# Week 5: Expected value and Betting systems

**Random variable** A random variable represents a "measurement" in a random experiment. We usually denote random variable with capital letter $X, Y, \cdots$. If $S$ is the sample space of the experiment then to each $i \in S$ the random variable $X$ assigns a certain value $\alpha$ ( a real number). The random variable is described by its probability distribution

$$P(X = \alpha)$$

for all possible values $\alpha$ that the random variable can take. Of course we have

$$0 \le P(X = \alpha) \le 1 \quad \text{and} \quad \sum_{\alpha} P(X = \alpha) \,=\, 1$$

**Example:** If you roll a pair of dice consider the random variable

$$X = \text{ sum of the two dice}$$

Then $X$ takes values $2, 3, \cdots, 12$ and $P(X = 2) = 1/36$, $P(X = 3) \,=\, 2/36$ etc.. ∎

**Expected value of a random variable** For a random variable $X$ the *expected value of* $X$ is the *average value* of $X$ which we denote by $E[X]$. It is given by

$$\boxed{\textbf{Expected value of } X: \quad E[X] \,=\, \sum_{\alpha} \alpha \, P(X = \alpha)}$$

**Chuck-a-luck** This game (found in fairgrounds) is played by rolling 3 dice and betting on a number between 1 and 6. You win your bet multiplied by the number of times your chosen appear on the the three dice. For example if you bet \$1 on 5 and roll $4, 5, 5$ you win \$2. A quick look at this game may make it appear reasonably fair. Since you roll 3 dice and there seems to be a probability 1/2 that your chosen number appears and so the odds should be in your favor. For a second look let us compute your expected gain $E[W]$ at this game. Suppose you bet on five, the probability to get 3 fives is $(1/6)^3$, the probability to get 2 fives is $3(1/6)^2(5/6)$, etc.... and we obtain

$$\begin{aligned} E[W] &= 3P(3 \text{ fives}) + 2P(2 \text{ fives}) + 1P(1 \text{ five}) - 1P(0 \text{ five}) \\ &= 3 \cdot \frac{1}{216} + 2 \cdot \frac{15}{36} + 1 \cdot \frac{75}{36} - 1 \cdot \frac{125}{216} \,=\, -\frac{17}{216} \,=\, -.079 \end{aligned}$$

That is you loose around 8 cents on the dollar at this game. ■

**The binomial random variable** The binomial random variable counts the number of successes when performing $n$ independent trials, each of which has a probability of success $p$. We have then

$$X \text{ takes values } 0, 1, 2 \cdots, n$$

and

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

To compute the expected value of $X$ we will use the binomial theorem which states that $(x+y)^n \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$. From this we see that

$$1 = (p + (1-p)) = (p + (1-p))^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

and this shows that $\sum_{k=0}^n P(X = k) = 1$ as it should be. To compute the expectation we will use the identity

$$\binom{n}{k} k = \binom{n-1}{k-1} n$$

seen in Chapter 1 (select a football team and a captain for the team). Furthermore we have

$$\begin{aligned}
E[X] &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j} \\
&= np
\end{aligned}$$

and the thus we have

The expected value of the binomial random variable is $np$

**Example: Expected gain at roulette**. At the Las Vegas roulette (with 38 numbers, 0,00,1,2,3, etc) you can do various bets (let's say the bet size is \$1).

1. Bet on red (or black) and a successful bet pays you \$1.
2. Bet on a number and a successful bet pays you \$35.
3. Bet on the first (or the second, the third) dozen of numbers a successful bet pays you \$2.
4. etc $\cdots$

You can find the list of all bets and payouts for Las Vegas and Monte-Carlo roulette at `http://en.wikipedia.org/wiki/Roulette`

You may wonder which of these bets is the more advantageous and so we compute the *expected gain* $E[W]$ for each bet

$$\begin{aligned}
\text{Red} \quad E[W] &= 1 \cdot \frac{18}{38} - 1\frac{20}{38} = -\frac{2}{38} = -0.0526 \\
\text{Number} \quad E[W] &= 35 \cdot \frac{1}{38} - 1\frac{37}{38} = -\frac{2}{38} \\
\text{Dozen} \quad E[W] &= 2 \cdot \frac{12}{38} - 1\frac{26}{38} = -\frac{2}{38}
\end{aligned}$$

All these bets (and all the other ones) are devised to give the same odds. It does not matter how you play, you shall loose on average around 5.3 cents for each dollar you bet.

If you bet on a group of $n$ numbers then the payout is

$$\text{Payout for a bet on } n \text{ numbers} = \frac{36}{n} - 1$$

and for such a bet the expected gain is

$$E[W] = \left(\frac{36}{n} - 1\right) \cdot \frac{n}{38} - 1 \cdot \frac{38-n}{38} = \frac{36}{38} - 1 = -\frac{2}{38}$$

■

**The hypergeometric random variable** The hypergeometric random variable appears in the game of Keno. In a general way imagine you have $N$ balls, $K$ of which are red and $N-K$ of which are black. Out of these $N$ balls you select $n$ balls without replacement. The hypergeometric random variable counts the number of red balls you select. So

$$\text{X takes the values } 0, 1, 2, \cdots, K$$

and

$$P(X = k) \;=\; \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

**Keno 10 spot card** Recall that in Keno the casino draws 20 numbers randomly out of 80 numbers. In a $m$ spot card you pick $m$ numbers and if $k$ of your $m$ numbers match the casino numbers you have a "catch of $k$". This is hypergeometric with $N = 80$, $K = m$ (the red balls are the number you chose) and $n = 20$. We have for $m = 10$

$$P(\text{catch of } k) \;=\; \frac{\binom{10}{k}\binom{70}{20-k}}{\binom{80}{20}}$$

The probability and payouts for Keno vary a bit from place to place: for a 10 spot card the payouts by the Massachusetts lottery are (see `http://www.masslottery.com/games/keno.html` for all the payouts )

| Match | Payout |
|---|---|
| 10 | 100'000 |
| 9 | 10'000 |
| 8 | 500 |
| 7 | 80 |
| 6 | 20 |
| 5 | 2 |
| 0 | 2 |

So for a bet of 1 dollar the expected amount paid by the lottery is

$$\begin{aligned}
E[W] \;&=\; 100000 \cdot \frac{\binom{10}{10}\binom{70}{10}}{\binom{80}{20}} + 10000 \cdot \frac{\binom{10}{9}\binom{70}{11}}{\binom{80}{20}} + 500 \cdot \frac{\binom{10}{8}\binom{70}{12}}{\binom{80}{20}} + 80 \cdot \frac{\binom{10}{7}\binom{70}{13}}{\binom{80}{20}} \\
&\quad +20 \cdot \frac{\binom{10}{6}\binom{70}{2014}}{\binom{80}{20}} + 2 \cdot \frac{\binom{10}{5}\binom{70}{15}}{\binom{80}{20}} + +2 \cdot \frac{\binom{10}{0}\binom{70}{20}}{\binom{80}{20}} \\
&= \; 0.0112211 \quad + \quad 0.0612064 \quad + \quad 0.0677096 \quad + \quad 0.1288914 \\
&\quad +0.2295878 \quad + \quad 0.1028553 \quad + \quad 0.0915814 \\
&= \; 0.6930534 \qquad (1)
\end{aligned}$$

The lottery keeps then more than 30 cents (!) of each dollar played on Keno and the number above tells you where the payout are. For example 22 cents on a dollar are given as payout for a catch of 6 while only about one cent as payout for a catch of 10, and so on.. ∎

**The martingale betting system**. Let us explain this betting system by an example. You just receive the news that you inherited from a long lost relative the nice sum of $\$2,550,000 = (2^8 - 1) \cdot 10,000)$. You move immediately to Atlantic city and devise the following gambling scheme. Every month you go to the craps table and bet $\$10,000$. If you win, you just won $\$10,000$ and you quit and live off your money for a month. Now if you loose play again you double your bet to $\$20,000$. If you win your second bet then your net win is $\$20,000 - \$10,000 = \$10,000$. Again if you loose you double you bet, etc.... The *martingale betting system* consists then of doubling your bet until your first win. In any case if you win at your $k^{th}$ bet then your net gain is

$$10'000 \left[-1 - 2 - 2^{k-1} + 2^k\right] \; = \; 10'000$$

using the geometric series $1 + x + \cdots + x^{k_1} = \frac{1-x^k}{1-x}$ with $x = 2$. So with the martingale betting system, you do win $10'000$ every time.

If you have unlimited resources (and if the casino has no betting limit) you could in principle make money using subfair games. But of course none of this condition is true. Suppose that like in our example you can bet at most $n$ bets in a row before running out of money. If the probability to lose any single game is $q$ then

$$\text{Probability to lose everything } = q^n$$

since to lose everything you need to loose $n$ times in a row. Let us compute the expected gain $W$ playing the game this way. We have

$$\boxed{E[W] \; = \; 10'000 \cdot (1 - q^n) - (2^n - 1)10'000 \cdot q^n \; = \; 10'000 \, [1 - (2q)^n]}$$

If $n = 8$ and the game were fair $q = 1/2$ then the probability to lose everything on a single month is $1/256 = 0.0039$ and the expected gain is 0. If you play craps for which $q = 251/295$ then the probability to lose everything 0.0044 and the expected gain is $-\$1188.92$ which is abut %11 of your bet size, which is not very good. To compare various $n$ note that for craps

| n | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|
| E[W] / 10,000 | -0.08 | -0.10 | -0.11 | -0.13 | -0.15 |

so actually the more money you have to play the martingale strategy, the more you lose on average. The idea behind this betting (and many other) betting system is to make sure that you win (a little) with high probability and to make you forget that when you eventually lose, you do lose a lot.

To analyze the martingale betting a bit better, let us compute how many times, on average, should one play to see $n$ loss in a row. We shall do this using a first order difference equation. We define

$$x(n) = \text{ Expected number of games until you lose n times in a row}$$

Let us start with $x(1)$: if the first game is a loss (with probability $q$) then $x(1) = 1$, while of the first game is a win( with probability $p$), then the expected number of games until 1 loss will be $1 + x(1)$ so that we have the equation

$$x(1) = 1 \times q + (x(1) + 1) \times p\,.$$

which is easily solved to give $x(1) = \frac{1}{q}$. For $x(n)$ let us concentrate what happen after we have sustained $n-1$ loss in a row (this took $x(n-1)$ games on average). If we lose (with probability $q$) then we need $x(n-1)+1$ games to loose $n$ in a row. But if we win that game, we start afresh and it will take then $x(n-1)+1+x(n)$ games to reach $n$ losses. That is we have the equation

$$x(n) \;=\; [x(n-1)+1] \times q + [x(n-1)+1+x(n)]p$$

which gives the difference equation

$$x(n) \;=\; \frac{1}{q}x(n-1) + \frac{1}{q}\,.$$

with the initial condition $x(1) = 1/q$. Using the methods in Chapter 3 we find the solution

$$x(n) \;=\; \frac{1}{p}\left[\left(\frac{1}{q}\right)^n - 1\right]\,.$$

For craps we get for example

| n | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|
| x(n) | 117.3 | 233 | 462 | 913 | 1803 |

So if you play at roulette 500 hundred time, you should expect to loose 8 times in a row! In the martingale betting system during a single game at the casino, you play until you win once, so it take on average $1/p$ games to achieve that. So on average you will visit the casino $(1/q)^n - 1$ times (227 times with $n$=8) before you go bust. You could as well spend \$10,000 every month and your money will last longer, namely 255 weeks.

**The coupon collecting problem and the geometric random variable**. The following problem appears in many different forms. Imagine that in every cereal box there is a free toy. Toys come in $n$ different types and are all equally likely to be put in any box. The goal being to collect all $n$ toys, how many cereal boxes will you have to buy, on average?

In order to do this let us consider a *geometric random variable* $N$: imagine a series of independent bets, each resulting in a success with probability $p$. The geometric random variable $N$ is equal to $n$ if your first success occurs on the $n^{th}$ bet. So we have

$$P\{N = n\} = (1-p)^{n-1}p\,, \quad n = 1, 2, 3, \cdots$$

The expected value of $N$ is

$$E[N] \;=\; \sum_{n=1}^{\infty} n(1-p)^{n-1}p\,.$$

To compute this series recall the geometric series $\sum_{n=0}^{\infty} x^n \;=\; (1+x+x^2+x^3+\cdots) \;=\; \frac{1}{1-x}$ which holds for $|x| < 1$. If we differentiate with respect to $x$ we have

$$\frac{d}{dx}\sum_{n=0}^{\infty} x^n \;=\; \sum_{n=1}^{\infty} nx^{n-1} \;=\; \frac{d}{dx}\left(\frac{1}{1-x}\right) \;=\; \frac{1}{(1-x)^2}\,.$$

Using this we find that

$$E[N] \;=\; \frac{1}{(1-(1-p))^2}p \;=\; \frac{1}{p}\,.$$

Returning to the coupon collecting problem we let $T$ be the total number of boxes needed to collect the $n$ toys and we write it as

$$T \;=\; T_1 + T_2 + \cdots + T_n$$

where

$$T_i \;=\; \text{ number of boxes needed to obtain the } i^{th} \text{ after having collected } i-1 \text{ toys}$$

We have $T_1 = 1$, $T_2$ is a geometric random variable with probability $p = \frac{n-1}{n}$ to find a new coupon in any box, $T_3$ is geometric with $p = \frac{n-2}{n}$, and so on. So we find

$$\begin{aligned} E[T] \;&=\; E[T_1 + T_2 + \cdots + T_n] \;=\; E[T_1] + \cdots + E[T_n] \\ &=\; 1 + \frac{n}{n-1} + \frac{n}{n-2} + \cdots \frac{n}{1} \;=\; n(1 + \frac{1}{2} + \cdots + \frac{1}{n}) \end{aligned}$$

As we learned in calculus the geometric series $1 + \frac{1}{2} + \cdots \frac{1}{n} \approx \ln(n)$ for large $n$ and so we have

$$E[T] \;\approx n\ln(n)\,.$$

We get for example

| $n$ | 10 | 20 | 100 | 500 | 1000 |
|---|---|---|---|---|---|
| $n\ln(n)$ | 23.02 | 59.91 | 460.51 | 3107.30 | 6907.75 |

**The Poisson random variable.** The Poisson random variables $X$ takes value $0, 1, 2, 3, \cdots$, and

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} .$$

We use the series for the exponential $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ to show that

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} e^{\lambda} = 1.$$

and for the expectation we have

$$\begin{aligned} E[X] &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda \end{aligned}$$

**The Poisson random variable and the law of small numbers**. One of the reason the Poisson random variable is ubiquitous in nature is that if we have *a very large number* $n \gg 1$ of independent or (almost) independent events, each of which has a *very small probability* $p \ll 1$ to successfully occur, then the the number of success is very well approximated by a Poisson random variable with parameter $\lambda = np$.

A formal way to see this is to consider a binomial random variable with $n$ trials and probability of success $p = p(n)$ and to take

$$n \to \infty \text{ and } p \to 0 \text{ assuming that } np \to \lambda.$$

We do this by choosing simply $p = \frac{\lambda}{n}$ and we have

$$\begin{aligned} \lim_{n \to \infty} \binom{n}{k} p^k (1-p)^{n-k} &= \lim_{n \to \infty} \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{n \to \infty} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n(n-1)\cdots(n-k+1)}{n^n} \left(1 - \frac{\lambda}{n}\right)^{-k} \end{aligned}$$

We use then from calculus that $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$ for any $x$. But if $k$ is fixed we have $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1$. In addition we have

$$\lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{n^n} = \lim_{n\to\infty} 1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) = 1$$

Putting this altogether gives

$$\lim_{n\to\infty} \binom{n}{k} p^k(1-p)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}\,.$$

which is called the *the law of small numbers.*

**Playing Powerball**. In its current version to win the jackpot at Powerball you need to select correctly 5 out of 69 numbers and 1 out 26 numbers (this is the powerball). So

$$\text{Probability to win the jackpot} = \frac{1}{26\binom{69}{5}} = \frac{1}{292,201,338}$$

In 2016 the jackpot hit the record 1.5 billion and 371 million tickets were sold and there was one single winner. Assuming that people choose numbers at random (which is not quite true) we have

$$n = 371,000,000 \quad p = \frac{1}{292,201,338} \text{ and so } \lambda = np = 1.2696$$

We can use the Poisson approximation for the number of winners $X$ and obtain

$$\text{Prob. of at least one winner} = P(X \geq 1) = 1 - P(X=0) = 1 - e^{-\lambda} = .719$$

and

$$\text{Prob. of at least two winners} = P(X \geq 2) = 1 - P(X=0) - P(X=1) = 1 - e^{-\lambda} - \lambda e^{-\lambda} = 0.3625$$

**Playing lottery and amazing coincidences**. In the 1980's a woman in New Jersey won twice at lottery. What an amazing coincidence no? At that time the lottery was to choose 6 out 42 number and so for one person to win twice at the lottery is

$$\frac{1}{\binom{42}{6}} \times \frac{1}{\binom{42}{6}} = \frac{1}{5,245,786} \times \frac{1}{5,245,786} = \frac{1}{2.75\times 10^{13}}$$

that is a chance of one in 27 trillions!

But this is not the relevant quantity we have many people playing lottery very many times. To fix the idea imagine that everyone plays lottery 1,000 times (for example buying five tickets a week for 4 years). Then to compute the probability that one given player will play at least twice is approximately Poisson with

$$\lambda_0 = 1000 \times \frac{1}{\binom{42}{6}} = 1.906 \times 10^{-4}$$

and thus

$$P(1 \text{ given player wins at least twice}) = 1 - e^{-\lambda_0} - e^{-\lambda_0}\lambda_0 = 1.817 \times 10^{-8}$$

Now if, say, 50 millions people play the lottery, each thousand times, the probability that one of those will win twice is is approximately Poisson with

$$\lambda = 50,000,000 \times 1.817 \times 10^{-8} = 0.9084$$

and so the probability of at least one double winner is about $1 - e^{-\lambda} = .5968$. Not so amazing.

**The birthday problem revisited** In the birthday problem we ask for the probability for at least 2 out of $m$ people to have the same birthday. To do this we form $n = \binom{m}{2} = \frac{m(m-1)}{2}$ pairs of people which is large if $m$ is not too small. For each pair the probability they have the same birthday is 1/365 which is small and if we assume that the pair are almost independent and use the Poisson approximation with $\lambda = \binom{m}{2}/365$ we obtain

$$P(\text{at least two same birthday}) = 1 - e^{-\frac{1}{2}m(m-1)}$$

which coincide what we computed with another method in Chapter 1.