Seminar Nr.3, Probabilistic Models

Theory Review

Binomial Model: The probability of k successes in n Bernoulli trials, with probability of success p ($q = 1 - p$), is

$P(n,k) = C_n^k p^k q^{n-k},\ k = \overline{0,n}.$

Hypergeometric Model: The probability that in n trials, we get k successes out of n_1 and $n-k$ failures out of $N - n_1$ ($0 \le k \le n_1$, $0 \le n-k \le N-n_1$), is

$$P(n;k) = \frac{C_{n_1}^k C_{N-n_1}^{n-k}}{C_N^n}.$$

Poisson Model: The probability of k successes ($0 \le k \le n$) in n trials, with probability of success p_i in the i^{th} trial ($q_i = 1 - p_i$), $i = \overline{1,n}$, is

$$P(n;k) = \sum_{1 \le i_1 < \ldots < i_k \le n} p_{i_1} \ldots p_{i_k} q_{i_{k+1}} \ldots q_{i_n}, \qquad i_{k+1}, \ldots, i_n \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$$

$$= \text{the coefficient of } x^k \text{ in the polynomial expansion } (p_1 x + q_1)(p_2 x + q_2) \ldots (p_n x + q_n).$$

Pascal (Negative Binomial) Model: The probability of the n^{th} success occurring after k failures in a sequence of Bernoulli trials with probability of success p ($q = 1 - p$), is

$P(n;k) = C_{n+k-1}^{n-1} p^n q^k = C_{n+k-1}^k p^n q^k.$

Geometric Model: The probability of the 1^{st} success occurring after k failures in a sequence of Bernoulli trials with probability of success p ($q = 1 - p$), is

$p_k = pq^k.$

---

1. Five percent of computer parts produced by a certain supplier are defective. What is the probability that a sample of 16 parts contains
a) exactly 3 defective parts (ev. A)?
b) more than 3 defective parts? (ev. B)?
c) at least one defective part (ev. C)?
d) less than 3 defective parts (ev. D)?

Solution:
A *trial* here is "selecting a computer part". There are 16 independent trials.
At each trial, the outcomes are "the computer part is defective" or "the computer part is good".
We are asked about a certain number of parts being *defective*, so we define "success": "a part is defective". Probability of success in every trial is then 0.05.
So, this is a **Binomial model** with parameters $n = 16$, $p = 0.05$.

a)
$$P(A) = P(k=3) = C_{16}^3 (0.05)^3 (0.95)^{13} = 0.0359.$$

b)
$$\begin{aligned} P(B) &= P(k>3) = 1 - P(k \le 3) = 1 - P\Big(\underbrace{(k=0) \cup (k=1) \cup (k=2) \cup (k=3)}_{\text{m.e.}}\Big) \\ &= 1 - \sum_{k=0}^{3} C_{16}^k (0.05)^k (0.95)^{16-k} = 0.007. \end{aligned}$$

c)
$$P(C) = P(k \ge 1) = 1 - P(k=0) = 1 - C_{16}^0 (0.05)^0 (0.95)^{16} = 0.5599.$$

d)

$$P(D) = P(k < 3) = 1 - P(k \geq 3) = 1 - P(\underbrace{A \cup B}_{\text{m.e.}}) = 1 - P(A) - P(B) = 0.9571.$$

2. There are 200 seats in a theater, 10 of which are reserved for the press. 150 people come to the show one night, and are seated randomly. What is the probability of all the seats reserved for the press to be occupied (ev. A)?

Solution:
A trial here is "taking a seat". There are 150 trials.
At each trial, the outcomes are "the seat is reserved for the press" or "the seat is *not* reserved for the press". There are 200 seats to choose from, 10 of which have the property of being reserved for the press and sampling is done **without** replacement.
Thus, this is a **Hypergeometric model** with parameters $N = 200, n_1 = 10, n = 150$.
Then the probability that among the 150 objects, 10 have the characteristic and 140 do not, is

$$P(A) = P(k = 10) = \frac{C_{10}^{10} \cdot C_{190}^{140}}{C_{200}^{150}} = 0.0521.$$

3. Among 10 laptop computers, seven are good, the rest have defects. Unaware of this, a customer buys 5 laptops.
a) What is the probability of exactly 2 defective ones among them (ev. A)?
b) Knowing that *at least* 2 purchased laptops are defective, what is the probability that *exactly* 2 are defective (ev. B)?

Solution:
a) There are 10 objects to choose from, 3 of which are defective and 5 are being selected, obviously *without* replacement. This is a **Hypergeometric model** with $N = 10$, $n_1 = 3$ (we are interested in the defective ones) and $n = 5$.

$$P(A) = P(k = 2) = \frac{C_3^2 \cdot C_7^3}{C_{10}^5} = 0.4167.$$

b) First off, this is *conditional* probability

$$P(B) = P(k = 2 | k \geq 2) = \frac{P(k = 2)}{P(k \geq 2)} = \frac{P(k = 2)}{P(k = 2) + P(k = 3)} = \frac{P(A)}{P(A) + P(k = 3)}.$$

For each probability, we use the same model with the same parameters.

$$P(B) = \frac{P(A)}{P(A) + \dfrac{C_3^3 \cdot C_7^2}{C_{10}^5}} = 0.8333.$$

4. A computer program is tested by 5 independent tests. If there is an error, these tests will detect it with probabilities 0.1, 0.2, 0.3, 0.4 and 0.5, respectively. Suppose that the program contains an error. What is the probability that it will be found by
a) at least one test (ev. A)?
b) more than two tests (ev. B)?
c) all five tests (ev. C)?

Solution:
A trial here is "a test is checking the computer program". There are 5 independent trials.
At each trial, the outcomes are "the test finds the error" or "it does not". Denote "success": "the test finds the error". What is different here from Problem 1. is that the probability of

success *differs* in every trial. Thus, this is a **Poisson model** with parameters $n = 5, p_1 = 0.1$, $p_2 = 0.2$, $p_3 = 0.3$, $p_4 = 0.4$, $p_5 = 0.5$. So we write the polynomial

$$\begin{aligned} & (0.1x + 0.9)(0.2x + 0.8)(0.3x + 0.7)(0.4x + 0.6)(0.5x + 0.5) \\ = \; & 0.0012x^5 + 0.0214x^4 + 0.1274x^3 + 0.3274x^2 + 0.3714x + 0.1512. \end{aligned}$$

Then we have
a)

$$P(A) = P(k \geq 1) = 1 - P(k = 0) = 1 - 0.1512 = 0.8488.$$

b)

$$P(B) = P(k > 2) = P(k = 3) + P(k = 4) + P(k = 5) = 0.1274 + 0.0214 + 0.0012 = 0.15.$$

c)

$$P(C) = P(k = 5) = 0.0012.$$

5. In a public library, 1 out of 10 people using the computers do not close Windows properly. What is the probability that Windows is closed properly only by the 3^{rd} user (event A)?

Solution:
A trial: a user closes his Windows session. How many trials are there? We *don't* know, theoretically, infinitely many! At each trial, the user either closes Windows correctly or not. Define "success" as "the user closes Windows properly", so that this matches a Geometric model. Then the probability of success is $p = 9/10 = 0.9$ and event A may be rephrased as: the first success in the 3^{rd} trial, i.e. after 2 failures. Now this is a **Geometric model** with parameter $p = 0.9$ and we compute

$$P(A) = P(k = 2) = 0.9(0.1)^2 = 0.009.$$

6. An engineer tests the quality of produced computers. Suppose that 5% of computers have defects and defects occur independently of each other. Find the probability
a) of exactly 3 defective computers in a shipment of 20 (ev. A);
b) that the engineer has to test at least 5 computers in order to find 2 defective ones (ev. B).

Solution:
a) This is a **Binomial model**, with "trial": engineer tests a computer, "success": the computer has defects, parameters $n = 20$, $p = 0.05$.

$$P(A) = P(k = 3) = C_{20}^{3}(0.05)^3(0.95)^{17} = 0.0596.$$

b) First off, note that the "in a shipment of 20" part refers to part a) *only*! In part b), a trial means the same thing, but there's no longer a given, finite number of trials. The experiment is repeated until something happens, namely, until the engineer finds 2 defective computers. This is about the rank of a success, so, again we define "success"as: the computer is defective. We rephrase event B as
B: the 2^{nd} success in 5 or more trials, i.e.
B: the 2^{nd} success after 3 or more failures.

This looks like a Pascal model, but there is a problem: the "3 *or more* failures" part, how many exactly are we talking about? A number greater than or equal to 3, i.e. $k = 3, 4, \ldots$. So that probability would be an infinite sum, a *series*. How do we deal with that? By looking at the *complementary* event,
$\overline{B}$: the 2^{nd} success after 2 or less failures;
So, for $\overline{B}$, we use the **Negative Binomial model** with parameters $n = 2, p = 0.05$ and we consider the values $k \in \{0, 1, 2\}$. Thus,

$$\begin{aligned} P(B) &= P(k \geq 3) = 1 - P(\overline{B}) = 1 - P(k \leq 2) \\ &= 1 - \sum_{k=0}^{2} C_{k+1}^{1}(0.05)^2(0.95)^k = 0.986. \end{aligned}$$

7. (Banach's Problem). A person buys 2 boxes of aspirin, each containing n pills. He takes one aspirin at a time, randomly from one of the two boxes. After a while, he realizes that one box is empty.

a) Find the probability of event A: when he notices that one box is empty, there are k ($k \leq n$) pills left in the other box.

b) Use part a) to find a formula for $S_n = C_{2n}^n + 2\cdot\, C_{2n-1}^n + \ldots + 2^n \cdot C_n^n$.

Solution:

a) Denote the events:

A_1: at the time he realizes that the 1st box is empty, there are k pills left in the 2nd box, and

A_2: the other way around.

Then A_1 and A_2 are disjoint and by symmetry, $P(A_1) = P(A_2)$. So,

$$P(A) = P(A_1 \cup A_2) = 2P(A_1).$$

Now, let us see, *when exactly* does he notice that the 1^{st} box is empty? It's not when he *takes* the last pill, but when he *attempts* to take another pill and there's none, i.e. when he attempts to take the $(n+1)^{st}$ pill out of it. This sounds like the rank of a success... Indeed, a trial is "he attempts to take a pill (and if he finds one, he will take it)" so, let us define "success" to be: the person opens the 1^{st} box (attempting to take a pill out of it), "failure": the 2^{nd} box. Event A_1 is then: obtain the $(n+1)^{\text{st}}$ success after $n-k$ failures.

This is then a **Negative Binomial model** with parameters $N = n+1$ and $p = P(\text{success}) = P(\text{failure}) = q = 1/2$ and we want to compute the probability for $K = n-k$. Thus,

$$P(A_1) = C_{(n+1)+(n-k)-1}^{n+1-1} \left(\frac{1}{2}\right)^{n+1} \left(\frac{1}{2}\right)^{n-k} = \frac{1}{2} C_{2n-k}^n \cdot \left(\frac{1}{2}\right)^{2n-k}$$

and

$$P(A) = C_{2n-k}^n \left(\frac{1}{2}\right)^{2n-k}.$$

b) If for $k = \overline{0,n}$, we denote by E_k the event A from a), then $\{E_k\}_{k=\overline{0,n}}$ form a partition, so $\sum_{k=0}^{n} P(E_k) = 1$, i.e. by a),

$$\begin{aligned} 1 &= \sum_{k=0}^{n} C_{2n-k}^n \cdot \frac{2^k}{2^{2n}} \\ &= \frac{1}{2^{2n}} \sum_{k=0}^{n} 2^k C_{2n-k}^n \\ &= \frac{1}{2^{2n}} \cdot S_n, \end{aligned}$$

so

$$S_n = 2^{2n}.$$