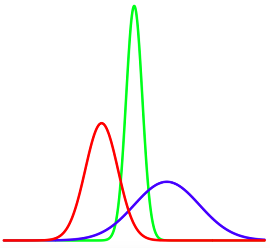
Statistics and Probability Theory

Revew Part 1



Please send your comments to: T. Levajković, tijana.levajkovic@tuwien.ac.at

Review

- Probability theory

  - Counting (Multiplication rule, permutations, combinations)
  - Computing probabilities, the Law of total probability, the Bayes theorem
  - Random variables (discrete and continuous)
  - Cumulative distribution function, p-quantiles,
    moments, central moments (in particular expectation and variance)
    some known random variables ($ber(p)$, $B(n, p)$, $\mathcal{P}(\lambda)$, geometric distribution,
    $\mathcal{U}(a, b)$, $\exp(\lambda)$, $\mathcal{N}(\mu, \sigma^2)$), transformations
  - Independence
  - Law of large numbers (LLN) and Central limit theorem (CLT)
  - Correlation between two random variables.

Examples

(1) There are two hospitals in a particular city. Every day 55 babies are born in the larger hospital and 18 babies are born in the smaller hospital. Over a year, each hospital recorded the days when more than 65% of babies born were boys.

(i) Let L be the number of days that more than 65% of babies born in the larger hospital were boys. What is the distribution of L? Calculate the expected value and the variance of L.

(ii) Let S be the number of days that more than 65% of babies born in the smaller hospital were boys. Find the distribution of S.

(iii) Use the CLT to approximate the 0.74-quantile of L.

(iv) Find the correlation between L and S.

(2) An accountant rounds to the nearest euro. We assume the error in rounding follows uniform distribution on $(-0.5, 0.5)$. Estimate the probability that the total error in 300 entries is more than 5 euro.

Examples

(3) Let $X_1, X_2, \ldots, X_{25}$ be i.i.d. random variables with $\mathcal{N}(1,4)$. Compute

$$P(X_1 + X_2 + \ldots\ldots X_{25} \geqslant 26).$$

(4) Let $X \sim \text{uniform}(0,1)$.

- (i) Compute EX and $Var\,X$.
- (ii) Find the median of $Y = -\log X$.

(5) Let X be a continuous random variable with pdf

$$f(x) = \begin{cases} kx^2, & x \in [0,1] \\ 0, & \text{else} \end{cases}$$

and let $Y = X^3$.

- (i) Find k and the cdf of X.
- (ii) Find the 30^{th} percentile of X.
- (iii) Compute EY and $Var\,Y$.

Some multiple-choice examples

(1) Let Z be a standard normal distributed random variable and let $X = 0.5 - 4Z$. Calculate

$$P(|X| \leqslant 2.1).$$

Use the values given in the table below.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051

Table: Cumulative distribution function of the standard normal distribution

a. 0.2422
b. 0.7578
c. 0.3976
d. 0.6024

(2) Let $Z_1, \ldots, Z_{100}$ be i.i.d. $\mathcal{N}(0,1)$ random variables. The correlation between $X = \sum_{i=1}^{98} Z_i$ and $Y = \sum_{i=3}^{100} Z_i$ is equal to

(a) 0
(b) $\frac{96}{98}$
(c) $\frac{98}{100}$
(d) 1

(3) Given IQ scores are approximatelly normally distributed $\mathcal{N}(100, 15^2)$, the proportion of people with IQs above 130 is

(a) 95%
(b) 68%
(c) 5%
(d) 2.5%

(4) Suppose we have a random variable X where $P(X = k) = \binom{15}{k} 0.29^k \, 0.71^{15-k}$ for $k = 0, \ldots, 15$. What is the expectation of X?

(a) 0.2059

(b) 3.0885

(c) 4.35

(d) 10.65

(5) Let X_1, X_2 and X_3 be uniform random variables on the interval $(0, 1)$ with $Cov(X_i, X_j) = \frac{1}{24}$ for $i, j = 1, 2, 3$, $i \neq j$. Then the variance of the sum $X_1 + 2X_2 - X_3$ equals

(a) $\frac{1}{6}$

(b) $\frac{5}{12}$

(c) $\frac{3}{8}$

(d) $\frac{3}{4}$

(6) Suppose that a test consists of 20 True-False questions. A student has not studied for the exam and will just randomly guess at all answers (with True and False equaly likely). How would you find the probability that the student will get at most 8 answers correct?

(a) Find the probability that $X = 8$ in a binomial distribution $bin(20, 0.5)$.
(b) Find $F(8)$, where F is the cdf of a binomial distribution $bin(20, 0.5)$.
(c) Find $F(8)$, where F is the cdf of a normal distribution $\mathcal{N}(10, 5^2)$.
(d) Find the area between 0 and 8 in a uniform $(0, 20)$-distribution.

(7) According to the empirical rule, the bell shaped distribution will have approximately 68% of the data within what number of standard deviations of the mean?

(a) one standard deviation
(b) two standard deviations
(c) three standard deviations
(d) none of the above

(8) A bowl contains 10 balls, of which 4 are red and 6 are black. Balls are randomly selected with replacement from the bowl until 4 red balls have been selected. Let X be the number of black balls drawn before the fourth red ball is selected. Then, the probability $P(X = 6)$ equals

(a) 0.0012
(b) 0.0446
(c) 0.1003
(d) 0.2508

(9) Let A and B be two events with $P(A) = 0.5$, $P(B) = 0.4$ and $P(A \cap B^c) = 0.2$. Then the probability $P(B^c | A \cup B)$ equals

(a) $\frac{1}{2}$
(b) $\frac{1}{3}$
(c) $\frac{1}{4}$
(d) 0

(10) Let $X_1, \dots, X_{100}$ be a random sample from a exponential distribution with the expectation $\frac{1}{2}$. Determine the approximative value of

$$P(\sum_{i=1}^{100} X_i > 57)$$

using the Central limit theorem.

(a) 0.08
(b) 0.16
(c) 0.31
(d) 0.46

(11) Let $X \sim \mathcal{N}(0, a)$, with $a > 0$. Compute $P(X^2 < a)$.

(a) 0.34
(b) 0.42
(c) 0.68
(d) 0.84

(12) Anna has an unfair coin. She knows that when fliping this coin, head is obtained twice as often as a tail. Anna rolls this coin until she obtains a tail. What is the probability that she rolled the coin at most three times?

a. 19/27
b. 26/27
c. 4/27
d. 5/9

(13) Suppose box A contains 4 red and 5 blue coins and box B contains 6 red and 3 blue coins. A coin is chosen at random from box A and placed in box B. Finally, a coin is chosen at random from among those that are now in box B. What is the probability a red coin was transferred from box A to box B given that the coin chosen from box B is blue?

a. 5/8
b. 3/8
c. 16/45
d. 2/9

(14) A medical treatment has a success rate of 0.75. Two patients will be treated with this treatment. Assuming the results are independent for the two patients, what is the probability that at most one of them will be successfully cured?

a. 0.500
b. 0.4375
c. 0.3750
d. 0.4750

(15) Assume that X is a binomial random variable with $n = 100$ and $p = 0.1$. Use the normal probability distribution to compute $P(X \geqslant 15)$.

a. 0.5336
b. 0.9664
c. 0.0336
d. 0.4664

(16) Let X be a random variable that takes values 0, 2 and 3 with probabilities 0.3, 0.1 and 0.6 respectively. Let $Y = 3(X-1)^2$. Then,

(a) The variance of X is larger than 2 and the variance of Y is less than 20.
(b) The variance of X is less than 2 and the variance of Y is less than 20.
(c) The variance of X is larger than 2 and the variance of Y is larger than 20.
(d) The variance of X is less than 2 and the variance of Y is larger than 20.

(17) As part of a large promotion, both you and your roommate are given free cellular phones from a batch of 50 phones. Unknown to you, five of the phones are faulty and do not work. Find the probability that one of the two phones is faulty.

a. 0.0082
b. 0.1837
c. 0.8082
d. 0.8164

(18) A fair die is rolled. Find the probability of getting an odd number or a number less than 5.

a. 7/12
b. 2/3
c. 1/3
d. 5/6

(19) Transportation officials tell us that 60% of the population wear their seatbelts while driving. A random sample of 1000 drivers has been taken. What is the probability that between 580 and 630 of the drivers were wearing their seatbelts?

a. 0.8822
b. 0.9756
c. 0.9066
d. 0.0934

(20) Temperatures in Vienna for the month of August follow a uniform distribution over the interval 22 to 27 degrees Celsius. Find the temperature which 90% of the August days exceed.

a. 22.5
b. 23.5
c. 24.5
d. 26.5

(21) The IQs of army volunteers in a given year are normally distributed $\mathcal{N}(110, 10^2)$. The army wants to give advanced training to the 25% of those recruits with the highest IQ scores. What is the lowest IQ score acceptable for the advanced training?

(a) 116.75
(b) 106.75
(c) 113.25
(d) 103.50

(22) A state energy agency mailed questionnaires on energy conservation to 1000 homeowners in Innsbruck. Five hundred questionnaires were returned. Suppose an experiment consists of randomly selecting one of the returned questionnaires. Consider the events:

A: The home is contructed of brick
B: The home is more than 30 years old.

A home that is constructed of brick and is less than or equal to 30 years old can be described in terms of events A and B as

(a) $A \cup B$
(b) $A \cap B$
(c) $(A \cap B)^c$
(d) $A \cap B^c$

(23) Calculate the sample mean for the following data: 3 9 8 5 7 9 4 6 5 5 4.

a. 5.5
b. 6
c. 5.94
d. 5.91

(24) A literature professor decides to give a 20-question true-false quiz. Each question is worth 1 point. He wants to choose the passing grade so that the probability of passing a student who guesses on every question is less than 0.10. What score should be set as the lowest passing grade?

(a) 12
(b) 13
(c) 14
(d) 15

(25) Which of the following statements is true for random variables?

(a) A random variable assumes numerical values associated with the random outcomes of an experiment; only one value can be assigned to each sample point.
(b) A random variable can only assume discrete values.
(c) A random variable assumes numerical values determined by a random number generator.
(d) A random variable assumes numerical values associated with the random outcomes of an experiment; more than one value can be assigned to each sample point.

Thank you for your attention!