Course: M362K Intro to Stochastic Processes
Term: Fall 2014
Instructor: Gordan Zitkovic

Lecture 4
The Simple Random Walk

We have defined and constructed a random walk $\{X_n\}_{n\in\mathbb{N}_0}$ in the previous lecture. Our next task is to study some of its mathematical properties. Let us give a definition of a slightly more general creature.

Definition 4.1. A sequence $\{X_n\}_{n\in\mathbb{N}_0}$ of random variables is called a **simple random walk** (with parameter $p \in (0,1)$) if

1. $X_0 = 0$,
2. $X_{n+1} - X_n$ is independent of $(X_0, X_1, \dots, X_n)$ for all $n \in \mathbb{N}$, and
3. the random variable $X_{n+1} - X_n$ has the following distribution

$$\begin{pmatrix} -1 & 1 \\ q & p \end{pmatrix}$$

where, as usual, $q = 1 - p$.

If $p = \frac{1}{2}$, the random walk is called **symmetric**.

The adjective *simple* comes from the fact that the size of each step is fixed (equal to 1) and it is only the direction that is random[1].

[1] One can study more general random walks where each step comes from an arbitrary prescribed probability distribution.

Proposition 4.2. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple random walk with parameter p. For $n \geq 1$ the distribution of the random variable X_n is discrete with support $\{-n, -n+2, \dots, n-2, n\}$, and probabilities $p_l = \mathbb{P}[X_n = l]$ given by*

$$p_l = \binom{n}{\frac{l+n}{2}} p^{(n+l)/2} q^{(n-l)/2}, \tag{4.1}$$

for $l = -n, -n+2, \dots, n-2, n$.

Proof. X_n is composed of n independent steps $\xi_k = X_k - X_{k-1}$, $k = 1, \dots, n$, each of which goes either up or down. In order to reach level l in those n steps, the number u of up-steps and the number d of downsteps must satisfy $u - d = l$ (and $u + d = n$). Therefore, $u = \frac{n+l}{2}$ and $d = \frac{n-l}{2}$. The number of ways we can choose these u up-steps from the total of n is $\binom{n}{\frac{n+l}{2}}$, which, with the fact the probability of any trajectory with exactly u up-steps is $p^u q^{n-u}$, gives the probability (4.1) above. Equivalently, we could have noticed that the random variable $\frac{n+X_n}{2}$ has the binomial $b(n,p)$-distribution. □

The proof of Proposition 4.2 uses the simple idea already hinted at in the previous lecture: view the random walk as a random trajectory in some space of trajectories, and, compute the required probability by simply counting the number of trajectories in the subset (event) you are interested in, and adding them all together, weighted by their probabilities. To prepare the ground for the future results, let C be the set of all possible trajectories:

$$C = \{(x_0, x_1, \dots, x_n) \,:\, x_0 = 0,\ x_{k+1} - x_k = \pm 1, k \leq n-1\}.$$

You can think of the first n steps of a random walk simply as a probability distribution on the state-space C.

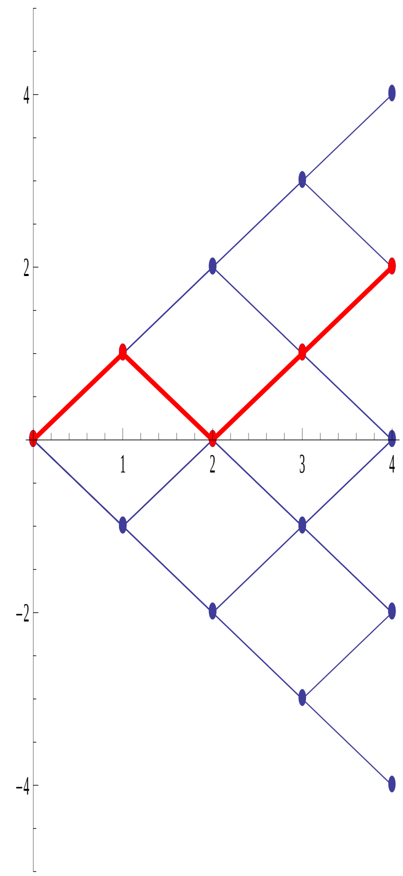

Figure 1: The superposition of all trajectories in C for $n = 4$ and a particular one - $(0,1,0,1,2)$ - in red.

Now we know how to compute the probabilities related to the position of the random walk $\{X_n\}_{n\in\mathbb{N}_0}$ at a fixed future time n. A mathematically more interesting question can be posed about the maximum of the random walk on $\{0,1,\dots,n\}$. More precisely, for $n \in \mathbb{N}_0$ we define its **running maximum process** $\{M_n\}_{n\in\mathbb{N}_0}$ by

$$M_n = \max(X_0, \dots, X_n), \quad \text{for } n \in \mathbb{N}_0.$$

A nice expression for the pmf of M_n is available for the case of *symmetric* simple random walks.

Proposition 4.3. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a symmetric simple random walk. Then $M_0 = 0$ and for $n \geq 1$, the support of M_n is $\{0,1,\dots,n\}$ and its probability mass function is given by*

$$\begin{aligned} p_l = \mathbb{P}[M_n = l] &= \mathbb{P}[X_n = l] + \mathbb{P}[X_n = l+1] \\ &= \binom{n}{\lfloor \frac{n+l+1}{2} \rfloor} 2^{-n}, \quad \textit{for } l = 0, \dots, n. \end{aligned}$$

Proof. Let us first pick a level $l \in \{0,1,\dots,n\}$ and compute the auxiliary probability $q_l = \mathbb{P}[M_n \geq l]$ by counting the number of trajectories whose maximal level reached is at least l. Indeed, the symmetry assumption ensures that all trajectories are equally likely. More precisely, let $A_l \subset C_0(n)$ be given by

$$\begin{aligned} A_l &= \{(x_0, x_1, \dots, x_n) \in C \,:\, \max_{k=0,\dots,n} x_k \geq l\} \\ &= \{(x_0, x_1, \dots, x_n) \in C \,:\, x_k \geq l, \text{ for at least one } k \in \{0,\dots,n\}\}. \end{aligned}$$

Then $\mathbb{P}[M_n \geq l] = \frac{1}{2^n} \# A_l$, where $\#A$ denotes the number of elements in the set A. When $l = 0$, we clearly have $\mathbb{P}[M_n \geq 0] = 1$, since $X_0 = 0$.

To count the number of elements in A_l, we use the following clever observation (known as the **reflection principle**):

Claim 4.4. For $l \in \mathbb{N}$, we have

$$\#A_l = 2\#\{(x_0, x_1, \dots, x_n) : x_n > l\} + \#\{(x_0, x_1, \dots, x_n) : x_n = l\}.$$

Proof Claim 4.4. We start by defining a bijective transformation which maps trajectories into trajectories. For a trajectory $(x_0, x_1, \dots, x_n) \in A_l$, let $k(l) = k(l, (x_0, x_1, \dots, x_n))$ be the smallest value of the index k such that $x_k \geq l$. In the stochastic-process-theory parlance, $k(l)$ is the **first hitting time of the set** $\{l, l+1, \dots\}$. We know that $k(l)$ is well-defined (since we are only applying it to trajectories in A_l) and that it takes values in the set $\{1, \dots, n\}$. With $k(l)$ at our disposal, let $(y_0, y_1, \dots, y_n) \in C$ be a trajectory obtained from $(x_0, x_1, \dots, x_n)$ by the following procedure:

1. do nothing until you get to $k(l)$:
   - $y_0 = x_0$,
   - $y_1 = x_1, \dots$
   - $y_{k(l)} = x_{k(l)}$.
2. use the flipped values for the coin-tosses from $k(l)$ onwards:
   - $y_{k(l)+1} - y_{k(l)} = -(x_{k(l)+1} - x_{k(l)})$,
   - $y_{k(l)+2} - y_{k(l)+1} = -(x_{k(l)+2} - x_{k(l)+1}), \dots$
   - $y_n - y_{n-1} = -(x_n - x_{n-1})$.

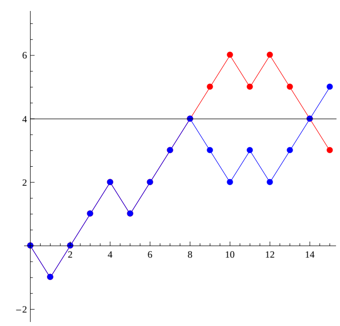

Figure 2: A trajectory (blue) and its image (red) under Φ.

The picture on the right shows two trajectories: a blue one and its reflection in red, with $n = 15$, $l = 4$ and $k(l) = 8$. Graphically, $(y_0, \dots, y_n)$ looks like $(x_0, \dots, x_n)$ until it hits the level l, and then follows its reflection around the level l so that $y_k - l = l - x_k$, for $k \geq k(l)$. If $k(l) = n$, then $(x_0, x_1, \dots, x_n) = (y_0, y_1, \dots, y_n)$. It is clear that $(y_0, y_1, \dots, y_n)$ is in C. Let us denote this transformation by

$$\Phi : A_l \to C, \ \Phi(x_0, x_1, \dots, x_n) = (y_0, y_1, \dots, y_n)$$

and call it the **reflection map**. The first important property of the reflexion map is that it is its own inverse: apply Φ to any $(y_0, y_1, \dots, y_n)$ in A_l, and you will get the original $(x_0, x_1, \dots, x_n)$. In other words $\Phi \circ \Phi = \text{Id}$, i.e. Φ is an involution. It follows immediately that Φ is a bijection from A_l onto A_l.

To get to the second important property of Φ, let us split the set A_l into three parts according to the value of x_n:

1. $A_l^{>} = \{(x_0, x_1, \dots, x_n) \in A_l : x_n > l\}$,

2. $A_l^= = \{(x_0, x_1, \dots, x_n) \in A_l : x_n = l\}$, and

3. $A_l^< = \{(x_0, x_1, \dots, x_n) \in A_l : x_n < l\}$,

So that

$$\Phi(A_l^>) = A_l^<,\ \Phi(A_l^<) = A_l^>,\ \text{and}\ \Phi(A_l^=) = A_l^=.$$

We should note that, in the definition of $A_l^>$ and $A_l^=$, the a priori stipulation that $(x_0, x_1, \dots, x_n) \in A_l$ is unncessary. Indeed, if $x_n \geq l$, you must already be in A_l. Therefore, by the bijectivity of Φ, we have

$$\#A_l^< = \#A_l^> = \#\{(x_0, x_1, \dots, x_n) : x_n > l\},$$

and so

$$\#A_l = 2\#\{(x_0, x_1, \dots, x_n) : x_n > l\} + \#\{(x_0, x_1, \dots, x_n) : x_n = l\},$$

just as we claimed. □

Now that we know that Claim 4.4 holds, we can easily rewrite it as follows:

$$\begin{aligned}\mathbb{P}[M_n \geq l] &= \mathbb{P}[X_n = l] + 2\sum_{j>l} \mathbb{P}[X_n = j] \\ &= \sum_{j>l} \mathbb{P}[X_n = j] + \sum_{j \geq l} \mathbb{P}[X_n = j].\end{aligned}$$

Finally, we subtract $\mathbb{P}[M_n \geq l+1]$ from $\mathbb{P}[M_n \geq l]$ to get the expression for $\mathbb{P}[M_n = l]$:

$$\mathbb{P}[M_n = l] = \mathbb{P}[X_n = l+1] + \mathbb{P}[X_n = l].$$

It remains to note that only one of the probabilities $\mathbb{P}[X_n = l+1]$ and $\mathbb{P}[X_n = l]$ is non-zero, the first one if n and l have different parity and the second one otherwise. In either case the non-zero probability is given by

$$\binom{n}{\lfloor \frac{n+l+1}{2} \rfloor} 2^{-n}.$$

□

Let us use the reflection principle to solve a classical problem in combinatorics.

Example 4.5 (The Ballot Problem)**.** Suppose that two candidates, Daisy and Oscar, are running for office, and $n \in \mathbb{N}$ voters cast their ballots. Votes are counted by the same official, one by one, until all n of them have been processed[2]. After each ballot is opened, the official records the number of votes each candidate has received so far. At the end, the official announces that Daisy has won by a margin of $m > 0$ votes, i.e., that Daisy got $(n+m)/2$ votes and Oscar the remaining $(n-m)/2$

[2] like in the old days.

votes. What is the probability that at no time during the counting has Oscar been in the lead?

We assume that the order in which the official counts the votes is completely independent of the actual votes, and that each voter chooses Daisy with probability $p \in (0,1)$ and Oscar with probability $q = 1 - p$. For $k \leq n$, let X_k be the number of votes received by Daisy *minus* the number of votes received by Oscar in the first k ballots. When the $k+1$-st vote is counted, X_k either increases by 1 (if the vote was for Daisy), or decreases by 1 otherwise. The votes are independent of each other and $X_0 = 0$, so X_k, $0 \leq k \leq n$ is (the beginning of) a simple random walk. The probability of an up-step is $p \in (0,1)$, so this random walk is not necessarily symmetric. The ballot problem can now be restated as follows:

What is the probability that $X_k \geq 0$ for all $k \in \{0, \dots, n\}$, given that $X_n = m$?

The first step towards understanding the solution is the realization that the exact value of p does not matter. Indeed, we are interested in the conditional probability $\mathbb{P}[F|G] = \mathbb{P}[F \cap G]/\mathbb{P}[G]$, where F denotes the family of all trajectories that always stay non-negative and G the family of those that reach m at time n. Each trajectory in G has $(n+m)/2$ up-steps and $(n-m)/2$ down-steps, so its probability weight is always equal to $p^{(n+m)/2}q^{(n-m)/2}$. Therefore,

$$\mathbb{P}[F|G] = \frac{\mathbb{P}[F \cap G]}{\mathbb{P}[G]} = \frac{\#(F \cap G)\ p^{(n+m)/2}q^{(n-m)/2}}{\#G\ p^{(n+m)/2}q^{(n-m)/2}} = \frac{\#(F \cap G)}{\#G}. \quad (4.2)$$

We already know how to count the number of paths in G - it is equal to $\binom{n}{(n+m)/2}$ - so "all" that remains to be done is to count the number of paths in $G \cap F$.

The paths in $G \cap F$ form a portion of all the paths in G which don't hit the level $l = -1$, so that $\#(G \cap F) = \#G - \#H$, where H is the set of all paths which finish at m, but cross (or, at least, touch) the level $l = -1$ in the process. Can we use the reflection principle to find $\#H$? Yes, we do. In fact, you can convince yourself that the reflection of any path in H around the level $l = -1$ after its first hitting time of that level poduces a path that starts at 0 and ends at $-m-2$. Conversely, the same procedure applied to such a path yields a path in H. The number of paths from 0 to $-m-2$ is easy to count - it is equal to $\binom{n}{(n+m)/2+1}$. Putting everything together, we get

$$\mathbb{P}[F|G] = \frac{\binom{n}{k} - \binom{n}{k+1}}{\binom{n}{k}} = \frac{2k+1-n}{k+1}, \text{ where } k = \frac{n+m}{2}.$$

The last equality follows from the definition of binomial coefficients $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

The Ballot problem has a long history (going back to at least 1887) and has spurred a lot of research in combinatorics and probability. In fact, people still write research papers on some of its generalizations. When posed outside the context of probability, it is often phrased as "*in how many ways can the counting be performed ...*" (the difference being only in the normalizing factor $\binom{n}{k}$ appearing in (4.2) above). A special case $m = 0$ seems to be even more popular - the number of $2n$-step paths from 0 to 0 never going below zero is called the **Catalan number**[3] and equals to

$$C_n = \frac{1}{n+1}\binom{2n}{n}. \tag{4.3}$$

[3] See Problem 4.6 for more information about Catalan numbers.

PROBLEMS

Problem 4.1. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a symmetric simple random walk. Compute the following

1. $\mathbb{P}[X_{2n} = 0]$, $n \in \mathbb{N}_0$,
2. $\mathbb{P}[X_n = X_{2n}]$, $n \in \mathbb{N}_0$,
3. $\mathbb{P}[\,|X_1 X_2 X_3| = 2]$,
4. $\mathbb{P}[X_7 + X_{12} = X_1 + X_{16}]$.

Solution: Let $\xi_k = X_k - X_{k-1}$, $k \in \mathbb{N}$.

1. $\mathbb{P}[X_{2n} = 0] = 2^{-n}\binom{2n}{n}$.
2. $\mathbb{P}[X_n = X_{2n}] = \mathbb{P}[\xi_{n+1} + \xi_{n+2} + \dots + \xi_{2n} = 0] = \begin{cases} 0, & n \text{ is odd}, \\ 2^{-n/2}\binom{n}{n/2}, & n \text{ is even}. \end{cases}$
3. $|X_1 X_2 X_3| = 2$ only in the following two cases: $X_1 = 1, X_2 = 2, X_3 = 1$ or $X_1 = -1, X_2 = -2, X_3 = -1$. The probability of each trajectory is $\frac{1}{8}$, so $\mathbb{P}[|X_1 X_2 X_3| = 2] = \frac{1}{4}$.
4. We write the event $\{X_1 + X_{12} = X_7 + X_{16}\}$ in terms of ξs:

   $$\{X_7 + X_{12} = X_1 + X_{16}\} = \{\xi_2 + \xi_3 + \dots + \xi_7 = \xi_{13} + \xi_{14} + \xi_{15} + \xi_{16}\}.$$

   All ξs are independent, have the same distribution and ξ_k has the same distribution as $-\xi_k$. Thus,

   $$\begin{aligned} &\mathbb{P}[\xi_2 + \xi_3 + \dots + \xi_7 = \xi_{13} + \xi_{14} + \xi_{15} + \xi_{16}] \\ &\quad = \mathbb{P}[\xi_2 + \xi_3 + \dots + \xi_7 - \xi_{13} - \xi_{14} - \xi_{15} - \xi_{16} = 0] \\ &\quad = \mathbb{P}[\xi_1 + \xi_2 + \dots + \xi_{10} = 0] = 2^{-10}\binom{10}{5}. \end{aligned}$$

Problem 4.2. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple symmetric random walk. Which of the following processes are simple random walks?

1. $\{2X_n\}_{n\in\mathbb{N}_0}$?
2. $\{X_n^2\}_{n\in\mathbb{N}_0}$?
3. $\{-X_n\}_{n\in\mathbb{N}_0}$?
4. $\{Y_n\}_{n\in\mathbb{N}_0}$, where $Y_n = X_{5+n} - X_5$?

How about the case $p \neq \frac{1}{2}$?

Solution:

1. No - the distribution of X_1 has support $\{-2, 2\}$ and not $\{-1, 1\}$.
2. No - $X_1^2 = 1$, and not ± 1 with equal probabilities.
3. Yes - check the definition.
4. Yes - check the definition.

The answers are the same if $p \neq \frac{1}{2}$, but, in 3., $-X_n$ comes with probability $1 - p$.

Problem 4.3. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple symmetric random walk. Given $n \in \mathbb{N}$, what is the probability that X does not visit 0 during the time interval $1, \dots, n$.

Solution: Let us denote the required probability by p_n, i.e.,

$$p_n = \mathbb{P}[X_1 \neq 0, X_2 \neq 0, \dots, X_n \neq 0].$$

For $n = 1$, $p_1 = 1$, since X_1 is either 1 or -1. For $n > 1$, let ξ_1 be the first increment $\xi_1 = X_1 - X_0 = X_1$. If $\xi_1 = -1$, we need to compute that probability that a random walk of length $n - 1$, starting at -1, does not hit 0. This probability is, in turn, the same as the probability that a random walk of length $n - 1$, starting from 0, never hits 1. By the symmetry of the increments, the same reasoning works for the case $\xi_1 = 1$. Therefore,

$$p_n = \mathbb{P}[X_1 \leq 0, X_2 \leq 0, \dots, X_{n-1} \leq 0] = \mathbb{P}[M_{n-1} = 0],$$

where $M_n = \max\{X_0, \dots, X_n\}$. By Proposition 4.3, this probability is given by

$$p_n = 2^{-n+1}\binom{n-1}{\lfloor n/2 \rfloor}.$$

Problem 4.4. (30pts) A fair coin is tossed repeatedly and the record of the outcomes is kept. Tossing stops the moment the total number of heads obtained so far exceeds the total number of tails by 3. For example, a possible sequence of tosses could look like *HHTTTHHTHHTHH*. What is the probability that the length of such a sequence is at most 10?

Solution: Let X_n, $n \in \mathbb{N}_0$ be the number of heads *minus* the number of tails obtained so far. Then, $\{X_n\}_{n\in\mathbb{N}_0}$ is a simple symmetric random walk, and we stop tossing the coin when X hits 3 for the first time. This will happen during the first 10 tosses, if and only if $M_{10} \geq 3$, where M_n denotes the (running) maximum of X. According to the reflection principle,

$$\begin{aligned}\mathbb{P}[M_{10} \geq 3] &= \mathbb{P}[X_{10} \geq 3] + \mathbb{P}[X_{10} \geq 4] \\ &= 2(\mathbb{P}[X_{10} = 4] + \mathbb{P}[X_{10} = 6] + \mathbb{P}[X_{10} = 8] + \mathbb{P}[X_{10} = 10]) \\ &= 2^{-9}\left[\binom{10}{3} + \binom{10}{2} + \binom{10}{1} + \binom{10}{0}\right] \; \left[= \tfrac{11}{32}\right].\end{aligned}$$

Problem 4.5. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple random walk with $\mathbb{P}[X_1 = 1] = p \in (0,1)$. Define

$$Y_n = \tfrac{1}{n}\sum_{k=1}^{n} X_k, \text{ for } n \in \mathbb{N}.$$

Compute $\mathbb{E}[Y_n]$ and $\operatorname{Var}[Y_n]$, for $n \in \mathbb{N}$.

Hint: You can use the following formulas:

$$\sum_{j=1}^{n} j = \frac{n(n+1)}{2}, \qquad \sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}$$

without proof.

Solution: Let us first represent Y_n in terms of the sequence $\{\xi_n\}_{n\in\mathbb{N}}$:

$$\begin{aligned}Y_n &= \frac{1}{n}\sum_{k=1}^{n} X_k = \frac{1}{n}\Big(\xi_1 + (\xi_1+\xi_2) + \dots + (\xi_1 + \xi_2 + \dots + \xi_n)\Big) \\ &= \frac{1}{n}\sum_{k=1}^{n}(n-k+1)\xi_k\end{aligned}$$

Remembering that $\mathbb{E}[\xi_k] = p - q$ and that $\operatorname{Var}[\xi_k] = 1 - (2p-1)^2 = 4pq$, we have

$$\begin{aligned}\mathbb{E}[Y_n] &= \tfrac{1}{n}\sum_{k=1}^{n}(n-k+1)\mathbb{E}[\xi_j] = \tfrac{p-q}{n}\sum_{k=1}^{n}(n-k+1) \\ &= \tfrac{p-q}{n}\sum_{k=1}^{n} k = (p-q)\frac{n+1}{2}.\end{aligned}$$

By the independence of $\{\xi_n\}_{n\in\mathbb{N}}$ we have

$$\begin{aligned}\operatorname{Var}[Y_n] &= \frac{1}{n^2}\sum_{k=1}^{n}\operatorname{Var}[(n-k+1)\xi_k] = \frac{1}{n^2}\sum_{k=1}^{n}(n-k+1)^2\operatorname{Var}[\xi_k]\\ &= \frac{1}{n^2}\sum_{k=1}^{n}k^2\operatorname{Var}[\xi_k] = \frac{4pq}{n^2}\frac{n(n+1)(2n+1)}{6}\\ &= \frac{2}{3}pq\frac{(n+1)(2n+1)}{n}.\end{aligned}$$

Problem 4.6 (Optional)**.** Let C_n denote the n-th Catalan number (defined in (4.3)).

1. Use the reflection principle to show that C_n is the number of paths $(x_0,\dots,x_{2n}) \in C$ such that $x_k \geq 0$, for all $k \in \{0,1,\dots,2n\}$ and $x_{2n} = 0$.
2. Prove the *Segner's recurrence formula* $C_{n+1} = \sum_{i=0}^{n} C_i C_{n-i}$. *Hint:* Don't compute - just think about paths.
3. Show that C_n is the number of ways that the vertices of a regular $2n$-gon can be paired so that the line segments joining paired vertices do not intersect.
4. Prove that
$$C_n = \binom{2n}{n} - \binom{2n}{n+1},$$
both algebraically (using the formula for the binomial coefficient) and combinatorially (by counting).