Examination Statistics I February 9, 2005

Otto-von-Guericke-Universität Magdeburg
Fakultät für Mathematik
Priv.-Doz. Dr. Th. Müller-Gronbach

Please note the following:

- The exam consists of 7 problems to be solved; the credit points you can get for each problem are given in brackets next to the number of the problem. You do not have to solve the individual problems completely, partial solutions are also possible. Except for problem 1, it is not enough to state the result only, but you should clearly display your approach and way to solution.
- You can reach a maximum of **50 points**. For passing the exam a total of **22 points** is sufficient.
- You are allowed to use: Pocket calculators, text books, mathematical and/or statistical tables, manuscripts and notes from lectures and/or exercises.

Problem 1 *(4 points)*
The following statements are TRUE or FALSE. So just answer TRUE or FALSE without giving any explanation.

(a) If A_1, A_2, A_3 and A_4 are independent events, then $A_1 \cup A_3$ and $A_2 \cap A_4$ are also statistically independent.

(b) If A and B are independent events, then

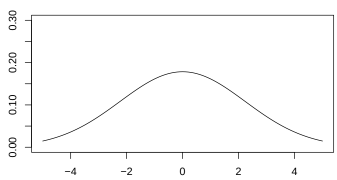
$$P(A \cup B) = P(A) + P(B).$$

(c) A probability density function can not take values greater than 1.

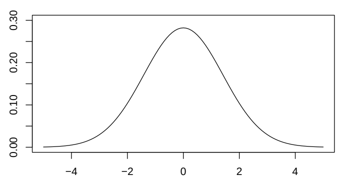
(d) If two random variables X and Y have the same distribution, then

$$P(X = Y) = 1.$$

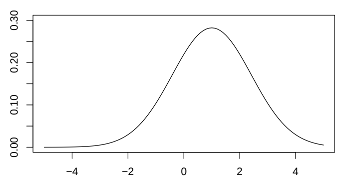
Problem 2 *(5 points)*
The following figures show density functions of three normally distributed random variables X, Y, Z. The expectations of these random variables are 0 or 1, the variances are 2 or 5.



X



Y



Z

For each of these random variables determine the corresponding expectation and variance. Justify your answer.

Problem 3 *(8 points)*
The following values are measured access times (in seconds) to a data storage device:

12.4 18.2 24.0 15.6 14.4 12.8 14.6 15.7 14.9 13.0 17.1

(a) Compute the values of the corresponding empirical distribution function at the points 14, 16 and 17.1.

(b) The arithmetic means of the data and of the squared data are 15.7 and 256.24, respectively. Use only this information to compute the empirical variance.

(c) Determine a sample median.

Problem 4 *(10 points)*
Mr. Miller has to use two different busses every morning to get to his office. He has to change from bus A to bus B at the central bus station. The time of arrival T_A (in minutes after 7:00 a. m.) of bus A at the station is normally distributed with mean 10 and variance 1. The time of departure T_B (in minutes after 7:00 a. m.) of bus B is also normally distributed and independent of the arrival time of bus A. The difference $T_B - T_A$ has mean 3 and variance 3.

(a) Compute the probability that bus A arrives at the station between 7:09 a. m. and 7:10 a. m.

(b) What is the expected time of departure of bus B? What is the variance of the departure time of bus B?

(c) What is the probability that Mr. Miller reaches bus B if the time needed to change the bus is 1 minute?

(d) If Mr. Miller misses bus B, he has to spend 6 Euro for a taxi. Compute the probability that within 5 days he has to spend more than 10 EUR for taxis. State the model you are using.

Problem 5 *(8 points)*
The electrical power supply of a satellite is established by three independently working sun sails. It is known that the lifetime of each of these sun sails is exponentially distributed. The mean lifetime of one sunsail is 120 months. The satellite is only fully operationable if all three sun sails are working. The satellite can still send data as long as at least one sun sail is working.

(a) Compute the probability that the satellite is fully operationable for at least 5 years.

(b) What is the probability that the satellite stops sending data within the first 10 years?

(c) Compute the maximum time up to which the satellite can send data with a probability of 99.9%.

Problem 6 *(8 points)*
The type of an incoming email in a company is modeled by a random variable E that can take the values "s" for spam, "p" for private mail and "b" for business mail. The company uses a spam filter that classifies each mail as "s" for spam or as "r" for regular mail. The classification of an email is modeled by a random variable C. The following table shows parts of the joint distribution of E and C.

		values of E			
		s	p	b	$\sum$
values of C	s	0.27			
	r				
	$\sum$		0.1	0.6	

It is known that 10% of the private mail and 5% of the business mail is falsely classified as spam.

(a) Compute the probability that a mail is classified as spam given that it really is spam.

(b) Compute the probability that an arbitrary mail is classified as spam.

(c) Determine the missing values in the table.

(d) Compute the probability that a mail is spam given that it is classified as spam.

Problem 7 *(7 points)*
The number of traffic accidents within one week at a certain dangerous crossing is modeled by a random variable with mean 2 and variance 2. Numbers of accidents for different weeks are assumed to be independent.

(a) Can you guarantee that with a probability of at least 0.1 the total number of accidents within 4 weeks is less than 12 and greater than 4?

(b) Compute approximately the probability that there are more than 328 accidents within 162 weeks.

State the models you are using in (a) and (b).