University of Mannheim
School of Social Sciences
Mathematics for Political Scientists, Fall 2022
Carlos Gueiros

Solutions Problem Set 6
Probability Theory I

1. Suppose we draw 5 balls from an urn containing 15. How many different sets of drawn balls are there? Consider permutation and combination both with and without repetition.

   Ordered, with repetition: $n^k = 15^5 = 759,375$
   Ordered, without repetition: $\binom{n}{k}k! = \frac{n!}{k!(n-k)!}k! = \frac{15!}{10!} = 360,360$
   Unordered, with repetition: $\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!} = 11,628$
   Unordered, without repetition: $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{15!}{5!(15-5)!} = 3,003$

2. Make a complete list of all the different subsets of the set $\{a, b, c\}$. How many are there if the empty set and the set itself are included? Do the same for the set $\{a, b, c, d\}$. In addition: How many different subsets are there for a set containing n elements?

   Subsets of $\{a, b, c\}$: $\{a\}$, $\{b\}$, $\{c\}$, $\{ab\}$, $\{ac\}$, $\{bc\}$, $\{abc\}$, $\emptyset$, that is 8 subsets.
   Subsets of $\{a, b, c, d\}$: $\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$, $\{ab\}$, $\{ac\}$, $\{ad\}$, $\{bc\}$, $\{bd\}$, $\{cd\}$, $\{abc\}$, $\{abd\}$, $\{acd\}$, $\{bcd\}$, $\{abcd\}$, $\emptyset$, which is 16 subsets overall.

   Let us think of the possible subsets of size k of a set of size n. We can calculate their number as a combination without repetition. For example, if we are interested in a subset with 4 elements of a set with 10 elements, there are $\binom{10}{4}$ possible combinations. The number of all possible subsets of a set containing n elements is, thus, given by $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

3. The World Health Organization (WHO) recently detected a new, deadly disease. This so called "CAT-Virus" is currently making its way to Europe and the US. The first symptoms are hallucination and dizziness. Quickly followed by pain, fever and shock. Death might follow within 2-4 days. The WHO developed a quick test that everyone can self-administer. The test is 99% accurate. That is, if you have the disease, there is a 99 percent chance that the test will detect it. If you don't have the disease, the test will be 99 percent accurate in saying that you don't. In the

general population, 0.1% (that is a tenth of a percentage point) of the people have the disease. When you get a positive test result, what is the probability that you actually have the CAT-Virus?

$$\begin{aligned} P(A) &= 0.01/100 \\ P(A^c) &= 1 - 0.01/100 \\ P(B|A) &= 0.99 \\ P(B|A^c) &= 1 - 0.99 \end{aligned} \tag{1}$$

Bayes Theorem says:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2}$$

Law of Total Probability says:

$$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c) \tag{3}$$

Substituting and computing:

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} \\ &= \frac{0.01/100 \cdot 0.99}{0.01/100 \cdot 0.99 + (1 - 0.01/100) \cdot (1 - 0.99)} \\ &= 0.09 \end{aligned} \tag{4}$$

4. Three steps to Jesus is a semi-popular teenager game in the USA. Given a random article from Wikipedia you have to get to the article from Jesus Christ in 3 steps at most. When you don't you loose. When there are 5,650 pages that link to the page of Jesus Christ and the total number of Wikipedia articles is 3,465,930, what is the probability of winning this game?

   Define p as the probability to get to the Jesus article as $p = 5650/3465930 = 0.0016$. The probability to land directly on the Jesus article is simply a. The probability to get there after the first step, $(1-a) \cdot a$, in the second step $(1-a)(1-a) \cdot a$ and in the third step, $(1-a)(1-a)(1-a) \cdot a$. You can either add these numbers up, or simply take the complementary probability $1 - ((1-a)^4) = 0.0064$. Comments: I made a number of assumptions here. Among them: The probability to get to a page is equal across all Wikipedia pages (consequential) and that one might go back to the page one came from (inconsequential). In essence, I assumed that someone plays this game without any though on where to click. In that sense, it is a worst-case estimate.