell}} = \frac{\pi_j P_{kj}}{\sum_\ell \pi_k P_{k\ell}} \implies q_{2i} = \pi_i$

(c) $\boldsymbol{q}_1$ is a probability vector and P is a stochastic matrix $\implies \boldsymbol{q}_2 = \boldsymbol{q}_3 = \boldsymbol{q}_1$.

(d) By definition of conditional probability $P_{kj} = \frac{\Pr(X_1 = k, X_2 = j)}{\Pr(X_1 = k)} = \frac{\pi_k P_{kj}}{\pi_k} = \frac{\pi_j P_{jk}}{\pi_k} = \pi_j$.

(e) none of these

Similar definitions are used for a sequence of random variables X_t, $t = 1, 2, 3, \ldots$. See chapter 7 in the book. Definition 7.1 is the general version of (1); equation (7.1) is similar to (b) above; Definition 7.8 defines an equilibrium distribution; and Theorem 7.10 is (c).