PRELIMINARIES: Combinatorial Analysis, Probability Spaces, and Random Variables

Marianna Bolla, Prof, DSc.
Institute of Mathematics, BME

February 13, 2024

Rules of Counting and Their Use in Combinatorial Probability Spaces

1. **Permutations.** How many different orders of n objects exist?

   - Without repetition (there are n different objects): $n!$
   - With repetition (there are n objects of which $n_1, \ldots, n_r$ are alike): $\frac{n!}{n_1! \ldots n_r!}$

2. **Variations.** How many different orders of k objects selected from a set of n objects exist?

   - Without repetition (an object is selected at most once): $n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}$
   - With repetition (an object may be selected several times): n^k

3. **Combinations.** How many different ways k objects can be selected from a set of n objects?

   - Without repetition (an object is selected at most once): $\frac{n!}{k!(n-k)!} = \binom{n}{k} = \binom{n}{n-k}, \quad k \le n$
   - With repetition (an object may be selected several times): $\binom{k+n-1}{k} = \binom{k+n-1}{n-1}, \quad k \in \mathbb{N}$

Identities containing binomial coefficients:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \qquad 2^n = \sum_{k=0}^{n} \binom{n}{k} \qquad 0 = \sum_{k=0}^{n} (-1)^k \binom{n}{k}$$

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k} \qquad \binom{n+m}{k} = \sum_{i=0}^{k} \binom{n}{i}\binom{m}{k-i} \quad \text{if } k \le \min\{n, m\}$$

Probability Space

$(\mathcal{S}, \mathcal{A}, \mathbb{P})$, where $\mathcal{S}$ is the *sample space* (set of all possible outcomes=*elementary events*), $\mathcal{A} = \{A \mid A \subset \mathcal{S}\}$ is the Boole-algebra of the all possible events (including $\emptyset$=impossible/null event and $\mathcal{S}$=certain/sure event), and the set function $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ satisfies the following **AXIOMS**:

1. For any $A \in \mathcal{A}$: $\quad 0 \le \mathbb{P}(A) \le 1$
2. $\mathbb{P}(\mathcal{S}) = 1$
3. For any sequence of mutually exclusive events $A_1, A_2, \ldots$: $\quad \mathbb{P}(\sum_i A_i) = \sum_i \mathbb{P}(A_i)$

Propositions implied by the axioms:

- $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$
- Probability is a monotone set function: if $A \subset B$, then $\mathbb{P}(A) \le \mathbb{P}(B)$
- $\mathbb{P}(\sum_{i=1}^n A_i) = \sum_{k=1}^n (-1)^{k-1} S_k, \quad S_k = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} \mathbb{P}(A_{i_1} A_{i_2} \ldots A_{i_k})$ (inclusion–exclusion)

De Morgan identities:

$$\overline{\sum_{i=1}^n A_i} = \prod_{i=1}^n \overline{A}_i \quad \text{and} \quad \overline{\prod_{i=1}^n A_i} = \sum_{i=1}^n \overline{A}_i$$

Examples of probability spaces:

- *Combinatorial*: the sample space has finite number of equally like outcomes, $\mathbb{P}(A) = |A|/|\mathcal{S}|$.
- *Geometric*: the sample space is a region with finite measure μ (length, area, volume), $\mathbb{P}(A) = \mu(A)/\mu(\mathcal{S})$.

Conditional Probability, Bayes Rule

- *Definition.* $\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}, \quad \mathbb{P}(B) > 0$. With B fixed, $\mathbb{Q}(A) := \mathbb{P}(A|B)$. $(\mathcal{S}, \mathcal{A}, \mathbb{Q})$ is also a probability space with all of its consequences.
- *Definition.* $B_1, B_2, \ldots$ is a complete set of mutually exclusive (disjoint) events, if $B_i B_j = \emptyset$ $(i \ne j)$ and $\sum_i \mathbb{P}(B_i) = 1$.
- **Theorem** (of complete probability). Let $B_1, B_2, \ldots$ be a complete set of mutually exclusive events and A be an arbitrary event. Then

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i).$$

- **Bayes' Theorem.** Let $B_1, B_2, \ldots$ be a complete set of mutually exclusive events and A be an arbitrary event. Then

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k) \cdot \mathbb{P}(B_k)}{\sum_i \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}, \qquad k = 1, 2, \ldots$$

- **Theorem** (factorization). Let $A_1, A_2, \ldots, A_n$ be arbitrary events. Then

$$\mathbb{P}(A_1 A_2 \ldots A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \ldots \mathbb{P}(A_n|A_1 \ldots A_{n-1}).$$

Independence

- *Definition.* A and B are independent if

$$\mathbb{P}(AB) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

  Remark: If $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$, then the independence of A and B means that $\mathbb{P}(A|B) = \mathbb{P}(A)$ and $\mathbb{P}(B|A) = \mathbb{P}(B)$; also, then A and B cannot be exclusive and independent at the same time. $\mathcal{S}$ and $\emptyset$ are independent of any other event.

- *Definition.* The events $A_1, \ldots, A_n$ are (completely) independent if

$$\mathbb{P}(A_1 \ldots A_n) = \mathbb{P}(A_1) \ldots \mathbb{P}(A_n).$$

  (If $\mathbb{P}(A_i A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for $i \neq j$, then $A_1 \ldots A_n$ are pairwise independent; this is weaker than independence.)

Random variables

- *Random Variable (r.v.):* an $X : \mathcal{S} \to \mathbb{R}$ stochastic function that is measurable with respect to $\mathcal{A}$. It means that

$$A = \{s \,|\, X(s) \in B\} \in \mathcal{A} \qquad \forall B \in \mathcal{B},$$

  where $\mathcal{B}$ denotes the set of Borel-sets of $\mathbb{R}$. *Distribution of* X: the collection of the probabilities $\mathbb{P}(A)$'s of the above A's. Of course, we need not give all of them.

- *Special types of random variables:*

  1. **Discrete probability distributions**: X takes on values $x_1, x_2, \ldots$. $\mathbb{P}(X = x_i) = p_i$, $i = 1, 2, \ldots$ $\quad (\sum_i p_i = 1)$. The collection of p_i's is called *probability mass function (p.m.f.)* of X. The *mode* of X: the value(s) taken on with the largest probability.
  2. **(Absolutely) continuous probability distributions:** The range of X is not countable and for any $x \in \mathbb{R}$: $\mathbb{P}(X = x) = 0$. However, there is an $f : \mathbb{R} \to \mathbb{R}$ nonnegative, integrable function such that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \quad \text{and} \quad \int_B f(x)\,dx = \mathbb{P}(X \in B), \quad \forall B \in \mathcal{B}.$$

f is called *probability density function (p.d.f.)* of X.
Cumulative distribution function (c.d.f.) of X: $F : \mathbb{R} \to \mathbb{R}$ such that

$$F(x) = \mathbb{P}(X < x) = \int_{-\infty}^{x} f(t)\ \mathrm{d}t, \quad x \in \mathbb{R}.$$

F is continuous, increasing, $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$. F is almost everywhere differentiable (at the points of continuity of f) and for such x's: $F'(x) = f(x)$.

$$\mathbb{P}(a < X < b) = F(b) - F(a) = \int_a^b f(x)\, dx \qquad (a < b).$$

(For discrete distributions the above F is a stepwise constant, increasing, left-continuous function.)

- **Expectation** of X (center of gravity of the mass distribution):
  1. $\mathbb{E}(X) = \sum_i x_i p_i$ (if it is absolutely convergent).
  2. $\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$ (if it is absolutely convergent).

  For $X \geq 0$: 1. $\mathbb{E}(X) = \sum_{i=0}^{\infty} \mathbb{P}(X > i)$ if $X \in \mathbb{N}$, 2. $\mathbb{E}(X) = \int_0^{\infty} (1 - F(x))\, dx$.

- **Variance** of X (inertia with resp. to the center of gravity of the mass distribution):

  $$\mathrm{Var}\,(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2) - \mathbb{E}^2(X), \quad \text{provided } \mathbb{E}(X^2) < \infty.$$

  Standard deviation of X: $\mathbb{D}(X) = \sqrt{\mathrm{Var}\,(X)} \geq 0$ and $=0$ if and only if $\mathbb{P}(X = cst.) = 1$.

- k-th *Moment* of X: $M_k(X) = \mathbb{E}(X^k)$, k-th *Central Moment* of X: $M_k^c(X) = \mathbb{E}(X - \mathbb{E}X)^k$ (if exists, then also exists for $1 \leq s < k$). $\mathbb{E}(X) = M_1(X)$, $\mathrm{Var}\,(X) = M_2^c(X) = M_2(X) - [M_1(X)]^2$.

- *Steiner's Theorem*: $\mathbb{E}(X - c)^2 = \mathbb{E}(X - \mathbb{E}X)^2 + (\mathbb{E}X - c)^2 \geq \mathrm{Var}\,(X)$, min. if $c = \mathbb{E}X$.

- p-**quantile** value or $100p$-*percentile* of X is x_p if $F(x_p) = p$. **Median**: 0.5-quantile value.

Examples to Notable Distributions, Prob. Models (see TABLES)

1. We have N balls, M red and $N - M$ white, mixed in an urn. n balls are selected randomly without replacement (or at the same time). Suppose that $n \leq \min\{M, N-M\}$. What is the probability that among the selected n balls there are k red ones ($k = 0, 1, \ldots, n$).

2. We have N balls, M red and $N - M$ white, mixed in an urn. n balls are selected randomly with replacement. What is the probability that among the selected (visited) n balls there are k red ones ($k = 0, 1, \ldots, n$).

3. What is the probability that by a 5-lottery ticket one wins a prize (one has at least a 2-hit)? (5 numbers are selected from $\{1, 2, \dots, 90\}$)

4. In a class of 20 students 8 are not prepared for the class. The teacher selects 5 students at random and asks them. Give the distribution of the number of students who are not able to answer the teacher's question among the selected 5.

5. What is the probability that we have a k-hit by filling in a TOTO ticket at random ($k = 0, 1, \dots, 13$)? (bet 1, 2, or x on the outcome of each of 13 soccer matches)

6. Give the distribution of the number of girls in a family having n children. Give the mode of this random variable! (The gender of the children is independent of each other with probability 1/2–1/2.) Equivalent problem: n fair coins are tossed, or a fair coin is tossed n times; give the distribution of the number of heads.

7. *Waiting for the first boy.* Consider the following population model: each family waits for a boy, and once they have him, they do not want more children. Give the boys/girls proportion in this population. (The gender of children is independent of each other with probability 1/2–1/2.)

8. Cookies are made in a big bakery: the blueberries are mixed into the mass and then the cookies are formed randomly. About how many blueberries have to be planned for a cookie, if they want to make the probability of possible complaints (of not having any blueberry in the cookie) as small as 0.01. Give the mode of the actual number of blueberries in a cookie!

9. Let X denote the lifetime of a radioactive isotope. Assuming, it has exponential distribution, prove the Markovian (ever-lasting) property of it:

$$\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t), \quad t, s > 0.$$

   If the halving time (median) is 100 years, find the parameter λ and the expected lifetime.

10. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ be Gaussian random variable. Calculate $\mathbb{P}(\mu - \sigma < X < \mu + \sigma)$ and $\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma)$.

The following theorem states that the Gaussian distribution has a distinguished importance. Let $X_1, X_2, \dots$ be i.i.d. r.v.'s (they are independent and identically distributed) with (existing) expectation μ and standard deviation σ. Let

$$\bar{X} := \frac{X_1 + \cdots + X_n}{n}, \quad n = 1, 2, \dots.$$

Obviously, $\mathbb{E}(\bar{X}) = \mu$ and $\mathbb{D}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.

Central Limit Theorem (CLT): In the above setup, the standardized $\bar{X}$ is approximately Gaussian:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma}\sqrt{n} \to \mathcal{N}(0, 1), \quad n \to \infty$$

in terms of the convergence in distribution (convergence of distribution functions).