INTRODUCTION TO PROBABILITY & STATISTICS. SYNOPSIS OF A COURSE AT OU WINTER SEMESTERS 2017/2018, 2018/2019, 2019/2020

PASHA ZUSMANOVICH

Each lecture lasted 1.5 hours.

Lecture 1. Basic combinatorics

Literature:

(1) Wikipedia: *Binomial coefficient*, *Pascal's triangle*, *Permutation*.
(2) R.L. Graham, D.E. Knuth, and O. Patashnik, *Concrete Mathematics*, 2nd ed., Addison-Wesley, 1994, Chapter 5.1.
(3) N.Ya. Vilenkin, *Combinatorics*, Academic Press, 1971.

A subject of probability and statistics.

To compute probabilities, we have to count, be it counting of discrete sets, like in combinatorics, or counting of areas of geometric figures, like in mathematical analysis.

Theorem 1.1. *The number of permutations of n elements is equal to $n!$.*

Proof. By induction. □

Variation without repetition = a way to choose k elements out of n elements, taking into account the order of elements.

Theorem 1.2. *The number of variations without repetition is equal to $\frac{n!}{(n-k)!}$.*

Variation with repetition = a way to choose k elements out of n elements, taking into account the order of elements, and with possible repetitions of elements.

Theorem 1.3. *The number variations with repetition is equal to n^k.*

Combination = a way to choose k elements out of n elements, without taking into account order of elements (and not allowing repetitions).

Theorem 1.4. *The number of combinations is equal to $\binom{n}{k}$.*

Proof. This is the same as doing variation without repetitions, but without accounting for different permutations of elements, i.e. the number in Theorem 1.2 should be divided by the number of permutations of k elements, which, according to Theorem 1.1, is equal to $k!$: $\frac{n!}{k!(n-k)!}$. □

These numbers are *binomial coefficients*, occurring in the binomial formula: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.

Properties of binomial coefficients: $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$, $\binom{n}{k} = \binom{n}{n-k}$.

Pascal's triangle.

A combinatorial meaning of the particular case of the binomial formula $2^n = \sum_{k=0}^{n} \binom{n}{k}$: all possible ways to choose elements out of n elements, i.e., the number of n-length binary sequences (like $0110\dots$, etc.)

Date: last modified October 24, 2019.

Inclusion-exclusion principle:

$$
\begin{aligned}
|A_1 \cup \cdots \cup A_n| = & \\
& |A_1| + \cdots + |A_n| \\
& -|A_1 \cap A_2| - |A_1 \cap A_3| - \cdots - |A_{n-1} \cap A_n| \\
& +|A_1 \cap A_2 \cap A_3| + \cdots + |A_{n-2} \cap A_{n-1} \cap A_n| \\
& - \cdots + \\
& +(-1)^{n+1}|A_1 \cap A_2 \cap \cdots \cap A_n|.
\end{aligned}
$$

Lecture 2. Probability function

Literature: [DKLM, Chapter 2].

Notions of sample space, event, elementary event.
Dictionary between set-theoretic and logical terms ($\cup$ = $\&$, $\cap$ = $\wedge$, complement = $\neg$).
Definition of probability function on finite and infinite sample space.

Claim 2.1. *If P is a probability function on a sample space Ω, then $P(\varnothing) = 0$.*

Proof. Follows from $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ for $A_1 \cap A_2 = \varnothing$ (take $A_2 = \varnothing$). $\square$

Claim 2.2. *If P is a probability function on a sample space Ω, then for any $A \subseteq \Omega$, $P(\Omega \backslash A) = 1 - P(A)$.*

Proof. Follows from $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ for $A_1 \cap A_2 = \varnothing$, and from $P(\Omega) = 1$ (take $A_1 = A$, $A_2 = \Omega \backslash A$). $\square$

Example: we are tossing an unbiased coin till the fist head. Our sample space is $\Omega = \{1, 2, 3, \ldots, n, \ldots\}$, where n signifies that the first head occurs at nth toss. Then $P(n) = \frac{1}{2^n}$, and

$$P(1) + P(2) + P(3) + \cdots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1,$$

as expected.

For arbitrary events $A_1, A_2, \ldots, A_n$, not necessary disjoint, we have

$$
\begin{aligned}
P(A_1 \cup \cdots \cup A_n) = & \\
& P(A_1) + \cdots + P(A_n) \\
& -P(A_1 \cap A_2) - P(A_1 \cap A_3) - \cdots - P(A_{n-1} \cap A_n) \\
& +P(A_1 \cap A_2 \cap A_3) + \cdots + P(A_{n-2} \cap A_{n-1} \cap A_n) \\
& - \cdots + \\
& +(-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n),
\end{aligned}
$$

what is exactly the inclusion-exclusion principle.

Lecture 3. Conditional probability, Bayes' formula, independent events

Literature:

(1) [DKLM, Chapter 3]
(2) Wikipedia: *Independence (probability theory)*

Definition and meaning of conditional probability.

Theorem 3.1. *Let Ω be a sample space, and B an event. Then $Q : \mathscr{P}(\Omega) \to [0, 1]$ defined as $Q(A) = P(A|B)$ is a probability function on Ω.*

Relationship between $P(A|B)$ and $P(B|A)$, Bayes' formula.

Theorem 3.2. *Let Ω be a sample space, $B_1, \dots, B_n$ events such that $B_1 \cup \dots B_n = \Omega$, and B_i's are pairwise disjoint. Then*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + \dots + P(A|B_n)P(B_n)}.$$

Proof. Using (several times) additivity of the conditional probability (what follows from Theorem 3.1), and Bayes' formula. □

Definition 3.1. Two events A and B are called independent, if one of the following six equivalent condition holds:

(i) $P(A|B) = P(A)$
(ii) $P(B|A) = P(B)$
(iii) $P(A \cap B) = P(A)P(B)$

and the same conditions (i)-(iii) with A and B being replaced by $\overline{A}$ and $\overline{B}$ (complements), respectively.

Generalization of this definition to the case of several events is not as straightforward as one might think at the first glance:

Definition 3.2. Events $A_1, \dots, A_n$ are called independent, if

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$$

for any $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$.

An alternative definition of independent events: events $A_1, \dots A_n$ are called independent, if

$$P(B_1 \cap \dots \cap B_n) = P(B_1) \dots P(B_n)$$

where each B_i's is either A_i or $\overline{A}_i$.

Equivalence of this definition with the definition from the previous lecture.

When events A and $\overline{A}$ are independent? Answer: if and only if $P(A) = 0$ or 1.

LECTURE 4. DISCRETE AND CONTINOUS RANDOM VARIABLES. DISTRIBUTION FUNCTION

Literature:

(1) [DKLM, Chapter 4, §5.1]
(2) Wikipedia: *Probability distribution*, *Probability mass function*.

Notion of discrete random variable, mass function, and distribution function. Example with "tossing two coins" with 0 and 1 (so $\Omega = \{0,1\} \times \{0,1\}$).

Properties of discrete distribution functions: piecewise constant, non-decreasing, $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to +\infty} F_X(x) = 1$.

Interplay between discrete and continuous. Continuous random variable as a limiting process of discrete random variables.

Continuous distribution function, density function, its properties. Density is not a probability!

Example: picking a point at a circle of radius R, the random variable $X(r)$ is the probability that the point will lie in a circle with radius r, $0 \le r \le R$.

LECTURE 5. EXPECTATION, QUANTILES

Literature: [DKLM, §§5.6, 7.1, 7.3].

Expectation of discrete and continuous random variables. Physical meaning of expectation as center of masses.

Formula for $E[g(X)]$.

Theorem 5.1. $E[aX+b] = aE[X]+b$.

Quantiles, median.

LECTURE 6. VARIANCE. UNIFORM DISTRIBUTION

Literature:

(1) [DKLM, §§5.2, 7.4]
(2) Wikipedia: *Discrete uniform distribution*, *Uniform distribution (continuous)*.

Definition and meaning of variance.

Theorem 6.1. $Var(X) = E[X^2] - E[X]^2$.

Proof goes separately for discrete and continuous distributions.

Theorem 6.2. $Var(aX+b) = a^2 Var(X)$.

Proof using definition of variance and Theorem 5.1.
Standard deviation.
Discrete and continuous uniform distributions, their expectation and variance.

LECTURE 7. BINOMIAL, GEOMETRIC, EXPONENTIAL DISTRIBUTIONS

Literature:

(1) [DKLM, §§4.3, 4.4, 5.3]
(2) Wikipedia: *Binomial distribution*, *Geometric distribution*, *Exponential distribution*.

Binomial distribution, its expectation and variance.

Geometric distributuon: the mass function is $f_X(k) = (1-p)^{k-1}p$.
$E[X] = \sum_{k=0}^{\infty}(1-p)^{k-1}pk = \frac{1}{p}$.
$Var(X) = E[X^2] - E[X]^2$.

$$E[X^2] = \sum_{k=0}^{\infty}(1-p)^{k-1}pk^2 = \sum_{k=1}^{\infty}(1-p)^{k-1}pk^2 = p\sum_{k=1}^{\infty}(-k\frac{d}{dp}((1-p)^k)) = -p\frac{d}{dp}\Big(\sum_{k=1}^{\infty}k(1-p)^k\Big)$$
$$= -p\frac{d}{dp}\Big(\frac{1-p}{p}\sum_{k=0}^{\infty}(1-p)^{k-1}pk\Big) = -p\frac{d}{dp}\Big(\frac{1-p}{p}E[X]\Big) = -p\frac{d}{dp}\Big(\frac{1-p}{p^2}\Big) = -p\frac{d}{dp}\Big(\frac{1}{p^2}-\frac{1}{p}\Big)$$
$$= -p(-\frac{2}{p^3}+\frac{1}{p^2}) = \frac{2}{p^2}-\frac{1}{p}.$$

$Var(X) = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$.
(Continuous) exponential distribution, its utility, expectation, and variance.

LECTURE 8. POISSON AND NORMAL DISTRIBUTIONS. SUM OF RANDOM VARIABLES

Literature:

(1) [DKLM, §§5.5, 10.1, 12.2]
(2) Wikipedia: *Poisson distribution*, *Normal distribution*.

(Discrete) Poisson distribition, its utility, expectation, and variance.
Normal distribution, its significance. Central limit theorem (imprecise, on an empirical level).

$E[X+Y] = E[X] + E[Y]$.
$E[a_1X_1 + \cdots + a_nX_n + b] = a_1E[X_1] + \cdots + a_nE[X_n] + b$.
Application of this formula: a short derivation of the formula $E[X] = pn$ for a binomial distribution with parameters p, n, without evaluating of the corresponding sums.

Lecture 9. covariance, correlation

Literature:

(1) [DKLM, Chapter 10]
(2) Wikipedia: *Covariance*, *Correlation and dependence*.

$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$.

If X and Y are independent random variables, then $E[XY] = E[X]E[Y]$. The opposite is not true: an example.

Property of correlation: $-1 \leq Cor(X,Y) \leq 1$. Correlation vs. dependence.

Lecture 10. Data analysis. Statistical models

Literature: [DKLM, §§15.2, 15.5, 16.1, 16.2, 16.3, 17.4, 22.1].

Graphical representation of datasets: histogram, scatterplot.

Numerical characteristics of datasets: mean, median, quantiles, variance.

Statistical models. Linear model (= linear regression). Method of least squares.

Lecture 11. Hypotheses testing. December 18, 2017

Literature: [DKLM, §25].

Hypotheses testing. Null and alternative hypotheses. Test statistic. p-values. Type I and type II errors.

References

[DKLM] F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, and L.E. Meester, *A Modern Introduction to Probability and Statistics*, Springer, 2005.

Email address: pasha.zusmanovich@gmail.com