Lecture 3

2. Hypergeometric Model

This is the version of the Binomial model, without replacement. That will make a *great* difference, not only in the computational formulas, but in the parameters of the model.

Model: There are N $(N \in \mathbb{N})$ objects, n_1 $(n_1 \leq N)$ of which have a certain trait (we could call that "success"). A number of n $(n \leq N)$ objects are selected, one at a time, **without** replacement. Find the probability $P(n; k)$ of exactly k $(0 \leq k \leq n)$ of the n objects selected, having that trait (i.e. k successes).

In the other setup, the model could be described as: There are N $(N \in \mathbb{N})$ balls in a box, n_1 $(n_1 \leq N)$ of which are white, the rest of them $(N - n_1)$ black. A number of n $(n \leq N)$ balls are extracted, one at a time, without putting them back. Find the probability $P(n; k)$ of exactly k $(0 \leq k \leq n)$ white balls being selected.

Remark 2.1.

The parameters in a Hypergeometric model are N (total number of objects), n_1 (number of objects with a certain property) and n (number of trials). Again, k is not a parameter of the model.

Proposition 2.2.

The probability $P(n;k)$ in a Hypergeometric model is given by

$$P(n;k) = \frac{C_{n_1}^{k}\, C_{N-n_1}^{n-k}}{C_N^n}, \quad k = 0, 1, \ldots, n. \tag{2.1}$$

Remark 2.3.

1. Intuitively, the probability $P(n;k)$ in (2.1) can be computed using the classical definition of probability:
The total number of possible outcomes for the experiment is C_N^n.
There are $C_{n_1}^k$ ways of choosing the k objects from the first category and $C_{N-n_1}^{n-k}$ ways of choosing the remaining $n-k$ objects from the rest (without replacement), and the two actions are independent of each other, so the number of favorable outcomes is $C_{n_1}^k C_{N-n_1}^{n-k}$.

2. As before,

$$\sum_{k=0}^{n} P(n;k) = 1, \text{ i.e. } \sum_{k=0}^{n} C_{n_1}^k C_{N-n_1}^{n-k} = C_N^n.$$

Example 2.4.

There are 15 boys and 20 girls in a probability class. Ten people are selected for a certain project. Find the probability that the group contains
a) an equal number of boys and girls (event A),
b) at least one girl (event B).

Solution. This is a Hypergeometric model with $N = 35$ and $n = 10$. If we choose "success" to mean "selecting a girl" (case I), then $n_1 = 20$, otherwise ("success" = "choosing a boy", case II) , $n_1 = 15$.

a) For event A, an equal number of boys and girls out of 10 people, means 5 boys and 5 girls. Therefore,
In case I,

$$P(A) = P(10; 5) = \frac{C_{20}^{5} C_{15}^{5}}{C_{35}^{10}} \approx 0.2536\,.$$

In case II,

$$P(A) = P(10; 5) = \frac{C_{15}^{5} C_{20}^{5}}{C_{35}^{10}} \approx 0.2536\,.$$

b) For event B, since the question is about the number of girls being selected, it is easier to go with case I.
"At least one girl" means the number of girls could be 1 or 2 or ... or 10. Let us look at the complementary event, which would be "at most 0 girls", or "0 girls". There are fewer numbers to consider, so it is easier to compute the probability of the contrary event. Thus,

$$P(B) = 1 - P\left(\overline{B}\right) = 1 - P(10; 0) = 1 - \frac{C_{20}^{0} C_{15}^{10}}{C_{35}^{10}} = 1 - \frac{C_{15}^{10}}{C_{35}^{10}} \approx 0.9999.$$

If we consider case II, the event would be "at most 9 boys" and again it is easier to compute the probability of the contrary event, i.e. "at least 10 boys", which means "10 boys". So,

$$P(B) = 1 - P\left(\overline{B}\right) = 1 - P(10; 10) = 1 - \frac{C_{15}^{10} C_{20}^{0}}{C_{35}^{10}} \approx 0.9999.$$

Note: whichever we consider as "success", of course the result should be the same.

3. Poisson Model

This model is a generalization of the Binomial model, in the sense that it allows the probability of success to vary at each trial. Everything else is the same. So, instead of one probability of success p, we will have probabilities of success $p_1, p_2, \ldots, p_n$, one for each of the n trials.

Model: Consider an experiment where in each trial there are two possible outcomes, "success", A, and "failure", $\overline{A}$. The probability of success in the ith trial is p_i (and, accordingly, the probability of failure is $q_i = 1 - p_i$). Find the probability $P(n; k)$ that in n independent such trials, exactly k $(0 \leq k \leq n)$ successes occur.

The parameters of a Poisson model are n and $p_1, p_2, \ldots, p_n$ (not k).

Proposition 3.1.

The probability $P(n;k)$ in a Poisson model is given by

$$P(n;k) = \sum_{1\leq i_1<\cdots<i_k\leq n} p_{i_1}\dots p_{i_k}q_{i_{k+1}}\dots q_{i_n}, \quad k=0,1,\dots,n, \tag{3.1}$$

where $i_{k+1},\dots,i_n \in \{1,\dots,n\}\setminus\{i_1,\dots,i_k\}$.

Remark 3.2.

1. The number $P(n;k)$ is the coefficient of x^k in the polynomial expansion

$$(p_1x+q_1)\dots(p_nx+q_n) \quad = \quad \sum_{k=0}^{n} P(n;k)x^k$$

and, for the Poisson model, this is the computational formula that we will use.

2. Again, as a consequence (let $x=1$ above),

$$\sum_{k=0}^{n} P(n;k) \quad = \quad 1.$$

3. If $p_i = p$ (and consequently, $q_i = q$), $\forall i = \overline{1,n}$, then this becomes the Binomial model and (3.1) is reduced to (1.7) in Lecture 2.

Example 3.3 (The Three Shooters Problem).

Three shooters aim at a target and they hit it (independently of each other) with probabilities 0.4, 0.5 and 0.7, respectively. Each of them shoots once. Find the probability p that the target is hit once.

Solution.

A trial is "a person shoots the target". Define "success" as "the target is hit".

Then we have a Poisson model with $n = 3$ independent trials and $p_1 = 0.4$, $p_2 = 0.5$, $p_3 = 0.7$.

We want the probability of 1 success occurring. Hence $p = P(3; 1)$ and it is equal to the coefficient of x in the polynomial

$$(0.4x + 0.6)(0.5x + 0.5)(0.7x + 0.3) = 0.14x^3 + 0.41x^2 + 0.36x + 0.09,$$

i.e.

$$p = 0.36.$$

4. Pascal (Negative Binomial) Model

This model is a little different from the previous ones, in the sense that, we are not only interested in *number* of successes and failures, but also how they occur, i.e. in the rank of a success. Another novelty is that in this model we have (theoretically) an infinite number of trials.

Model: Consider an infinite sequence of Bernoulli trials with probability of success p (and probability of failure $q = 1 - p$) in each trial. Find the probability $P(n, k)$ of the nth success occurring after k failures ($n \in \mathbb{N},\ k \in \mathbb{N} \cup \{0\}$).

Remark 4.1.

For the Pascal model, again the parameters are n (rank of the success we want) and p (probability of success), but n has a different meaning than the one in the Binomial model. Again k is not a parameter of the model, it varies from 0 to ∞.

Proposition 4.2.

The probability $P(n,k)$ in a Negative Binomial model is given by

$$P(n,k) \quad = \quad C^k_{n+k-1} p^n q^k, \;\; k = 0, 1, \ldots . \tag{4.1}$$

Remark 4.3.

1. The probability $P(n;k)$ is the coefficient of x^k in the expansion

$$\left(\frac{p}{1-qx}\right)^n \quad = \quad \sum_{k=0}^{\infty} P(n,k) x^k, \;\; |qx| < 1,$$

hence the name.

2. As before,

$$\sum_{k=0}^{\infty} P(n,k) \quad = \quad 1.$$

5. Geometric Model

Although a particular case for the Pascal Model (case $n = 1$), the Geometric model comes up in many applications and deserves a place of its own.

Model: Consider an infinite sequence of Bernoulli trials with probability of success p (and probability of failure $q = 1 - p$) in each trial. Find the probability p_k that the first success occurs after k failures ($k \in \mathbb{N} \cup \{0\}$).

There is only one parameter for this model, p.

Proposition 5.1.

The probability p_k in a Geometric model is given by

$$p_k = pq^k, \quad k = 0, 1, \dots. \tag{5.1}$$

Remark 5.2.

1. The number p_k is the coefficient of x^k in the Geometric expansion (series)

$$\frac{p}{1-qx} \quad = \quad \sum_{k=0}^{\infty} p_k x^k, \ |qx| < 1,$$

hence the name.
2. Again,

$$\sum_{k=0}^{\infty} p_k \quad = \quad \sum_{k=0}^{\infty} pq^k \quad = \quad 1,$$

(the Geometric series).

Remark 5.3.

In a Geometric model setup, one might count the number of trials (not just *failures*) needed to get the 1^{st} success. The model would then be: In an infinite sequence of Bernoulli trials with probability of success p (and probability of failure $q = 1 - p$), find the probability $\tilde{p}_k$ that it takes k trials to get the first success ($k \in \mathbb{N}$). Then that would be

$$\tilde{p}_k \quad = \quad pq^{k-1}, \;\; k = 1, 2, \ldots .$$

Of course, if X is the number of failures and Y the number of trials, then we simply have $Y = X + 1$ (the number of failures plus the one success).

Example 5.4.

When a die is rolled, find the probability of the following events:
a) A: the first 6 appears after exactly 5 throws;
b) B: the 3$^{\text{rd}}$ even appears after exactly 5 throws.

Solution.

a) For event A, success means that face 6 appears, hence $p = 1/6$. We want the first success to occur after 5 failures, so this is a Geometric model. By (5.1), we have

$$P(A) = p_5 = \frac{1}{6}\left(\frac{5}{6}\right)^5 \approx 0.067.$$

b) For event B, success means that an even number shows, so $p = 1/2$. This fits the Pascal model with $n = 3$ and $p = 1/2$. The 3$^{\text{rd}}$ even appears after 5 throws (on the 6$^{\text{th}}$ throw), which means after 3 odds, i.e. after 3 failures. Thus, using (4.1), we have

$$P(B) = P(3,3) = C_5^3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^3 \approx 0.1562.$$

Chapter 3. Random Variables and Random Vectors

- to do a more rigorous study of random phenomena, we need to give them a more general quantitative description;
- that materializes in **random variables**, variables whose observed values are determined by chance;
- random variables are the fundamentals of modern Statistics;
- they fall into one of two categories:
  - *discrete* or
  - *continuous*.

1. Discrete Random Variables and Probability Distribution Function

Definition 1.1.

Let $(S, \mathcal{K}, P)$ be a probability space. A **random variable** is a function $X : S \to \mathbb{R}$ satisfying the property that for every $x \in \mathbb{R}$, the event

$$(X \leq x) \quad := \quad \{e \in S \mid X(e) \leq x\} \in \mathcal{K}. \tag{1.1}$$

Definition 1.2.

A random variable $X : S \to \mathbb{R}$ is a **discrete random variable** if the set of values that it takes, $X(S)$, is at most countable (i.e., finite or countably infinite) in $\mathbb{R}$.

Example 1.3.

Consider the experiment of rolling a die. Then the sample space is

$$S = \{e_1, \dots, e_6\},$$

where e_i represents the event that face i shows on the die, $i = \overline{1, 6}$.
Let $\mathcal{K} = \mathcal{P}(S)$ (all subsets of S) and P be given by classical probability.
Define $X : S \to \mathbb{R}$ by

$$X(e_i) \;=\; i,\; i = 1, \dots, 6.$$

Let us check that this is a discrete random variable.
For any $x \in \mathbb{R}$, the event (set) $(X \le x) \subseteq S$, so it obviously belongs to $\mathcal{K}$.
Thus X is a well-defined random variable (it satisfies (1.1)).
Since the set of values that it takes $X(S) = \{1, \dots, 6\}$ is *finite*, X is also a discrete random variable.

Example 1.4 (The indicator of an event).

Consider a probability space $(S, \mathcal{K}, P)$ over the sample space S of some experiment. For any event $A \in \mathcal{K}$, define $X_A : S \to \mathbb{R}$ by

$$X_A(e) = \begin{cases} 0, & e \notin A \ \ (e \in \overline{A}) \\ 1, & e \in A \end{cases} \tag{1.2}$$

First off, $X_A(S) = \{0, 1\}$, which is obviously *countable*.
Let us check condition (1.1).

- Let $x < 0$. Since all the values that X_A takes are nonnegative, there is no way that $X_A(e)$ could be $\leq x$, i.e.

$$(X_A \leq x) = \{e \in S \mid X_A(e) \leq x\} = \emptyset \in \mathcal{K},$$

since any σ-field contains the impossible event (empty set).

- If $0 \leq x < 1$, the event from (1.1) is

$$\begin{aligned}(X_A \leq x) &= \{e \in S \mid X_A(e) \leq x\} \\ &= \{e \in S \mid X_A(e) = 0\} \\ &= \overline{A} \in \mathcal{K},\end{aligned}$$

  because $A \in \mathcal{K}$.
- Finally for $x \geq 1$,

$$(X_A \leq x) = \{e \in S \mid X_A(e) \leq x\} = A \cup \overline{A} = S \in \mathcal{K},$$

  again, by the properties of a σ-field.

So X_A is a discrete random variable.

Remark 1.5.

A discrete random variable that takes only a finite set of values is called a **simple discrete random variable**. All of the examples above are simple discrete random variables.

The previous example can easily be generalized to any countable partition of the sample space S.

Example 1.6.

Let I be a countable set of indexes, $\{A_i\}_{i\in I} \subseteq \mathcal{K}$ a partition of S and $\{x_i\}_{i\in I} \subseteq \mathbb{R}$ a sequence of distinct real numbers. Define $X : S \to R$ by

$$X(e) = \sum_{i\in I} x_i X_{A_i}(e), \tag{1.3}$$

where X_{A_i} is the indicator of $A_i, i \in I$. Then X is a discrete random variable satisfying

$$X(e) = x_i <=> e \in A_i, \tag{1.4}$$

for all $i \in I$.

This is more than just an example, relation (1.3) gives the general expression of a discrete random variable.

Any discrete random variable can be written in the form (1.3).

Having the set of values that X takes, $\{x_i\}_{i\in I}$, X can be written as in (1.3), with $A_i = (X = x_i)$.

This justifies the next definition. Instead of defining a discrete random variable as a function $X : S \to \mathbb{R}$, we emphasize directly the values $\{x_i\}_{i\in I}$ that it takes and the probabilities of taking each value, $p_i = P(A_i) = P(X = x_i)$.

Definition 1.7.

Let $X : S \to \mathbb{R}$ be a discrete random variable. The **probability distribution function (pdf)**, or **probability mass function (pmf)** of X is an array of the form

$$X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}, \tag{1.5}$$

where $x_i \in \mathbb{R},\ i \in I$, are the values that X takes and $p_i = P(X = x_i)$ are the probabilities that X takes each value x_i.

Remark 1.8.

1. All values $x_i, i \in I$, in (1.5) are distinct. If some are equal, they only appear once, with the added corresponding probability.
2. All probabilities $p_i \neq 0, i \in I$. If for some $i \in I,\ p_i = 0$, then the corresponding value x_i is not included in the pdf (1.5).
3. If X is a discrete random variable with pdf (1.5), then

$$\sum_{i \in I} p_i = 1,$$

(a necessary and sufficient condition for such an array to represent a pdf of a discrete random variable). Indeed, since the events $\{(X = x_i)\}_{i \in I}$ form a partition of S, we have

$$\sum_{i \in I} p_i = \sum_{i \in I} P(X = x_i) = P(S) = 1.$$

4. Henceforth, we will identify a discrete random variable with its pdf and use (1.5) to describe it.

Example 1.9.

The pdf of the random variable in Example 1.3 (rolling a die) is

$$X \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

Example 1.10.

The pdf of the random variable in Example 1.4 (the indicator of an event) is

$$X_A \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}, \ p = P(A).$$