Math 340 HW 4

Asa Royal (ajr74) [collaborators: none]

March 1, 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 $\mathbb{E}[X] = \mu$. Let $\ell(x)$ be the line tangent to the graph of f at $(\mu, f(\mu)$. Suppose the graph of f lies above the graph of ℓ everywhere except the point of tangency. Prove that $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$.

Proof. Since the graph of f lies above the graph of l everywhere except where they touch (at $x = \mu$), f is a convex function, and $(\mu, f(\mu))$ is its global minimum. Thus for all values $x \in R(X), \mathbb{E}[f(X)] \geq f(\mu) = f(\mathbb{E}[X]$. □

In particular, this conclusion applies to $f(x) = e^x$, implying that $\mathbb{E}[e^X] \geq e^{\mathbb{E}[X]}$.

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

$$\mathbb{E}\left[n\left(\frac{n-1}{n}\right)^g n1nX_n\right] = \left(\frac{n-1}{n}\right)^n \tag{4}$$

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

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

Per (4), $\mathbb{E}[X_n] = n\left(\frac{n-1}{n}\right)^n$, so $\mathbb{E}[X_n]^2 = n^2\left(\frac{n-1}{n}\right)^{2n}$

(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?

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{n}X_n\right) &= \left(\frac{1}{n}\right)^2\mathrm{Var}(X_n) && (1)\\
&= \frac{1}{n^2}\mathbb{E}[X^2] - \mathbb{E}[X]^2 && (2)
\end{aligned}$$

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,

$$\mathbb{P}(|Y - \mathbb{E}[Y]| \geq 500 \leq \frac{1}{500^2}\mathrm{Var}(Y) \tag{3}$$

Y can be represented as the sum of 10,000 indicator functions for the event of each coin flip. Thus, by linearity $\mathrm{Var}(Y) = np(1-p)$. For a fair coin with 10000 flips, $\mathrm{Var}(Y) = (0.5)(0.5)(10000) = 2500$. Plugging this into (3), 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

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

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

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

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 purchase. 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,

$$\begin{aligned}
\mathbb{E}[X_n] &= \mathbb{E}\left[\sum_{k=1}^{n} N_k\right] \\
&= \sum_{k=1}^{n} \mathbb{E}\left[N_k\right] && \text{by linearity} \\
&= \mathbb{E}[N_1] + \sum_{k=2}^{n} \mathbb{E}\left[N_k\right] && \text{split up sum} \\
&= \mathbb{E}[N_1] + \sum_{k=1}^{n-1} \mathbb{E}\left[N_{k+1}\right] && \text{adjust summation bounds} \\
&= 1 + \sum_{k=1}^{n-1} \frac{1}{\left(\frac{n-k}{n}\right)} && \mathbb{E}\text{ of geo r.v.} \\
&= 1 + n\sum_{k=1}^{n-1} \frac{1}{n-k} && \text{factor, rearrange frac} \\
&= 1 + n\left(\frac{1}{n-1} + \frac{1}{n-2} + \ldots + \frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right)
\end{aligned}$$

The term inside the parentheses is a harmonic series bounded by $ln(n)+1$, so $\mathbb{E}[X_n]$ in fact grows logarithmically–more slowly as $n \to \infty$.

7. **Meester 2.7.21**. Let (X, Y) be a random vector with probability mass function $\mathbb{P}(X = i, Y = j) = 1/10$ for $1 \leq i \leq j \leq 4$.

(a) Show that this is a probability mass function. Let $\Omega = (x, y)$, pairs of outcomes of the random variables with probability mass > 0. $|\Omega| = 10 : \Omega = \{\omega | \omega \in \{(1,1),(1,2),(1,3),(1,4),(2,2),(2,3),(2,4),(3,3),(3,4),(4,4)\}\}$. Per the problem setup, each outcome has mass $1/10$, so the entire probability space has mass $(1/10) * 10$ as required.

(b) Compute the marginal distributions of X and Y.
Per the partition rule,

$$\mathbb{P}(X = 1) = \sum_{y \in Y} \mathbb{P}(X = 1, Y = y) = (1/10) + (1/10) + (1/10) + (1/10) = 4/10$$
$$\mathbb{P}(X = 2) = \sum_{y \in Y} \mathbb{P}(X = 1, Y = y) = 0(1/10) + (1/10) + (1/10) + (1/10) = 3/10$$
$$\mathbb{P}(X = 3) = \sum_{y \in Y} \mathbb{P}(X = 1, Y = y) = 0(1/10) + 0(1/10) + (1/10) + (1/10) = 2/10$$
$$\mathbb{P}(X = 4) = \sum_{y \in Y} \mathbb{P}(X = 1, Y = y) = 0(1/10) + 0(1/10) + 0(1/10) + (1/10) = 1/10$$

As expected,

$$\sum_{i=1}^{4} \mathbb{P}(X = i) = 1$$

$$\mathbb{P}(Y = 1) = \sum_{x \in x} \mathbb{P}(X = x, Y = 1) = (1/10) + 0(1/10) + 0(1/10) + 0(1/10) = 1/10$$
$$\mathbb{P}(Y = 2) = \sum_{x \in x} \mathbb{P}(X = x, Y = 2) = (1/10) + (1/10) + 0(1/10) + 0(1/10) = 2/10$$
$$\mathbb{P}(Y = 3) = \sum_{x \in x} \mathbb{P}(X = x, Y = 3) = (1/10) + (1/10) + (1/10) + 0(1/10) = 3/10$$
$$\mathbb{P}(Y = 4) = \sum_{x \in x} \mathbb{P}(X = x, Y = 4) = (1/10) + (1/10) + (1/10) + (1/10) = 4/10$$

As expected,

$$\sum_{i=1}^{4} \mathbb{P}(Y = i) = 1$$

(c) Are X and Y independent?
No. Counterexample:

$$\begin{aligned} \mathbb{P}(X = 4, Y = 1) &= 0 \\ \mathbb{P}(X = 4)\mathbb{P}(Y = 1) &= (1/16)(25/192) \neq 0 \end{aligned}$$

(d) Compute $\mathbb{E}[XY]$

$$\begin{aligned} \mathbb{E}[XY] &= \sum_{\substack{x,y \in (R(X), R(Y)) \\ x \leq y}} XY\mathbb{P}(X = x, Y = y) \\ &= \frac{1}{10} \sum_{x,y} XY \\ &= \frac{1}{10}[(1*1) + (1*2) + (1*3) + (1*4) + (2*2) + (2*3) + (2*4) + (3*3) + (3*4) + (4*4)] \\ &= 6.5 \end{aligned}$$

8. **Meester 2.7.24**. We roll two fair dice. Find the joint probability mass function of X and Y when

(a) X is the largest value obtained and Y is the sum of the values

Consider the event $Y = 3$. All combinations of two dice occur with equal probability, and two such outcomes lead to $Y = 3$: $(1, 2)$ and $(2, 1)$. The probability of seeing either of these outcomes is $\mathbb{P}(Y) = 2/36$. If $Y = 2$, we know intuitively that the only possible value of X is 1. Thus $\mathbb{P}(X = 1|Y = 2) = 1$. To find $\mathbb{P}(Y = 2, X = 1)$, we apply the definition of conditional probability and see $\mathbb{P}(Y = 2, X = 1) = \mathbb{P}(Y = 2|X = 1)\mathbb{P}(Y = 1) = 1/36$. We can replicate this approach to find the joint probability for all values with non-zero probability mass in the ranges of X and Y.

The following table enumerates $\mathbb{P}(X = x, Y = y)$ for $x \in R(X), y \in R(Y)$. Blank cells represent events with probability 0.

	X=1	X=2	X=3	X=4	X=5	X=6
Y=2	1/36					
Y=3		2/36				
Y=4		1/36	2/36			
Y=5			2/36	2/36		
Y=6			1/36	2/36	2/36	
Y=7				2/36	2/36	2/36
Y=8				1/36	2/36	2/36
Y=9					2/36	2/36
Y=10					1/36	2/36
Y=11						2/36
Y=12						1/36

(b) X is the value on the first die and Y is the largest value
$\forall x \in R(X), \mathbb{P}(X = x) = 1/6$. Thus, $\mathbb{P}(Y = y, X = x) = \mathbb{P}(Y = y|X = x)\mathbb{P}(X = x) = (1/6)\mathbb{P}(Y = y|X = x)$.

The following table enumerates $\mathbb{P}(X = x, Y = y)$ for $x \in R(X), y \in R(Y)$. Blank cells represent events with probability 0.

	X=1	X=2	X=3	X=4	X=5	X=6
Y=1	1/36					
Y=2	1/36	1/18				
Y=3	1/36	1/36	1/12			
Y=4	1/36	1/36	1/36	1/9		
Y=5	1/36	1/36	1/36	1/36	5/36	
Y=6	1/36	1/36	1/36	1/36	1/36	1/6

(c) X is the smallest value and Y is the largest
Note that $\mathbb{P}(Y = y|X = x) = 0$ if $y < x$ and that $\mathbb{P}(Y = y|X = x) = 2/36\ \forall y > x$ since there are two ways to roll the values $\{x, y\}$.

The following table enumerates $\mathbb{P}(X = x, Y = y)$ for $x \in R(X), y \in R(Y)$. Blank cells represent events with probability 0.

	X=1	X=2	X=3	X=4	X=5	X=6
Y=1	1/36					
Y=2	2/36	1/36				
Y=3	2/36	2/36	1/36			
Y=4	2/36	2/36	2/36	1/36		
Y=5	2/36	2/36	2/36	2/36	1/36	
Y=6	2/36	2/36	2/36	2/36	2/36	1/36

9. **Meester 2.7.25** Note that

$$\mathbb{E}[Y|X = x] = \sum_{y \in R(Y)} y\mathbb{P}(Y = y|X = x) = \sum_{y \in R(Y)} y\frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)}$$

So to calculate $\mathbb{E}[Y|X = x]$, we can calculate the weighted average of the conditional Y values, then normalize that weighted average by dividing by the probability that $X = x$.

(a) X is the largest value obtained and Y is the sum of the values

$$
\begin{aligned}
\mathbb{E}[Y|X=1] &= \frac{2(1/36)}{1/36} = 2 \\
\mathbb{E}[Y|X=2] &= \frac{3(2/36)+4(1/36)}{3/36} = 10/3 \\
\mathbb{E}[Y|X=3] &= \frac{4(2/36)+5(2/36)+6(1/36)}{3/36} = 24/5 \\
\mathbb{E}[Y|X=4] &= \frac{5(2/36)+6(2/36)+7(2/36)+8(1/36)}{7/36} = 44/7 \\
\mathbb{E}[Y|X=5] &= \frac{6(2/36)+7(2/36)+7(2/36)+8(2/36)+9(2/36)+10(1/36)}{9/36} = 84/9 \\
\mathbb{E}[Y|X=6] &= \frac{7(2/36)+8(2/36)+9(2/36)+10(2/36)+11(2/36)+12(1/36)}{11/36} = 102/11
\end{aligned}
$$

(b) X is the value on the first die and Y is the largest value

$$
\begin{aligned}
\mathbb{E}[Y|X=1] &= \frac{(1/36)[1+2+3+4+5+6]}{6/36} = 21/6 \\
\mathbb{E}[Y|X=2] &= \frac{2(2/36)+3(1/36)+4(1/36)+5(1/36)+6(1/36)}{6/36} = 22/6 \\
\mathbb{E}[Y|X=3] &= \frac{3(3/36)+4(1/36)+5(1/36)+6(1/36)}{6/36} = 4 \\
\mathbb{E}[Y|X=4] &= \frac{4(4/36)+5(1/36)+6(1/36)}{6/36} = 27/6 \\
\mathbb{E}[Y|X=5] &= \frac{5(5/36)+6(1/36)}{6/36} = 31/6 \\
\mathbb{E}[Y|X=6] &= \frac{6(6/36)}{6/36} = 1
\end{aligned}
$$

(c) X is the smallest value and Y is the largest

$$
\begin{aligned}
\mathbb{E}[Y|X=1] &= \frac{1(1/36)+2(2/36)+3(2/36)+4(2/36)+5(2/36)+6(2/36)}{11/36} = 41/11 \\
\mathbb{E}[Y|X=2] &= \frac{2(1/36)+3(2/36)+4(2/36)+5(2/36)+6(2/36)}{9/36} = 38/9 \\
\mathbb{E}[Y|X=3] &= \frac{3(1/36)+4(2/36)+5(2/36)+6(2/36)}{7/36} = 33/7 \\
\mathbb{E}[Y|X=4] &= \frac{1(4/36)+5(2/36)+6(2/36)}{5/36} = 26/5 \\
\mathbb{E}[Y|X=5] &= \frac{5(1/36)+6(2/36)}{3/36} = 17/3 \\
\mathbb{E}[Y|X=6] &= \frac{6(1/36)}{1/36} = 6
\end{aligned}
$$

10. **Meester 2.7.32** Let X and Y be independent and geometrically distributed with the same parameter p. Compute the probability mass function of $X - Y$. Can you also compute $P(X = Y)$ now?

Compute probability mass function

$$
\mathbb{P}(X - Y = k) = \sum_{x \in R(X)} \mathbb{P}(X - Y = k | X = x)\mathbb{P}(X = x)
$$

So rearranging the random variables,

$$\begin{aligned}
\mathbb{P}(X-Y=k) &= \sum_{x\in R(X)} \mathbb{P}(Y=X-k|X=x)\mathbb{P}(X=x) \\
&= \sum_{x\in R(X)} (1-p)^{x-k-1}p(1-p)^{x-1}p && \text{marginal of geo r.v.} \\
&= \sum_{x\in R(X)} (1-p)^{2x-k-2}p^2 && \text{combine terms} \\
&= p^2(1-p)^{-k-2}\sum_{x\in R(X)} (1-p)^{2x} && \text{factor} \\
&= p^2(1-p)^{-k-2}\sum_{x=1}^{\infty} \left[(1-p)^2\right]^x && \text{separate exponents} \\
&= \frac{p^2(1-p)^{-k-2}}{1-(1-p)^2} && \text{evaluate geo series}
\end{aligned}$$

Compute $\mathbb{P}(X=Y)$

$X=Y$ occurs when when $X-Y=0$. Per the probability mass function occurs, the probability of this event is

$$\mathbb{P}(X-Y=0) = \frac{p^2(1-p)^{-2}}{1-(1-p)^2}$$

11. A bag has 14 marbles: 10 are red, 4 are blue. Consider the following two-stage experiment: I roll a standard 6-sided die one time. Let Y be the value rolled. Then I draw Y marbles randomly from the bag without replacement.

   (i) What is the probability that all 4 blue marbles are drawn?
   Per the partition rule,

$$\mathbb{P}(X=4) = \sum_{y\in R(Y)} \mathbb{P}(X=4|Y=y)\mathbb{P}(Y=y) \tag{5}$$

   If we draw 4 blue marbles, Y must have been ≥ 4, so (5) can be refined to:

$$\mathbb{P}(X=4) = \sum_{y=4}^{6} \mathbb{P}(X=4|Y=y)\mathbb{P}(Y=y) \tag{6}$$

   We can count the conditional probability that $X=4|Y=k$:

$$\mathbb{P}(X=4|Y=4) = \frac{\binom{4}{4}}{\binom{14}{4}} \tag{1}$$

$$\mathbb{P}(X=4|Y=5) = \frac{\binom{4}{4}\binom{10}{1}}{\binom{14}{5}} \tag{2}$$

$$\mathbb{P}(X=4|Y=6) = \frac{\binom{4}{4}\binom{10}{2}}{\binom{14}{6}} \tag{3}$$

   $\mathbb{P}(Y=y)$ for any $y\in\{4,5,6\}$ is $1/6$. Plugging this information and (1), (2), (3) into (6), we find that

$$\mathbb{P}(X=4) = \frac{1}{6}\left[\frac{\binom{4}{4}}{\binom{14}{4}} + \frac{\binom{4}{4}\binom{10}{1}}{\binom{14}{5}} + \frac{\binom{4}{4}\binom{10}{2}}{\binom{14}{6}}\right] = \frac{1}{286}$$

   (ii) What is the expected number of red marbles drawn? Let R be a random variable denoting how many red marbles are drawn. Applying the standard formula for expected value,

$$\mathbb{E}[R] = \sum_{r\in\text{Range}(R)} r\mathbb{P}(R=r) = \sum_r r \sum_{y\in\text{Range}(y)} \mathbb{P}(R=r|Y=y)\mathbb{P}(Y=y) \tag{4}$$

   R is naturally limited by Y and $1\leq Y\leq 6$, so

$$\mathbb{E}[R] = \sum_{y=1}^{6}\sum_{r=1}^{y} r\mathbb{P}(R=r|Y=y)\mathbb{P}(Y=y) = \frac{1}{6}\sum_{y=1}^{6}\sum_{r=1}^{y} r\mathbb{P}(R=r|Y=y)$$

For each combination of r, y, we can determine $\mathbb{P}(R = r|Y = y)$ through counting principles. E.g., if $r = 3, y = 4$, we are trying to find the probability that we draw three red marbles given we draw four marbles in total. There are $\binom{4}{3}$ ways to choose 3 red marbles and $\binom{10}{1}$ ways to choose the last marble, so there are, in total, $\binom{4}{3}\binom{10}{1}$ ways to choose three red marbles in a draw of four. In comparison, there are $\binom{14}{4}$ ways to draw any four marbles, so $\mathbb{P}(R = 3|Y = 4) = \frac{\binom{4}{3}\binom{10}{1}}{\binom{14}{4}}$. Applying this to all, cases,

$$
\begin{aligned}
\mathbb{E}[R] = \frac{1}{6}\Bigg[& 1\frac{\binom{10}{1}}{\binom{14}{1}} \\
& +1\frac{\binom{10}{1}\binom{4}{1}}{\binom{14}{2}} + 2\frac{\binom{10}{2}}{\binom{14}{3}} \\
& +1\frac{\binom{10}{1}\binom{4}{2}}{\binom{14}{3}} + 2\frac{\binom{10}{2}\binom{4}{1}}{\binom{14}{3}} + 3\frac{\binom{10}{3}}{\binom{14}{3}} \\
& +1\frac{\binom{10}{1}\binom{4}{3}}{\binom{14}{4}} + 2\frac{\binom{10}{2}\binom{4}{2}}{\binom{14}{2}} + 3\frac{\binom{10}{3}\binom{4}{1}}{\binom{14}{4}} + 4\frac{\binom{10}{4}}{\binom{14}{4}} \\
& +1\frac{\binom{10}{1}\binom{4}{4}}{\binom{14}{5}} + 2\frac{\binom{10}{2}\binom{4}{3}}{\binom{14}{5}} + 3\frac{\binom{10}{3}\binom{4}{2}}{\binom{14}{5}} + 4\frac{\binom{10}{4}\binom{4}{1}}{\binom{14}{5}} + 5\frac{\binom{10}{5}}{\binom{14}{5}} \\
& +1\cancel{\frac{\binom{10}{1}\binom{4}{5}}{\binom{14}{6}}} + 2\frac{\binom{10}{2}\binom{4}{4}}{\binom{14}{6}} + 3\frac{\binom{10}{3}\binom{4}{3}}{\binom{14}{6}} + 4\frac{\binom{10}{4}\binom{4}{2}}{\binom{14}{6}} + 5\frac{\binom{10}{5}\binom{4}{1}}{\binom{14}{6}} + 6\frac{\binom{10}{6}}{\binom{14}{6}}\Bigg]
\end{aligned}
$$

12. Roll two fair 6-sided dice, one aftere the other. Let X be the number on the first roll. Let Y be the number on the second roll. Let $Z = X - Y$.

 (i) What is $\mathbb{P}(X > Y)$?

$$\mathbb{P}(X > Y) = \sum_{y=1}^{5}\sum_{\substack{x\le 6\\ x>y\\ x\in\mathbb{Z}}} \mathbb{P}(X = x, Y = y) = (5+4+3+2+1)\mathbb{P}(X = x, Y = y) = 15\mathbb{P}(X = x, Y = y$$

 ? Since all events $X = k$ and $Y = j$ for $k, j \in \{1, 2, \ldots, 6\}$ are equally likely and there are 36 such events, $\mathbb{P}(X = k, Y = j) = 1/36$ and $\boxed{\mathbb{P}(X > Y) = 15/36}$.

 (ii) What is the joint distribution of X and Z?
 This question is essentnially asking how likely it is that we see $X = k$ and some $Y = y$ that is exactly $Z = j$ smaller than k. All events $(X = k) \cap (Y = y)$ are equally likely so long as k and y are valid dice values. Thus, so long as $k - j > 0$ (i.e. $Y > 0$) and $k - j \le 6$ (i.e. $Y \le 6$), $\mathbb{P}(X = k, Z = j) = 1/36$. Succinctly,

$$\mathbb{P}(X = k, Z = j) = \begin{cases} 1/36, & 0 < k - j \\ 0, & \text{otherwise} \end{cases}$$

 (iii) What is the distribution of Z?
 Per the partition rule,

$$\mathbb{P}(Z = k) = \sum_{j=1}^{6} \mathbb{P}(Z = k, X = j)$$

 We can find $\mathbb{P}(Z = 1)$ by thinking about $\mathbb{P}(Z = k, X = j)$, for each $j \in R(X)$. When $j = 1$, to fulfill k=1, $1 = 1 - Y$, so $Y = 0$ which is impossible. Thus $\mathbb{P}(Z = 1, X = 1) = 0$. If $j = 2$, to fulfill $k = 2$, $1 = 2 - Y$, so $Y = 1$. $\mathbb{P}(X = 2, Y = 1) = \mathbb{P}(X = 2, Z = 1) = 1/36$. Continuing to evaluate joint probabilities in this way, we find

$$
\begin{aligned}
\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = 5/36 \\
\mathbb{P}(Z = 2) = \mathbb{P}(Z = -2) = 4/36 \\
\mathbb{P}(Z = 3) = \mathbb{P}(Z = -3) = 3/36 \\
\mathbb{P}(Z = 4) = \mathbb{P}(Z = -4) = 2/36 \\
\mathbb{P}(Z = 5) = \mathbb{P}(Z = -5) = 1/36
\end{aligned}
$$

(iv) What is the conditional distribution of X given $Z = 2$?
We first note that $\mathbb{P}(Z = 2) = 4/36$. Now since $\mathbb{P}(X = k, Z = j)$ occurs with equal probability for each valid k, j, and $k > j$ given the construction of Z,

$$\mathbb{P}(Z = 2, X = 3) = \mathbb{P}(Z = 2, X = 4) = \mathbb{P}(Z = 2, X = 5) = \mathbb{P}(Z = 2, X = 6) = 1/36$$
$$\mathbb{P}(Z = 2, X = 1) = \mathbb{P}(Z = 2, X = 2) = 0/36$$

Normalizing these probabilities using $\mathbb{P}(Z = 2)$, we find

$$\mathbb{P}(X|Z = 2) = \begin{cases} 1/4, & x \in \{3, 4, 5, 6\} \\ 0, & \text{otherwise} \end{cases}$$

(v) Are X and Z independent?
No. For a counterexample, consider $X = 1, Z = 2$. $\mathbb{P}(X = 1, Z = 2) = 0$ since $Z < X$ by construction. But $\mathbb{P}(X = 1)\mathbb{P}(Z = 2) = (1/6)(4/36) \neq 0$.