

7. Lotteries

Lotteries are among the simplest and most widely played of all games of chance, and unfortunately for the gambler, among the worst in terms of expected value. Lotteries come in such an incredible number of variations that it is impractical to analyze all of them. Thus, in this section, we will study some of the more common lottery formats.

The Basic Lottery

The basic lottery is a [random experiment](#) in which the gambling house (in many cases a government agency) selects n numbers at random, without replacement, from the integers from 1 to N . The integer parameters N and n vary from one lottery to another, and of course, n cannot be larger than N . The order in which the numbers are chosen usually does not matter, and thus in this case, the [sample space](#) S of the experiment consists of all subsets ([combinations](#)) of size n chosen from the population $\{1, 2, \dots, N\}$.

$$S = \{x \subseteq \{1, 2, \dots, N\} : \#(x) = n\}$$

1. Recall, or show that

$$\#(S) = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Naturally, we assume that all such combinations are equally likely, and thus, the chosen combination X , the basic [random variable](#) of the experiment, is [uniformly distributed](#) on S .

$$\mathbb{P}(X = x) = \frac{1}{\binom{N}{n}}, \quad x \in S$$

The player of the lottery pays a fee and gets to select m numbers, without replacement, from the integers from 1 to N . Again, order does not matter, so the player essentially chooses a combination y of size m from the population $\{1, 2, \dots, N\}$. In many cases $m = n$, so that the player gets to choose the same number of numbers as the house. In general then, there are three parameters in the [basic \$\(N, n, m\)\$ lottery](#).

The player's goal, of course, is to maximize the number of matches (often called [catches](#) by gamblers) between her combination y and the random combination X chosen by the house. Essentially, the player is trying to guess the outcome of the random experiment before it is run. Thus, let U denote the number of catches.

2. Show that the number of catches U in the (N, n, m) , lottery has [probability density function](#)

$$\mathbb{P}(U = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}, \quad k \in \{0, 1, \dots, m\}$$

The distribution of U is the [hypergeometric distribution](#) with parameters N , n , and m , and is studied in detail in the

chapter on [Finite Sampling Models](#). In particular, from this section, it follows that the [mean](#) and [variance](#) of the number of catches U is

$$\mathbb{E}(U) = n \frac{m}{N}, \quad \text{var}(U) = n \frac{m}{N} \left(1 - \frac{m}{N}\right) \frac{N-n}{N-1}$$

Note that $\mathbb{P}(U = k) = 0$ if $k > n$ or $k < (n + m) - N$. However, in most lotteries, $m \leq n$ and N is much larger than $n + m$. In these common cases, the density function is positive for the values of k given in [Exercise 2](#).

We will refer to the special case where $m = n$ as the **(N, n) , lottery**; this is the case in most state lotteries. In this case, the probability density function of the number of catches U is

$$\mathbb{P}(U = k) = \frac{\binom{n}{k} \binom{N-n}{n-k}}{\binom{N}{n}}, \quad k \in \{0, 1, \dots, n\}$$

The mean and variance of the number of catches U in this special case are

$$\mathbb{E}(U) = \frac{n^2}{N}, \quad \text{var}(U) = \frac{n^2 (N-n)^2}{N^2 (N-1)}$$

3. Explicitly give the probability density function, mean, and standard deviation of the number of catches in the (47, 5) lottery.



4. Explicitly give the probability density function, mean, and standard deviation of the number of catches in the (49, 5) lottery.



5. Explicitly give the probability density function, mean, and standard deviation of the number of catches in the (47, 7) lottery.



The analysis above was based on the assumption that the player's combination y is selected deterministically. Would it matter if the player chose the combination in a random way? Thus, suppose that the player's selected combination Y is a random variable taking values in S . (For example, in many lotteries, players can buy tickets with combinations randomly selected by a computer; this is typically known as **Quick Pick**). Clearly, X and Y must be independent, since the player (and her randomizing device) can have no knowledge of the winning combination X . In the following exercise, you will show that such randomization makes no difference.

6. Let U denote the number of catches in the (N, n, m) lottery when the player's combination Y is a random variable, independent of the winning combination X . Show that U has the same distribution as in Exercise 1. *Hint*: condition on the value of Y .

There are many web sites that publish data on the frequency of occurrence of numbers in various state lotteries. Some gamblers evidently feel that some numbers are luckier than others.

7. Given the assumptions and analysis above, do you believe that some numbers are luckier than others? Does it make any mathematical sense to study historical data for a lottery?

The prize money in most state lotteries depends on the sales of the lottery tickets. Typically, about 50% of the sales money is returned as prize money; the rest goes for administrative costs and profit for the state. The total prize money is divided among the winning tickets, and the prize for a given ticket depends on the number of catches U . For all of these reasons, it is impossible to give a simple mathematical analysis of the expected value of playing a given state lottery. Note however, that since the state keeps a fixed percentage of the sales, there is essentially no risk for the state.

From a pure gambling point of view, state lotteries are bad games. In most casino games, by comparison, 90% or more of the money that comes in is returned to the players as prize money. Of course, state lotteries should be viewed as a form of voluntary taxation, not simply as games. The profits from lotteries are typically used for education, health care, and other essential services. A discussion of the value and costs of lotteries from a *political and social* point of view (as opposed to a *mathematical* one) is beyond the scope of this project.

Bonus Numbers

Many state lotteries now augment the basic (N, n) , format with a **bonus number**. The bonus number T is selected from a specified set of integers, in addition to the combination X , selected as before. The player likewise picks a bonus number s , in addition to a combination y . The player's prize then depends on the number of catches U between X and y , as before, and in addition on whether the player's bonus number s matches the random bonus number T chosen by the house. We will let I denote the indicator variable of this latter event. Thus, our interest now is in the **joint distribution** of (I, U) .

In one common format, the bonus number T is selected at random from the set of integers $\{1, 2, \dots, M\}$, **independently** of the combination X of size n chosen from $\{1, 2, \dots, N\}$. Usually $M < N$. Note that with this format, the game is essentially two independent lotteries, one in the (N, n) , format and the other in the $(M, 1)$, format.

8. Explicitly compute the joint probability density function of (I, U) for the (47, 5) lottery with independent bonus number from 1 to 27. This format is used in the **California lottery**, among others.



9. Explicitly compute the joint probability density function of (I, U) for the (49, 5) lottery with independent bonus number from 1 to 42. This format is used in the **Powerball lottery**, among others.



In another format, the bonus number T is chosen from 1 to N , and is distinct from the numbers in the combination X . To model this game, we assume that T is uniformly distributed on $\{1, 2, \dots, N\}$, and given $T = t$, X is uniformly distributed on the set of combinations of size n chosen from $\{1, 2, \dots, N\} \setminus \{t\}$. For this format, the joint probability density function is harder to compute.

10. Show that

$$\mathbb{P}(I = 1, U = k) = \frac{\binom{n}{k} \binom{N-1-n}{n-k}}{N \binom{N-1}{n}}, \quad k \in \{0, 1, \dots, n\}$$

11. Condition on whether or not T is in $\{y_1, y_2, \dots, y_n\}$ to show that

$$\mathbb{P}(I = 0, U = k) = (N - n + 1) \frac{\binom{n}{k} \binom{N-1-n}{n-k}}{N \binom{N-1}{n}} + n \frac{\binom{n-1}{k} \binom{N-n}{n-k}}{N \binom{N-1}{n}}, \quad k \in \{0, 1, \dots, n\}$$

12. Explicitly compute the joint probability density function of (I, U) for the (47, 7) lottery with bonus number chosen as described above. This format is used in the **Super 7 Canada lottery**, among others.

Keno

Keno is a lottery game played in casinos. For a fixed N (usually 80) and n (usually 20), the player can play a range of basic (N, n, m) games, as described in the first subsection. Typically, m ranges from 1 to 15, and the payoff depends on m and the number of catches V . In this section, you will compute the density function, mean, and standard deviation of the random payoff, based on a unit bet, for a typical keno game with $N = 80$, $n = 20$, $m \in \{1, 2, \dots, 15\}$. The payoff tables are based on the keno game at the *Tropicana* casino in Atlantic City, New Jersey.

Recall that the probability density function of the number of catches U , from [Exercise 2](#) above, is given by

$$\mathbb{P}(U = k) = \frac{\binom{m}{k} \binom{80-m}{20-k}}{\binom{80}{20}}, \quad k \in \{0, 1, \dots, m\}$$

13. The payoff table for $m = 1$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 1$

Catches	0	1
Payoff	0	3

14. The payoff table for $m = 2$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 2$

Catches	0	1	2
Payoff	0	0	12

15. The payoff table for $m = 3$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 3$

Catches	0	1	2	3
---------	---	---	---	---

Payoff	0	0	1	43
--------	---	---	---	----



16. The payoff table for $m = 4$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 4$

Catches	0	1	2	3	4
Payoff	0	0	1	3	130



17. The payoff table for $m = 5$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 5$

Catches	0	1	2	3	4	5
Payoff	0	0	0	1	10	800



18. The payoff table for $m = 6$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 6$

Catches	0	1	2	3	4	5	6
Payoff	0	0	0	1	4	95	1500



19. The payoff table for $m = 7$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 7$

Catches	0	1	2	3	4	5	6	7
Payoff	0	0	0	0	1	25	350	8000



20. The payoff table for $m = 8$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 8$

Catches	0	1	2	3	4	5	6	7	8
Payoff	0	0	0	0	0	9	90	1500	25,000



21. The payoff table for $m = 9$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 9$

Catches	0	1	2	3	4	5	6	7	8	9
Payoff	0	0	0	0	0	4	50	280	4000	50,000



22. The payoff table for $m = 10$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 10$

Catches	0	1	2	3	4	5	6	7	8	9	10
Payoff	0	0	0	0	0	1	22	150	1000	5000	100,000



23. The payoff table for $m = 11$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 11$

Catches	0	1	2	3	4	5	6	7	8	9	10	11
Payoff	0	0	0	0	0	0	8	80	400	2500	25,000	100,000



24. The payoff table for $m = 12$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 12$

Catches	0	1	2	3	4	5	6	7	8	9	10	11	12
Payoff	0	0	0	0	0	0	5	32	200	1000	5000	25,000	100,000



25. The payoff table for $m = 13$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 13$

Catches	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Payoff	1	0	0	0	0	0	1	20	80	600	3500	10,000	50,000	100,000



26. The payoff table for $m = 14$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 14$

Catches	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Payoff	1	0	0	0	0	0	1	9	42	310	1100	8000	25,000	50,000	100,000



27. The payoff table for $m = 15$ is given below. Compute the probability density function, mean, and standard deviation of the payoff.

Pick $m = 15$

Catches	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Payoff	1	0	0	0	0	0	0	10	25	100	300	2800	25,000	50,000	100,000	100,000



In the exercises above, you should have noticed that the expected payoff on a unit bet varies from about 0.71 to 0.75, so the expected profit (for the gambler) varies from about -0.25 to -0.29 . This is quite bad for a casino game, but as always, the lure of a very high payoff on a small bet for an extremely rare event overrides the expected value analysis for most players.

28. With $m = 15$, show that the top 4 prizes (25,000, 50,000, 100,000, 100,000) contribute only about 0.017 (less than 2 cents) to the total expected value of about 0.714.

On the other hand, the standard deviation of the payoff varies quite a bit, from about 1 to about 55.

29. Although the game is highly unfavorable for each m , with expected value that is nearly constant, which do you think is better for the gambler--a format with high standard deviation or one with low standard deviation?