Homework 4

- Homework 4 is due January 31 in class.
- Exercises are taken from R.V. Hogg, E. Tanis, D. Zimmerman, *Probability and Statistical Inference*, 10th Ed., Pearson.
- The starred problems will not be graded.

2.3 Special Mathematical Expectations

Exercises

1. Find the mean, variance, and index of skewness for the following discrete distributions:

(a) $p(x) = 1, \qquad x = 5.$

(b) $p(x) = 1/5, \qquad x = 1, 2, 3, 4, 5.$

(c) $p(x) = 1/5, \qquad x = 3, 5, 7, 9, 11.$

(d) $p(x) = x/6, \qquad x = 1, 2, 3.$

(e*) $p(x) = (1 + |x|)/5, \qquad x = -1, 0, 1.$

(f) $p(x) = (2 - |x|)/4, \qquad x = -1, 0, 1.$

4.* Let μ and σ^2 denote the mean and variance of the random variable X. Determine

$$E\left[\frac{X-\mu}{\sigma}\right] \qquad \text{and} \qquad E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right].$$

5. Consider an experiment that consists of selecting a card at random from an ordinary deck of cards. Let the random variable X equal the value of the selected card, where Ace $=$ 1, Jack $=$ 11, Queen $=$ 12, and King $=$ 13. Thus, the space of X is $S = \{1, 2, 3, \cdots, 13\}$. If the experiment is performed in an unbiased manner, assign probabilities to these 13 outcomes and compute the mean μ of this probability distribution.

8. Let X equal the larger outcome when a pair of fair four-sided dice is rolled. The pmf of X is

$$p(x) = \frac{2x-1}{16}, \qquad x = 1, 2, 3, 4.$$

Find the mean, variance, and standard deviation of X.

10. Let X be a discrete random variable with pmf

$$p(x) = \begin{cases} 1/16, & x = -5, \\ 5/8, & x = -1, \\ 5/16, & x = 3. \end{cases}$$

Find the index of skewness of X. Is this distribution symmetric?

11. If the moment-generating function of X is

$$M(t) = \frac{2}{5}e^t + \frac{1}{5}e^{2t} + \frac{2}{5}e^{3t},$$

find the mean, variance, and pmf of X.

16.* Let X equal the number of flips of a fair coin that are required to observe the same face on consecutive flips.

(a) Find the pmf of X. Hint: Draw a tree diagram.

(b) Find the moment-generating function of X.

(c) Use the mgf to find the values of **(i)** the mean and **(ii)** the variance of X.

(d) Find the values of **(i)** $P(X \leq 3)$, **(ii)** $P(X \geq 5)$, and **(iii)** $P(X = 3)$.

17. Let X equal the number of flips of a fair coin that are required to observe heads–tails on consecutive flips.

(a) Find the pmf of X. Hint: Draw a tree diagram.

(b) Show that the mgf of X is $M(t) = e^{2t}/(e^t - 2)^2$.

(c) Use the mgf to find the values of **(i)** the mean and **(ii)** the variance of X.

(d) Find the values of **(i)** $P(X \leq 3)$, **(ii)** $P(X \geq 5)$, and **(iii)** $P(X = 3)$.

18. Let X have a geometric distribution. Show that

$$P(X > k + j \mid X > k) = P(X > j),$$

where k and j are nonnegative integers. *Note:* We sometimes say that in this situation there has been loss of memory.

Hint: Show first that $P(X > j) = q^j$ for each non-negative integer j.

19.* Given a random permutation of the integers in the set $\{1, 2, 3, 4, 5\}$, let X equal the number of integers that are in their natural position. The moment-generating function of X is

$$M(t) = \frac{44}{120} + \frac{45}{120}e^t + \frac{20}{120}e^{2t} + \frac{10}{120}e^{3t} + \frac{1}{120}e^{5t}.$$

(a) Find the mean and variance of X.

(b) Find the probability that at least one integer is in its natural position.

(c) Draw a graph of the probability histogram of the pmf of X.

2.4. The Binomial Distribution

3. On a six-question multiple-choice test there are five possible answers for each question, of which one is correct (C) and four are incorrect (I). If a student guesses randomly and independently, find the probability of

(a) Being correct only on questions 1 and 4 (i.e., scoring C, I, I, C, I, I).

(b) Being correct on two questions.

6. It is believed that approximately 75% of American youth now have insurance due to the health care law. Suppose this is true, and let X equal the number of American youth in a random sample of $n = 15$ with private health insurance.

(a) How is X distributed?

(b) Find the probability that X is at least 10.

(c) Find the probability that X is at most 10.

(d) Find the probability that X is equal to 10.

(e) Give the mean, variance, and standard deviation of X.

Note: In each of (b) and (c), you may leave your answer as a sum.

11. Find the index of skewness for the $b(n, p)$ distribution, and verify that it is negative if $p < 0.5$, zero if $p = 0.5$, and positive if $p > 0.5$.

15. A hospital obtains 40% of its flu vaccine from Company A, 50% from Company B, and 10% from Company C. From past experience, it is known that 3% of the vials from A are ineffective, 2% from B are ineffective, and 5% from C are ineffective. The hospital tests five vials from each shipment. If at least one of the five is inef- fective, find the conditional probability of that shipment's having come from C.

18. In group testing for a certain disease, a blood sample was taken from each of n individuals and part of each sample was placed in a common pool. The latter was then tested. If the result was negative, there was no more testing and all n individuals were declared negative with one test. If, however, the combined result was found positive, all individuals were tested, requiring n+1 tests. If $p = 0.05$ is the probability of a person's having the disease and $n = 5$, compute the expected number of tests needed, assuming independence.

2.5. The Hypergeometric Distribution

1. In a lot (collection) of 100 light bulbs, there are five bad bulbs. An inspector inspects ten bulbs selected at random. Find the probability of finding at least one defective bulb. *Hint:* First compute the probability of finding no defectives in the sample.

6.* To find the variance of a hypergeometric random variable in Equation 2.5-2, we used the fact that

$$E[X(X-1)] = \frac{N_1(N_1-1)(n)(n-1)}{N(N-1)}.$$

Prove this result by making the change of variables $k = x - 2$ to

$$E[X(X-1)] = \sum_{\substack{x \in S \\ x \neq 0,1}} x(x-1)\frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}$$

and noting that

$$\binom{N}{n} = \frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}.$$

8. Forty-four states, Washington D.C., and the Virgin Islands have joined for the Mega Millions lottery game. For this game the player selects five white balls numbered from 1 to 70, inclusive, plus a single gold Mega Ball numbered from 1 to 25, inclusive. There are several different prize options including the following.

(a) What is the probability of matching all five white balls plus the Mega Ball and winning the jackpot?

(b) What is the probability of matching five white balls but not the Mega Ball and winning \$1,000,000?

- **(c)** What is the probability of matching four white balls plus the MEga Ball and winning \$10,000?
- **(d)** What is the probability of matching four white balls but not the Mega Ball and winning \$400?
- **(e)** What is the probabiity of matching the Mega Ball only and winning \$2?

9. Suppose there are three defective items in a log (collection) of 50 items. A sample of size ten is taken at random and without replacement. Let X denote the number of defective items in the sample. Find the probability that the sample contains

- **(a)** Exactly one defective item.
- **(b)** At most one defective item.

Solution of some selected problems

Solution of Exercise 2.3-1.(e) Plugging the values $x = -1, 0, 1$, we get

x	-1	0	1
$p(x)$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

Now let X be any random variable with the above PMF. Then the mean is computed as

$$\begin{aligned} \mu = E(X) &= (-1) \cdot p(-1) + 0 \cdot p(0) + 1 \cdot p(1) \\ &= (-1) \cdot \frac{2}{5} + 0 \cdot \frac{1}{5} + 1 \cdot \frac{2}{5} = \boxed{0}. \end{aligned}$$

Then using this, the variance can be computed as

$$\begin{aligned} \sigma^2 = \mathrm{Var}(X) = E[(X-\mu)^2] &\overset{(\mu=0)}{=} E(X^2) \\ &= (-1)^2 \cdot p(-1) + 0^2 \cdot p(0) + 1^2 \cdot p(1) = \boxed{\frac{4}{5}}. \end{aligned}$$

To compute the index of skewness $\gamma = E[(X-\mu)^3]/\sigma^3$, we first compute

$$E[(X-\mu)^3] \overset{(\mu=0)}{=} E(X^3) = (-1)^3 \cdot p(-1) + 0^3 \cdot p(0) + 1^3 \cdot p(1) = 0.$$

Dividing both sides by $\sigma^3 = 1$, we get

$$\gamma = \boxed{0}.$$

Note that this is also implied by the fact that the distribution is symmetric. □

Solution of Exercise 2.3-4. For the first one, the linearity of expectation gives

$$E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}(\underbrace{E(X)}_{=\mu} - \mu) = 0.$$

For the second one, the linearity of expectation together with the definition of the variance $\mathrm{Var}(X) = \sigma^2$ tells that

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma^2}\underbrace{E[(X-\mu)^2]}_{=\sigma^2} = 1.$$

This tells that the new random variable $Y = (X-\mu)/\sigma$ has zero mean and unit variance. This procedure of transforming a variable to have a mean of zero and a standard deviation of 1 is often called *standardization*. □

Solution of Exercise 2.3-16.

(a) We stop flipping the coin as soon as we have observed the consecutive patterns HH or TT. So the tree diagram would look like as in Figure 1. In particular, we can only observe patterns of the form HTHT$\cdots$HTT or THTH$\cdots$THH. So we deduce that the support of X is

$$S = \{2, 3, 4, \cdots\},$$

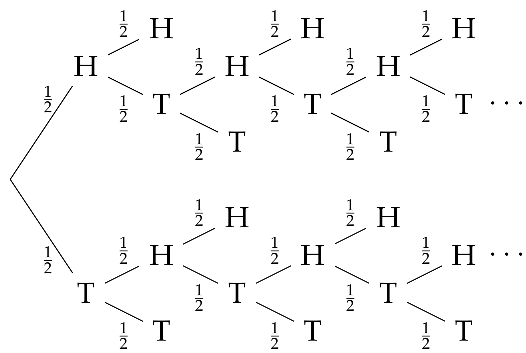


Figure 1: Tree diagram for Exercise 2.3-16.

reflecting the fact that you have to flip the coin at least twice to complete either pattern. Moreover,

$$\begin{aligned}
p(2) &= P(\text{pattern HH or TT is observed}) = \frac{1}{2^2} + \frac{1}{2^2} = \frac{1}{2}, \\
p(3) &= P(\text{pattern HTT or THH is observed}) = \frac{1}{2^3} + \frac{1}{2^3} = \frac{1}{2^2}, \\
p(4) &= P(\text{pattern HTHH or THTT is observed}) = \frac{1}{2^4} + \frac{1}{2^4} = \frac{1}{2^3}, \\
&\vdots
\end{aligned}$$

In general, for each $x = 2, 3, 4, \cdots$ there are exactly two possible patterns of length x, each of which appearing with probability $1/2^x$. So the PMF is given by

$$p(x) = \frac{1}{2^{x-1}}, \qquad x = 2, 3, 4, \cdots .$$

(b) The MGF of X is computed as:

$$\begin{aligned}
M(t) = E(e^{tX}) &= \sum_{x \in S} e^{tx} p(x) = \sum_{x=2}^{\infty} e^{tx} \cdot \frac{1}{2^{x-1}} \\
&= \frac{e^{2t}}{2} \left[1 + \frac{e^t}{2} + \left(\frac{e^t}{2} \right)^2 + \cdots \right] \\
&= \frac{e^{2t}}{2} \cdot \frac{1}{1 - \frac{e^t}{2}} = \frac{e^{2t}}{2 - e^t},
\end{aligned}$$

provided $e^t/2 < 1$, or equivalently, $t < \ln 2$. In the third line, we utilized the geometric series $\frac{1}{1-z} = 1 + z + z^2 + \cdots$ which is valid if and only if $|z| < 1$.

(c) **(i)** Using the MGF, the mean μ of X is computed as

$$\mu = E(X) = M'(0) = \frac{\mathrm{d}}{\mathrm{d}t} \frac{e^{2t}}{2 - e^t} \bigg|_{t=0} = \frac{e^{2t}(4 - e^t)}{(2 - e^t)^2} \bigg|_{t=0} = 3.$$

(ii) In order to compute, we first evaluate the second moment $E(X^2)$:

$$E(X^2) = M''(0) = \frac{\mathrm{d}^2}{\mathrm{d}t^2} \frac{e^{2t}}{2-e^t}\bigg|_{t=0} = \frac{e^{2t}(16-6e^t+e^{2t})}{(2-e^t)^3}\bigg|_{t=0} = 11.$$

Then by the formula $\sigma^2 = E(X^2) - \mu^2$ together with the computation in part (i), we get

$$\sigma^2 = E(X^2) - \mu^2 = 2.$$

(d) **(i)** $P(X \le 3) = p(2) + p(3) = \frac{3}{4}$.

(ii) $P(X \ge 5) = 1 - P(X < 5) = 1 - (p(2) + p(3) + p(4)) = \frac{1}{8}$.

(iii) $P(X = 3) = p(3) = \frac{1}{4}$.

□

Solution of Exercise 2.3-19.

(a) Since

$$M'(t) = \frac{45}{120}e^t + \frac{40}{120}e^{2t} + \frac{30}{120}e^{3t} + \frac{5}{120}e^{5t},$$
$$M''(t) = \frac{45}{120}e^t + \frac{80}{120}e^{2t} + \frac{90}{120}e^{3t} + \frac{25}{120}e^{5t},$$

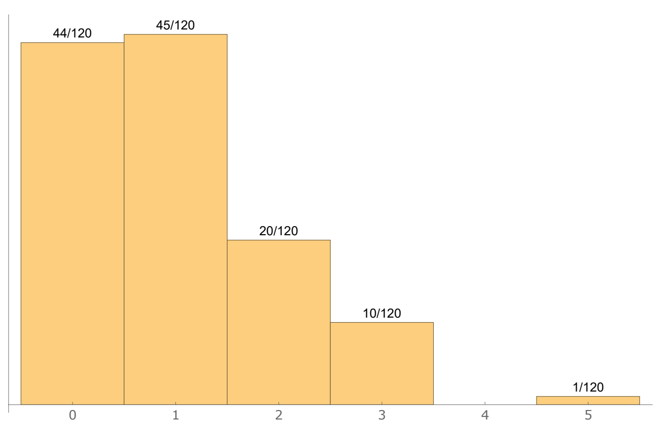
we get $E(X) = M'(0) = 1$ and $E(X^2) = M''(0) = 2$. From this, we compute the mean and variance as $\mu = E(X) = M'(0) = 1$ and $\sigma^2 = 1$.

(b) The desired probability is $P(X \ge 1) = 1 - P(X = 0) = 1 - p(0)$, where $p(x)$ denotes the PMF of X. From the expansion of the MGF, we read out that $p(0)$ is the value of the constant term in the expansion, which is $\frac{44}{120}$. Therefore $P(X \ge 1) = 1 - \frac{44}{120} = \frac{19}{30}$.

(c) The values of $p(x)$ can be read out from the coefficient of e^{tx} in the expansion of the MGF. So

x	0	1	2	3	4	5
$p(x)$	$\frac{44}{120}$	$\frac{45}{120}$	$\frac{20}{120}$	$\frac{10}{120}$	0	$\frac{1}{120}$

(The column for $x = 4$ is included to emphasize that $p(4) = P(X = 4) = 0$.) Accordingly, a probability histogram looks like:



□

Solution of Exercise 2.5-6. When $n = 0$ or $n = 1$, it is clear that X takes values only in $\{1, 0\}$, hence $X(X-1) = 0$ identically. Therefore both sides of the desired identity reduces to 0. For $x \geq 2$. In light of this, we may assume that $n \geq 2$. Then $x \geq 2$, we may write

$$\binom{N_1}{x} = \frac{N_1!}{(N_1 - x)!x!} = \frac{N_1(N_1-1)}{x(x-1)}\binom{N_1-2}{x-2}$$

Plugging this and $\binom{N}{n} = \frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}$ to the sum for $E[X(X-1)]$, we get

$$E[X(X-1)] = \frac{N_1(N_1-1)n(n-1)}{N(N-1)} \sum_{\substack{x \in S \\ x \neq 0,1}} \frac{\binom{N_1-2}{x-2}\binom{N_2}{n-x}}{\binom{N-2}{n-2}}$$

So it suffices to show that the summation part equals 1. Indeed, substituting $k = x - 2$,

$$\sum_{\substack{x \in S \\ x \neq 0,1}} \frac{\binom{N_1-2}{x-2}\binom{N_2}{n-x}}{\binom{N-2}{n-2}} = \sum_{k \geq 0} \frac{\binom{N_1-2}{k}\binom{N_2}{n-2-k}}{\binom{N-2}{n-2}}.$$

However, we recognize the right-hand side as the total sum of the PMF of $HG(N_1 - 2, N_2, n - 2)$ distribution, thus the sum equals 1 as required. $\square$