Lectures 1–5
Probability Models

Analogy with Geometry: abstract model for chance phenomena

Language and Symbols of Chance Experiments:
Sample space $S$, consisting of all possible outcomes (elements) $e, f, \ldots$,
events (subsets of $S$) $E, F, \ldots$. $E$ occurs when one of its outcomes occurs.
Concepts of set theory with names that reflect the chance experiments.

Examples of different types of sample spaces:
• roll two dice, gambling was the cradle of probability theory
• random point in a circle of radius one around $(0, 0)$,
• all sequences of dice throws that end with a first 6.

Derived Events (Sets):
illustrate with two dice example
union: $E \cup F = \text{all outcomes in } E \text{ or } F$
intersection: $E \cap F = EF = \text{all outcomes in } E \text{ and } F$
empty set: $\emptyset = \text{impossible event}$
$E$ and $F$ are mutually exclusive events when $EF = \emptyset$
subset: $E \subset F$, $E$ implies $F$ or occurrence of $E$ implies occurrence of $F$
complement: $E^c = \text{all outcomes in } S \text{ which are not in } E$. Note $(E^c)^c = E$.
finite (infinite) unions and intersections
Venn diagrams

Algebra of Events:
Note analogy between $\cap$ and $\cup$ and multiplication and addition.
Illustrate with Venn diagrams.
Commutative law: $E \cup F = F \cup E$ and $EF = FE$
Associative law: $(E \cup F) \cup G = E \cup (F \cup G) = E \cup F \cup G$
(Distributive law: $(E \cup F)G = EG \cup FG$
(special case) $(E \cup F)(G \cup F) = EG \cup EF \cup FG \cup FH$
(De Morgan’s Laws: $(\bigcup E_i)^c = \bigcap E_i^c$ and $(\bigcap E_i)^c = \bigcup E_i^c$
for finite or infinite unions/intersections.

Axioms of Probability:
Long run frequency notion (intuitive guidance), qualitative regularity in chaos.
Postulate probabilities $P(E)$ for each event $E \subset S$, satisfying the following (plausible)
axioms:

---

1 Games, Gods and Gambling by F.N. David, Dover 1998
2 See wikipedia under Venn Diagram
3 Kolmogorov (1933), Grundbegriffe der Wahrscheinlichkeitstheorie
Axiom 1: \( 0 \leq P(E) \leq 1 \) for all \( E \subset S \)

Axiom 2: \( P(S) = 1 \), the sure event \( S \)

Axiom 3: \( P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i) \) for any sequence of events \( E_i, i = 1, 2, \ldots \), which are mutually exclusive (\( E_i E_j = \emptyset \) for \( i \neq j \)). **Countable additivity.** It seems that all subsets could be events, but we have a measurability caveat.

Consequences:
\( P(\emptyset) = 0 \) since \( 1 \geq P(\emptyset) = P(\emptyset) + P(\emptyset) + \ldots \)

for any events \( E_i, i = 1, 2, \ldots, n \), which are mutually exclusive we have
\[
P \left( \bigcup_{i=1}^{n} E_i \right) = \sum_{i=1}^{n} P(E_i)
\]

**Finite additivity**, follows simply from Axiom 3, using \( E_i = \emptyset \) for \( i > n \).

\[
P(E^c) = 1 - P(E) \text{ since } E \cup E^c = S
\]

\( E \subset F \Rightarrow P(E) \leq P(F)(= P(E) + P(FF^c)) \)

\( P(E \cup F) = P(E) + P(F) - P(EF) \)

**Boole’s inequality:** \( P(E \cup F) \leq P(E) + P(F) \), useful\(^1\) when \( P(E) + P(F) \approx 0 \).

**Bonferroni’s inequality:** \( P(E \cap F) \geq P(E) + P(F) - 1 \), useful when \( P(E) + P(F) \approx 2 \).

Follows from \( P(EF) = P(E) + P(F) - P(E \cup F) \geq P(E) + P(F) - 1 \)

By induction: \( P \left( \bigcap_{i=1}^{n} E_i \right) \geq \sum_{i=1}^{n} P(E_i) - (n-1) \) and \( P \left( \bigcup_{i=1}^{n} E_i \right) \leq \sum_{i=1}^{n} P(E_i) \)

\[
P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EGF)
\]

The following further generalizes \( P(E \cup F) = P(E) + P(F) - P(EF) \):
\[
P \left( \bigcup_{i=1}^{n} E_i \right) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \ldots + (-1)^{r+1} \sum_{i_1 < i_2 \ldots < i_r} P(E_{i_1} E_{i_2} \ldots E_{i_r})
\]

\[
+ \ldots + (-1)^{n+1} P(E_1 E_2 \ldots E_n)
\]

(1)
where \( \sum_{i_1 < i_2 \ldots < i_r} \) is over all \( \binom{n}{k} \) combinations of \( r \) indices from \( \{1, 2, 3, \ldots, n\} \)

The proof consists of representing the union \( E_1 \cup E_2 \cup \ldots \cup E_n \) as a disjoint union of patches skip of the type \( E_{j_1}^c E_{j_2} \ldots E_{j_k}^c E_{j_{k+1}} \ldots E_{j_N}^c \) with probability denoted by

\[
p = P(E_{j_1} E_{j_2} \ldots E_{j_k} E_{j_{k+1}}^c \ldots E_{j_N}^c)
\]

Such a patch makes one contribution \( p \) to the left side of (1) and
\[
\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \ldots \pm \binom{k}{k} = 1
\]

\(^1\)10\(^{-9}\) target risk per system in aircraft industry. 100 critical systems ⇒ risk \leq 100 \times 10^{-9} = 10^{-7}. This upper bound holds no matter how such system failures might interact.
contribution of \( p \) on the right side of (1). Here the counting of contributions will be explained just for the second term \(-\binom{k}{2}\). The minus comes from the sign of the sum of probabilities \( P(E_{i_1}E_{i_2}) \), and the count \( \binom{k}{2} \) from the fact that we can form that many intersection pair probabilities \( P(E_{i_1}E_{i_2}) \) with events \( E_{i_1}, E_{i_2} \) taken from \( E_{j_1}, E_{j_2}, \ldots, E_{j_k} \) (as long as \( 2 \leq k \)), to which the patch \( E_{j_1}E_{j_2} \ldots E_{j_k}E_{j_{k+1}}^c \ldots E_{j_N}^c \) would contribute probability \( p \), since then \( E_{j_1}E_{j_2} \ldots E_{j_k}E_{j_{k+1}}^c \ldots E_{j_N}^c \subset E_{i_1}E_{i_2} \) for any such choice. Any other choice \( E_{i_1}E_{i_2} \) does not intersect with the patch.

The identity (2) follows from the binomial expansion of \((1-1)^k = 0\). Since these contributions of patches holds for any such patch, the identity (1) follows.

### Equally Likely Outcomes

In finite sample spaces there are often symmetry considerations that suggest that all outcomes be treated equally, i.e. be assigned the same probability.

If \( S = \{1, 2, \ldots, N\} \) then \( P(\{1\}) = \cdots = P(\{N\}) \) & \( P(\{1\}) + \cdots + P(\{N\}) = NP(\{1\}) = 1 \)
\[ \Rightarrow P(i) = 1/N, \quad i = 1, 2, \ldots, N \]
\[ \Rightarrow P(E) = \#(E)/\#(S) = \#(E)/N, \quad \text{where} \quad \#(E) \text{ is the number of outcomes in } E. \]

Thus we need to be able to count the number of elements in an event (set).

#### Quick Counting Interlude (Ch. 1, read on your own)

It helps to have good visuals or symbols to organize your counting. The important thing is to count each possibility for an outcome, and only count it once. First we need to count all outcomes in \( S \) to get \( \#(S) \), and then see how many of them satisfy the restrictions imposed by the description of \( E \).

**No adjacent defective rocket motors:** Six rocket motors, arranged in a circular pattern, are used to separate two stages in space. Engineering analysis determined that up to two rocket motors can be allowed to fail, as long as they are not adjacent. Suppose two motors are defective, what is the chance of successful separation, given that the six motors are randomly arranged in the circular pattern.

How many circular arrangements with 2 defects, and 4 nondefects? Think of the motors as distinct items being placed in positions 1, 2, \ldots, 6, with 1 and 6 considered adjacent. The first defective motor \( D_1 \) can wind up in any of 6 possible positions. The second defective motor \( D_2 \) can be in any of the remaining 5 positions. Similarly, the nondefective or good motors \( G_1, \ldots, G_4 \) have 4, 3, 2, 1 possible choices, respectively. Combining all the choices with each other gives us \( 6! = “6! factorial“ = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720 = \#(S) \) possible arrangements of 6 motors in a circle. This is the number of permutations of 6 distinct items in some sequential order. For example, \( 3! = 3 \times 2 \times 1 = 6 \) counts the number of permutations of 1, 2, 3 in all orders: \((1,2,3) \) (2,1,3) (3,1,2) \((1,3,2) \) (2,3,1) (3,2,1) \( \)

This is a special illustration of the more general multiplication rule: If we have \( n_1 \) possibilities to make one choice, then \( n_2 \) possibilities to make a second or subsequent choice (the possibilities may depend on the first choice, but \( n_2 \) does not), then \( n_3 \) possibilities to make a third or subsequent choice (the possibilities may depend on the first two choices, but \( n_3 \) does not), \ldots, and finally \( n_k \) possibilities to make a \( k^{th} \) or subsequent and last choice
(the possibilities for this may depend on the first \( k - 1 \) choices, but \( n_k \) does not), then the total number of \( k \) combined choices (choice_1, \ldots, choice_k) that can be made is \( n_1 \times \ldots \times n_k \).

Now back to the rocket motors: how many of these 720 arrangements of motors have no defects next to each other? \( D_1 \) had 6 choices, say it goes into position 1. Then \( D_2 \) should avoid positions 2 and 6, thus has 3 choices to stay clear, say it goes into position 5. The good motors now have 4! = 4 \times 3 \times 2 \times 1 = 24 choices to be placed into positions 2,3,4,6.

Thus in total we have \( 6 \times 3 \times 24 = \#(E) \) combined choices avoiding adjacent defects. Thus \( P(E) = \#(E)/\#(S) = 6 \times 3 \times 24/(6 \times 5 \times 24) = 3/5 \).

When you deal with ratios of products, it is best to cancel common factors first! You avoid the work of laborious and superfluous multiplications.

By good visuals or symbols, think of a way to represent a typical outcome: 

\[ (G_1, D_2, G_4, G_2, G_3, D_1) \]

expressing the fact that \( G_1, G_2, G_3, G_4 \) go into positions 1, 4, 5, 3 and \( D_1, D_2 \) go into 6 and 2. Writing down such example cases helps you organize your thinking about the problem. You typically don’t want to write down all choice combinations! The numbers get big in a hurry. That’s why you need to learn how to count efficiently. End

License Plates: How many license plates are possible with the new State of Washington scheme (3 letter followed by 4 digits)? \( 26 \times 26 \times 26 \times 10 \times 10 \times 10 \times 10 = 175,760,000 \) (Multiplication rule). That should last for a while, even after removing some taboos.

Counting Words: What if we have items that are not all distinguishable? For example, in a “word” like MISSISSIPPI we can permute same letters among each other and the word still “reads” the same. How many unique words can we form from the same set of 11 letters MISSISSIPPI? If \( x \) denotes that number of distinct words, then we can get all \( 11! \) permutations of the 11 letters by traversing through all \( x \) distinguishable word choices and for each such word choice make all \( 4! \times 1! \times 2! \times 4! \) choices of permuting same letters among each other. Thus \( 11! = x \times 4! \times 1! \times 2! \times 4! \) or \( x = 11!/(4! \times 1! \times 2! \times 4!) = 34650 \). Of course most of those words have no meaning. While meaning matters in Scrabble™, there are plenty of applications where the counting of “words” is relevant. Say, we want to see how many ways 4 placebos, 4 pills each of type A, B, C can be assigned to 16 patients. Each such assignment is a “word.” We don’t care which of the 4 patients getting pill \( B \), gets which of the 4 \( B \) pills. Thus there are \( 16!/(4!)^4 = 63,063,000 \) such distinguishable pill assignments.

Combinations: How many ways can we pick \( k \) out of \( N \) patients to receive treatment \( A \), while the rest gets treatment \( B \)? This is just a special case of the previous word count. The answer is:

\[
\binom{N}{k} = \frac{N!}{k!(N-k)!} = \frac{N(N-1) \cdots (N-k+1)(N-k)!}{k!(N-k)!} = \frac{N(N-1) \cdots (N-k+1)}{k!}
\]

The symbol \( \binom{N}{k} \) is pronounced “\( N \) choose \( k \).” We think of this number as the number of grabs of \( k \) items from \( N \) without regard to order. Any such grab can be represented as a set of \( k \) distinct indices \( \{i_1, \ldots, i_k\} \) with \( i_j \) representing distinct integers taken from 1, 2, \ldots, \( N \). We can view the \( i_j \) in \( \{i_1, \ldots, i_k\} \) as ordered, i.e., \( i_1 < \ldots < i_k \).

If we select \( k \) items in sequence from \( N \) and the order thus matters, i.e., we represent such a selection as a \( k \)-tuple \( (i_1, \ldots, i_k) \), there are \( N(N-1) \cdots (N-k+1) \) such \( k \)-tuples, with distinct elements \( i_j \) taken from 1, 2, \ldots, \( N \). The \( i_j \) in these \( k \)-tuples are not necessarily ordered. This is a special case of the multiplication rule, where subsequent choice possibilities are affected by previous choices, but the number of any of the choices is not.
The Binomial Theorem: For integer $n \geq 0$ and real $x$ and $y$ we have

$$(x + y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

Proof: Each of the $n$ factors $(x+y)$ contributes either an $x$ or a $y$ when multiplying them out. Say there are $i$ such $x$’s and $n-i$ such $y$’s, then they could arise in $\binom{n}{i}$ such ways, by choosing the $i$ terms $(x+y)$ which contribute an $x$, while the others contribute a $y$. Since $i$ can take on any one of the values 0, 1, \ldots, $n$ and since these choices are mutually exclusive we get the above formula.

The Multinomial Theorem: For integer $n \geq 0$ and real $x_1, \ldots, x_r$ we have

$$(x_1 + \ldots + x_r)^n = \sum_{(n_1, n_2, \ldots, n_r) \in \mathbb{Z}^r} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

Proof: Each of the $n$ factors $(x_1 + \ldots + x_r)$ contributes some $x_i$, say $n_i$ of them, when multiplying them out. Say there are $n_i$ such $x_i$’s for $i = 1, \ldots, r$, then they could arise in $\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! \cdots n_r!}$ such ways, by choosing, for $i = 1, \ldots, r$, the $n_i$ terms $(x_1 + \ldots + x_r)$ which contribute an $x_i$. Note that the multinomial coefficient $\binom{n}{n_1, n_2, \ldots, n_r}$ just represents the number of words that can be made from $n_i$ of $x_i$, for $i = 1, \ldots, r$. The words indicate which terms $(x_1 + \ldots + x_r)$ contribute each respective “letter” $x_i$. Since the $n_i$ are integers adding to $n$ and since these choices are mutually exclusive and exhaustive we get the above formula.

Sums of $r$ Positive Integers with Same Total: How many ways to write $n = x_1 + \ldots + x_r$ with integer $x_i > 0$? Between the $n$ units there are $n - 1$ gaps where we can place our $r - 1$ plus signs. Answer $\binom{n-1}{r-1}$

Sums of $r$ Nonnegative Integers with Same Total: How many ways to write $n = x_1 + \ldots + x_r$ with integer $x_i \geq 0$? Think of $n$ units and the apportioning $r - 1$ plus signs as $n + r - 1$ items from which we have to select the $r - 1$ items that should be plus signs. Answer $\binom{n+r-1}{r-1}$

Equally Likely Outcomes (continued)

Example 1 (Two Fair Dice): improvise

Example 2 (Bowl Draw): Two balls drawn randomly without replacement from a bowl with six white and five black balls. If $E$ = ”one white and one black ball is drawn” what is $P(E)$?

Two solution paths, ordered and unordered drawings. Answer: 6/11
**Example 3 (Poker Hands):** A hand consists of 5 cards taken from a deck of 52, consisting of 4 suits, each having face values: ace, 2, . . . , 10, jack, queen, king.

Straight: five in sequence not of same suit = $E_1$

Straight flush: five in sequence from same suit = $E_2$

Full house: three of one face value and two of another = $E_3$

Compute the probabilities $P(E_i)$

Answers: \(\binom{52}{5} = 2,598,960\) possible hands

\[
P(E_1) = \frac{(4^5 - 4) \cdot 10}{\binom{52}{5}} = .0039, \quad P(E_2) = \frac{4 \cdot 10}{\binom{52}{5}} = .000015, \quad P(E_3) = \frac{13 \cdot 12 \cdot \binom{4}{3} \cdot \binom{4}{2}}{\binom{52}{5}} = .0014
\]

**Example 4 (Treatment Randomization):**

In a randomized trial of a new drug, 5 patients out of 10 are randomly assigned to receive the new drug while the other 5 get a placebo (look alike drug). If the new drug acts just like the placebo, we could treat the measurable outcome for the 10 patients as being preordained, i.e., have nothing to do with the treatment assignment. What then is the chance that the treatment group will contain at least four of the five highest scores\(^1\)?

Assume all scores are distinct and can be ranked 1, 2, . . . , 10, from lowest to highest.

There are \(\binom{10}{5} = 252\) choices of 5 treatment scores from the 10 available scores. Either the top 5 are chosen (one way), or exactly 4 of the top 5 are chosen. Thus the total number of ways to get at least 4 of the top 5 scores is

\[
1 + \binom{5}{4} \binom{5}{1} = 26 \quad \text{with probability} \quad \frac{26}{252} = 0.1032
\]

However, if we judge the drug performance by looking at the sum of the 5 score ranks for the treatment group, what is the chance of getting at least the third highest possible sum? The highest sum is obtained with \(10 + 9 + 8 + 7 + 6 = 40\). The second highest sum is \(10 + 9 + 8 + 7 + 5 = 39\) and the third highest sum is obtained in two possible ways as \(10 + 9 + 8 + 7 + 4 = 10 + 9 + 8 + 6 + 5 = 38\). Thus the desired chance is \((1 + 1 + 2)/252 = 0.0159\), which might induce you to reconsider your position that the drug has nothing to do with high scores. This is an example of the Wilcoxon test procedure.

**Example 5 (Lottery):**

Among \(N\) tickets in a lottery there is a unique special prize. \(k\) tickets are sold in random order one by one. What is the chance that the prize is among them?

All \(N(N - 1) \cdots (N - k + 1)\) draws are equally likely. The prize could be part of these \(k\) tickets if it was the first or second, . . . , or \(k\)th ticket drawn, each one of these could be combined with \((N - 1)(N - 2) \cdots (N - k + 1)\) choices for the remaining \(k - 1\) tickets to be drawn. Thus the desired probability is

\[
\frac{(N - 1)(N - 2) \cdots (N - k + 1)}{N(N - 1) \cdots (N - k + 1)} + \cdots + \frac{(N - 1)(N - 2) \cdots (N - k + 1)}{N(N - 1) \cdots (N - k + 1)} = \frac{k}{N}
\]

The same effect could be had if the number of the prize ticket is randomly chosen after the \(k\) tickets are sold, as in a normal lottery. Then the answer \(k/N\) is more directly evident.

---

\(^1\)different from containing at least the 4 highest scores, chance 6/252 = 0.02381.
Example 6 (Birthday Problem):

$E = \text{all } n \text{ persons in a room have distinct birthdays.}$

e = (x_1, \ldots, x_n), \ x_i \in \{1, 2, \ldots, 365\}

There are $365^n$ such possible outcomes $e$, ignoring leap years. Assume all outcomes are equally likely.

$$P(E) = \frac{(365)_n}{365^n} = \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n} = 1 \cdot \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right)$$

which becomes .524305, .492703, .029626, .000000307 for $n = 22$, $n = 23$, $n = 50$ and $n = 100$ respectively. It can be be shown that equally likely birthdays is least favorable for having a birthday match.

Birthday Problem Extension: The probability of the event $A^c$ of getting at least one matching or neighboring pair of birthdays is:

$$P_n = P(A^c) = 1 - P(A) = 1 - \frac{365(n-1)!}{365^n} \binom{365-n-1}{n-1}$$

$$= 1 - \frac{(365-2n+1)(365-2n+2)\cdots(365-2n+n-1)}{365^{n-1}}.$$ 

Hence $P_{13} = .4829$, $P_{14} = .5375$, $P_{23} = .8879$, and $P_{40} = 0.9991$. The expression for $P(A)$, the probability of no equal or neighboring birthdays derives from:

1) 365 choices of person 1’s birthday. 2) There are $365 - n$ non-birthdays (NB), two of which flank person 1. Into the $365 - n - 1$ remaining places between NB’s we need to place the remaining $n - 1$ birthdays (B). 3) This gives us an NB-B pattern with no two B’s next to each other. 4) There are $(n-1)!$ ways of assigning the $n - 1$ B’s to specific persons.

Lec4 ends

Example 7 (Matching Problem, Present Exchange):
\(N\) distinct items, \(1, 2, \ldots, N\), are arranged in random order. A match occurs if the position number coincides with the item number. \(F_N = \text{event that no match occurs}\). \(P(F_N) = ?\)

Solution: Let \(E_i = \text{event that item } i \text{ is in the } i^{th}\text{ position, i.e. gives us a match.} \) For distinct \(i_1, i_2, \ldots, i_k\) we have

\[
P(E_{i_1}E_{i_2}\cdots E_{i_k}) = \frac{(N-k)!}{N!}
\]

so that

\[
P(F_N) = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{N-k} \frac{1}{(N-k)!}
\]

which for large \(N\) becomes \(\approx e^{-1/k!}, k = 0, 1, 2, \ldots\) (Poisson Distribution).
Example 8 (Monte Hall Cups): Monte Hall\(^1\) lets you choose one of three doors, one of which hides a fancy car, the others just a toy goat. You get to keep what is behind your chosen door. Just as you reach for your door handle, he says: wait, I will open one of the other doors, with a goat behind it. He swears that he would always give you this intervention information, regardless of what you chose. What should you do, and what are the best chances for your choice?

Originally we had chance \(\frac{1}{3}\) of getting the car, assuming that we picked our door randomly. With the given information we could improve on that to \(\frac{1}{2}\) by picking one of the two remaining doors at random. An even better choice is to switch to the other remaining door, relying on the \(\frac{2}{3}\) chance of having missed the door with the car on our original random choice. By switching we would get the car with probability \(\frac{2}{3}\). The power of \(P(E^c) = 1 - P(E)\).

Example 9 (Axiom 3): Illustrate Axiom 3 through the event \(E\) of getting a 6 before a 5 when you roll a die until you get a 5 or 6. How to assign probabilities to all possible outcomes? Any sequence \((x_1, \ldots, x_n)\) with \(x_i \in \{1, 2, \ldots, 6\}\) should get probability \(\frac{1}{6^n}\).

\[
E = \{(6), \{(x_1, 6), x_1 = 1, \ldots, 4\}, \{(x_1, x_2, 6), x_1, x_2 = 1, \ldots, 4\}, \ldots\}
\]

\[
P(E) = \frac{1}{6} + \frac{4}{6^2} + \frac{4^2}{6^3} + \frac{4^3}{6^4} + \ldots = \frac{1}{4} \left( \sum_{i=0}^{\infty} \left(\frac{4}{6}\right)^i - 1 \right) = \frac{1}{4} \left( \frac{1}{1 - 2/3} - 1 \right) = \frac{1}{2}
\]

Example 10 (Craps): In the game of craps (Ch. 2, Problem 26) you win outright if you roll a sum of 7 or 11 with two dice, you lose, if you roll a sum of 2, 3, or 12, and if you rolled a sum \(i = 4, 5, 6, 8, 9, 10\), then you win when you roll the same sum again before a 7 shows up. Thus one of the ways to win is to get a 4 and then another 4 before a 7 shows up. What is the probability of winning the game that way? The outcomes of this type consist of

\[
E = \{(4, *, \ldots, *, 4), \,* \neq 4 \text{ or } 7, \, n = 0, 1, 2, \ldots\}
\]

Instead of

\[
(4, *, \ldots, *, 4), \,* \neq 4 \text{ or } 7
\]

I should really write more precisely

\[
(x_1, y_1) : x_1 + y_1 = 4, (x_2, y_2) : x_2 + y_2 \neq 4 \text{ or } 7, \ldots, (x_{n+1}, y_{n+1}) : x_{n+1} + y_{n+1} \neq 4 \text{ or } 7, (x_{n+2}, y_{n+2}) : x_{n+2} + y_{n+2} = 4
\]

but as you can see, that notation is a bit cumbersome. Thus I will continue with the original telegraphic notation, with the understanding that in counting outcomes I count the number of pairs \((x_i, y_i)\) with the given sum properties. I hope this makes it all clearer.

There are \(36 - 9 = 27\) sums with \(* \neq 4 \text{ or } 7\). For fixed \(n\), the probability of all the \((n+2)\)-tuples in \(E\) is

\[
P(\{(4, *, \ldots, *, 4), \,* \neq 4 \text{ or } 7\}) = \frac{3 \cdot 27 \cdot \ldots \cdot 27 \cdot 3}{36^{n+2}}
\]

\(^1\)Jason Rosenhouse (2009), The Monte Hall Problem, The Remarkable Story of Math’s Most Contentious Brain Teaser, Oxford University Press.

Lec5 ends
\[ P(E) = \frac{9}{36^2} \cdot \sum_{n=0}^{\infty} \left( \frac{27}{36} \right)^n = \frac{9}{36^2} \cdot \frac{1}{1 - \frac{27}{36}} = \frac{9}{36^2} \cdot \frac{36}{9} = \frac{1}{36} \]

Similarly you can work out the probabilities of all other ways of winning at craps this way, where you have to get an \( i \) and then another \( i \) before a 7 shows up, for \( i = 4, 5, 6, 8, 9, 10 \). Of course, you need to combine this with getting a 7 or 11 on the first roll and thus winning with the first roll outright in order to get the probability of winning at craps.