Exercises to
Probability Theory and Statistics
(sf1901)

These are essentially translations of the Swedish pdf-exercises, but I have replaced a few of them with – in my opinion – more relevant ones. Harald Lang, 2016.

1.1 Let A and B be two events. Express the following events in words, and display them in a Venn diagram:

a) $A \cap B$ b) $A \cap B^*$ c) $A^* \cap B^*$

d) Prove (using a Venn diagram) the rule

$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$

e) Prove (using a Venn diagram) de Morgan's rules:

$(A \cup B)^* = A^* \cap B^*$ and $(A \cap B)^* = A^* \cup B^*$

1.2 At the production of a certain item, two types of defects, A and B, can occur. We know that P(A) = 0.1, P(B) = 0.2 and $\mathrm{P}(A \cap B) = 0.05$. Compute the probability that a produced unit has

a) at least one of the defects

b) defect A but not defect B

c) none of the defects

d) precisely one of the defects A and B

1.3 We take at random three cards (without replacement) from an ordinary deck of cards (of 52 cards). Compute (using the classical probability definition) the probabilities for the events

a) all three are hearts

b) none of the cards is hearts

c) all three are aces.

1.4 (Challenging!) Compute the probability that out of 23 randomly selected individuals, at least two of them have the same birthday. (Assume that all birthdays are equally common, and that there are 365 days in a year.)

1.5 Discuss the *Monty Hall problem* (see Wikipedia). Answer the following:

a) If you decide in advance to always stick to your originally chosen door, what is your chance to win the car?

b) If you decide in advance to always change door (when there are two closed doors,) what is your chance to win the car?

2.1 The two events A and B are independent with $P(A) = 0.1$ and $P(B) = 0.05$. Compute $\mathrm{P}(A^* \cap B^*)$. What can we say about the events A^* and B^*?

2.2 The two events A and B have both positive probabilities.

a) If they are disjoint, can they be independent?

b) If they are independent, can they be disjoint?

2.3 The two events A and B have probabilities $\mathrm{P}(A) = 0.1$, $\mathrm{P}(B) = 0.05$ and $\mathrm{P}(A \cup B) = 0.8$. Determine if A and B are independent events.

2.4 We have three lottery tickets; one is a winning ticket and two are blanks. The three individuals Joe, Kim and Mary draw a ticket each in order, (Joe first, Mary last.)

a) What is the probability that Joe wins?

b) that Kim wins?

c) that Mary wins?

2.5 From a sign with the text MALMO two random letters fall to the ground. An individual, who is ignorant of the original text, puts them up again at random. What is the probability that the original text will be restored?

2.6 Assume that the probability that a born baby is a boy is 0.5, and that gender is independent between births. A family has four children. Compute

a) the probability that they have two boys and two girls, conditional that the first (oldest) is a boy

b) the probability that they have two boys and two girls, conditional that at least one is a boy.

2.7 For the three events A, B and C it holds that

$P(A \cap B \cap C) = 0.1$, $P(A) = 0.5$ and $P(B \mid A) = 0.4$.

Compute the probability $P(C \mid A \cap B)$.

3.1 At a production process, the produced items are tested for defects. A defective unit is classified as such with probability 0.9, whereas a correct unit is classified as such with probability 0.85. Furthermore, 10% of the produced units are defective.

Compute the conditional probability that a unit is defective, given that is has been classified as such.

3.2 *A* and *B* are two independent events with probabilities 0.5 and 0.4 respectively. Compute the conditional probability that both *A* and *B* occur, when it is known that at least one of the events *A* and *B* has occurred.

3.3 In a certain risk group, individuals are tested for a certain disease *S*. A person who has the disease gets the correct diagnosis with probability 0.99, whereas a person who does not suffer from *S* gets the correct diagnosis with probability 0.95. Furthermore, it is known that 6% of the individuals in the group get the diagnosis "suffer from *S*".

Determine a) the proportion of individuals in the group who suffer from *S*, and b) the probability that a person who gets the diagnosis "suffer from *S*" actually carries the disease.

3.4 We have two urns, *A* and *B*, who contain black and white balls. *A* contains two black and three white balls, whereas *B* contains two black and two white balls. We draw one ball at random from *A* and put it in *B*, without noticing its colour. Next we draw at random one ball from *B*, and notice that it is white.

Compute the probability that the ball we moved from *A* to *B* was black.

4.1 A library has on average 78 visitors on an ordinary Friday. Compute the probability that the library has more than 90 visitors on an ordinary Friday. Motivate carefully!

4.2 Fifteen persons toss coins; two coins each.

a) What distribution has the number of individuals who get the same result on the two coins?

b) Compute the probability that at most six individuals get the same result on the two coins.

4.3 Two defective units have accidentally been mixed in among three flawless units. In order to find the defective units, the units are tested one at a time, in order, until either the two defective units have been found, or the three flawless units have been found.

a) Determine the distribution for X, the number of tested units.

b) Compute the expected value $E(X)$ and the standard deviation $SD(X)$.

4.4 A certain Friday evening 1% of the car drivers are intoxicated.

a) At a road check 50 drivers are tested. Compute the probability that at least one intoxicated driver is caught.

b) At a road check drivers are tested. Let X be the number of tested drivers when the first intoxicated driver is caught. Compute the expected number $E(X)$ and the standard deviation $SD(X)$.

4.5 Prove that if $X \in \text{Po}(\lambda)$ and $Y \in \text{Po}(\mu)$ are independent, then $X \mid (X+Y=n) \in \text{Bin}(n, \lambda/(\lambda+\mu))$.

5.1a) Determine the constant c such that the function

$$f(x)=\begin{cases} cx^2 & 0<x<6 \\ 0 & \text{otherwise} \end{cases}$$

becomes a probability density function.

b) Determine the corresponding distribution function.

c) Compute the expected value and variance.

5.2 A random variable X has the distribution function

$$F_X(x)=\begin{cases} 0 & x<0 \\ x^3 & 0\le x\le 1 \\ 1 & x>1 \end{cases}$$

Compute E(X) and SD(X).

5.3 The length X of a telephone call can often be approximated as an $\exp(\lambda)$ random variable. If $\lambda = 2/3\,\text{min}^{-1}$, compute

a) the probability $\text{P}(1<X\le 10)$.

b) E(X) and SD(X).

5.4 It is a fact that if $X\in\exp(\lambda)$, then $2\lambda X\in\chi^2(2)$. Use this fact to compute the probability in 5.3.a) with your pocket calculator.

5.5 In a coordinate system, the interval [–a, a] on the x-axis represents a fault fissure. An earth quake will occur such that its epicentre X on that interval follows a uniform probability distribution. At (0,d) is a dam. Let Y be the distance from the dam to the epicentre X. Determine the density function for Y.

5.6 $X\in\exp(\lambda)$.

a) Determine the density function for $Y=e^X$.

b) Use a calculator to compute $\text{E}\left(\sqrt{X}\right)$.

6.1 X and Y are independent, discrete random variables whose probability functions are given in the tables below:

k	1	2	3
$p_X(k)$	$1/3$	$1/2$	$1/6$

k	1	2	3
$p_Y(k)$	$2/3$	$1/6$	$1/6$

a) Compute the probability $P(X+Y=4)$.

b) Compute the conditional probability $P(X \leq 2 \mid X+Y=4)$.

6.2 Two friends A and B have decided to meet at a cafe "somewhat after eight o'clock". A arrives X minutes past eight, and B arrives Y minutes past eight.

Assume now that $X \in U(0,4)$ and $Y \in U(0,6)$, and that X and Y are independent. What is the probability that A has to wait for B?

6.3 X and Y are independent exponentially distributed random variables with expected values 2. Let $U = \max(X, Y)$ and $V = \min(X, Y)$. Compute $E(U - V)$.

6.4 X and Y are as in 6.3. Determine the density function for $Z = X + Y$.

6.5 $X \in \exp(\lambda)$. We draw a random value x of X, and let $Y \in \exp(x)$. Compute the probability $P(Y > 1)$.

7.1 The two random variables X and Y have expected values $\mathrm{E}(X) = -1$ and $\mathrm{E}(Y) = 1$, and variances $\mathrm{V}(X) = 2$ and $\mathrm{V}(Y) = 4$. The covariance is $\mathrm{Cov}(X, Y) = -1$. Compute the standard deviation $\mathrm{SD}(X - 2Y + 1)$.

7.2 The random variable $X \in \exp(2.5)$. Compute $\mathrm{E}\left(e^{-X}\right)$ and $\mathrm{V}\left(e^{-X}\right)$.

7.3 $X \in \mathrm{U}(1, 2)$. We draw a random value x of X, and let $Y \in \exp(x)$. Compute the expected value $\mathrm{E}(Y)$.

7.4 The random variables X_1, X_2 and X_3 are independent and have all the expected value 2 and standard deviation 3. Let

$$Y = 3X_1 - 2X_2 + X_3 - 6.$$

a) Compute $\mathrm{E}(Y)$ and $\mathrm{SD}(Y)$.

b) Compute the correlation coefficient $\rho(Y, X_1 + X_2)$.

7.5 The random variable X has density function $f_X(x) = 0.5x,\ 0 \le x \le 2$. Compute $\mathrm{V}(X^2)$.

7.6 In certain breakfast cereal packs, plastic toys are placed. Each pack contains one toy, and there are in total six different toys that are distributed evenly among packs. You don't know which toy the pack contains until you open it.

You buy packs, one at a time, and let X be the number of packs bought when you have just collected at least one of each toy. Compute $\mathrm{E}(X)$.

8.1 $X \in \mathrm{N}(0,1)$. Compute

a) $\mathrm{P}(X \le 1.8)$

b) $\mathrm{P}(X \le -1.35)$

c) $\mathrm{P}(-1.2 < X < 0.5)$

d) a, such that $\mathrm{P}(X > a) = 0.05$

e) a, such that $\mathrm{P}(|X| < a) = 0.95$

8.2 $X \in \mathrm{N}(5, 2)$. Compute

a) $\mathrm{P}(X \le 6)$

b) $\mathrm{P}(1.8 < X < 7.2)$

c) a, such that $\mathrm{P}(X \le a) = 0.05$

8.3 X, Y and Z are independent normally distributed random variables with expected values 2, 1 and 0, respectively. They all have the sane variance 2. Compute $\mathrm{P}(4X - 3Y > 5Z)$.

8.4 A lift in a department store has a sign "*a maximum 10 persons, or at most 800 kg.*" Assume that a random person's weight is N(70, 10) kg. What is the probability that 10 customers will overload the lift?

8.5 A new residential area is planned for 1'000 families. The probability that a family owns 0, 1, 2 and 3 cars is, 0.2, 0.4, 0.3 and 0.1, respectively. How many parking places have to be planned for, in order that with 90% probability all cars have a parking place?

8.6 $X \in \mathrm{Bin}(2300, 0.25)$ and $Y \in \mathrm{Po}(640)$. X and Y are independent. Compute (approximately) the probability $P(X > Y)$.

9.1 A measuring instrument uses a special kind of battery. Five batteries of this type have been used to measure the time spell they deliver enough power. The result was 5, 4, 6, 4, 7 hours. These values can be regarded as observations of a random variable X with expected value μ and variance σ^2.

a) Estimate μ and σ^2 in some suitable way.

b) Show that estimate of μ is unbiased.

c) compute the standard deviation and standard error of the estimate of μ.

9.2 We have a random sample $x_1, \ldots, x_n$ ($n > 2$) of an N(μ ,σ) random variable. Consider the two estimates

$$\mu^* = \frac{1}{n}(x_1 + \ldots + x_n) \text{ and } \hat{\mu} = \frac{1}{2}(x_1 + x_n).$$

a) Show that both estimates are unbiased.

b) Which of the two estimates is most efficient?

9.3 $x_1, \ldots, x_n$ are independent observations of a *Maxwell* distributed random variable, i.e. its density function is

$$f(x) = \sqrt{\frac{2}{\pi}} \frac{x^2}{\alpha^{3/2}} e^{-x^2/(2\alpha)}, \quad x > 0,$$

where α is a parameter. This distribution has expected value $\sqrt{8\alpha/\pi}$ and variance $\alpha(3 - 8/\pi)$.

a) Compute the ML-estimate $\hat{\alpha}_{ML}$ of α.

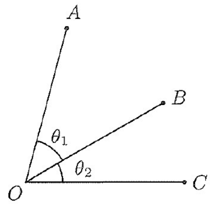
b) Is $\hat{\alpha}_{ML}$ an unbiased estimate of α?

9.4 We have three independent measurements x_1, x_2, x_3 of the angle AOC, and two independent measurements x_4, x_5 of AOB (see picture below.) The measurements are unbiased with a random error with standard deviation σ.

a) Compute the LS estimates of θ_1 and θ_2.

b) Are the estimates unbiased?

c) Compute the variances of the estimates.



9.5 The discrete random variable X has probability function $p(j) = \theta(1-\theta)^{j-1}$, $j = 1, 2, 3, \ldots$ where $0 < \theta < 1$. We have six independent observations of X: 4, 5, 4, 6, 4, 1.

a) Compute the ML-estimate of θ.

b) Compute the LS-estimate of θ.

10.1 Four independent measurements of a distance have been performed with an instrument which gives an unbiased result, but with a random error which is normally distributed with standard deviation $\sigma = 0.005$ mm. The measurements were (mm)

1132.155 1132.158 1132.145 1132.163

a) Compute a 95% confidence interval for the true distance.

b) As a) but assume that σ is unknown.

10.2 Two nursery schools are located in the same town, but one of them, "The Clearing" is located close to a forest, whereas the other, "The Brick Wall" is located downtown. The environmental manager suspects that the children at "The Brick Wall" might get higher content of lead in their blood, due to the location. In order to find out, five children were randomly drawn from each nursery school, and their lead contents were measured (ng/ml). The result was

The Clearing:	0.96	0.43	0.93	0.85	0.48
The Brick Wall:	0.93	0.63	1.21	1.30	0.58

Do these data indicate a difference between the two nursery schools? Construct a 95% confidence interval to answer the question.

a) Assume that the variation in lead content between children is normally distributed with the same variance for the two nursery schools.

b) As in a) but do not assume equal variances.

10.3 In order to compare two scales, six items were weighed with each of the two scales. The results were:

item	1	2	3	4	5	6
scale A	1.0	7.7	9.6	21.0	32.3	22.6
scale B	3.1	8.8	12.0	19.5	35.5	32.5

The measurements can be regarded as observations of normal random variables. Compute a 95% confidence interval for the difference μ_{B-A}

a) when the standard deviations are known to be $\sigma_A = 2$ and $\sigma_B = 3$

b) when the standard deviations are unknown.

10.4 During 30 days at an upstream location of the Mississippi river a certain contamination was measured. Since the measurements were made at different days, they can be considered as independent. The sample mean value was $\overline{x} = 13.2$ and sample standard deviation $s = 2.8$. At a location downstream, 40 measurements were made at different days. Here the sample mean was $\overline{x} = 86.1$ and the sample standard deviation $s = 38.7$. Compute a 95% confidence interval that can be used to assess the difference in contamination between the two locations. (Note that we do not assume normal distributions.)

10.5 A physicist has made five measurements in order to determine a certain physical constant. These measurements can be regarded as observations of a normally distributed random variable with known variance. He computes a 90% confidence interval to be (7.02, 7.14). How many measurements would he need in order to get

a) a confidence interval with half the length?

b) a 99% confidence interval of the same length?

c) a 99% confidence interval with half the length?

11.1 Two firms, *A* and *B*, offer telephone support. One wanted to estimate the difference in wait time before a call is answered between the two firms. The wait time for 420 calls to firm *A* was on average 26.0 minutes, and for 376 calls to firm *B* the wait time was on average 31.6 minutes. Assume that wait time is exponentially distributed with expected values μ_A and μ_B respectively. Construct a 95% confidence interval for the difference $\mu_B - \mu_A$. Can we draw any conclusion (with error risk 5%) as to which firm has the longer expected wait times?

11.2 The traffic on a certain road can be described such that the number of cars that pass a certain point on the road during a time period of *t* minutes is $\mathrm{Po}(\lambda t)$. During a time period of 10 minutes, 400 cars passed. Compute a 95% confidence interval for λ.

11.3 In a psychological experiment 198 economics students were asked "*Should Sweden join the EMU and introduce the euro as currency?*" 107 said "Yes" and 91 "No". Another 198 students were asked "*Should Sweden join the EMU and abandon the Swedish krona as currency?*" 95 answered "Yes" and 103 "No".

Compute a 95% confidence interval for the difference in proportions of students who answer "yes" in the entire student population between the two questions. Can we draw any conclusion whether the formulation of the question influences the answer?

11.4 Assume that the number of fatal traffic accidents during two consecutive years was 537 and 453, respectively. The minister in charge described this decline in accidents as a result of his clever policy, whereas his opponents claimed it was pure chance.

Compute a 95% confidence interval for the difference in intensity of fatal accidents between the two years. Can we reject the hypothesis that the outcome was pure chance?

11.5 The number X of calls to a telephone exchange during the busiest hour is Po(λ). During eight days the following observations on X were made:

$$115, 82, 108, 106, 118, 87, 99, 92.$$

Compute a 95% confidence interval for λ.

12.1 Joe plays on a slot machine which gives a win with probability p. The claim is that $p = 0.2$, but Joe suspects that $p < 0.2$.

Define the hypotheses H_0: $p = 0.2$ and H_1: $p < 0.2$. Joe wants to test H_0 against H_1, so he plays on the machine until he wins the first time. This happens at his 11:th attempt. Can he – with risk level at most 0.1 – reject H_0 in favour of H_1?

12.2 Eighteen pairs of iron rods have been buried in the soil, where in each pair one rod has been given anti corrosion treatment, whereas the other has not. The task is to figure out if the treatment has any real effect. The rods will be examined after some prescribed time and x, = the number of pairs where the treated rods have corroded more than the untreated, will be noted. It is decided that the null hypothesis H_0: "no effect" will be rejected in favour of H_1: "the treatment has some anti corrosion effect" if $x \leq 5$.

a) Decide the significance level of the test

b) What is the conclusion if $x = 1$? $x = 5$? $x = 15$?

12.3 We have 10 independent data from a N(μ, 0.2)-distribution:

3.9 4.1 4.4 4.0 3.8 4.0 3.9 4.3 4.2 4.4

a) Test, with a risk level 0.05, the hypothesis H_0: H_0: "$\mu = 4$" against the alternative H_1: "$\mu > 4$".

b) As a), but assume that the standard deviation (0.2) is unknown.

12.4 A researcher has made a new alloy and has theoretically calculated its melting temperature to 1050 centigrade. In order to test this result, she has measured the melting temperature for ten samples, and come up with the following results (centigrade:)

1054.8 1052.9 1051.0 1049.8 1051.6

1047.9 1051.8 1048.5 1050.2 1050.7

The variation in these numbers depends on imperfections of the thermometer employed. It is known that the errors are independent between measurements, unbiased and have a standard deviation of 2.3 centigrade

a) Test the hypothesis that the melting temperature is 1050 centigrade against the alternative that it is different from that. Use error level 5%.

b) The same as in a), but assume that the standard deviation is unknown.

12.5 The traffic police sets up a road check to test drivers for intoxication. The test is as a null hypothesis test: H_0: “the driver is sober” is tested against the alternative H_1: “the driver is intoxicated” with an error risk of 1%. What will happen in the long run? Is it true that

a) at most 1% of those acquitted are in fact intoxicated

b) at most 1% of those who are intoxicated are acquitted?

c) at most 1% of those who are caught are in fact sober.

13.1 In exercise 12.2, compute the power of the test against the alternative that $p = 0.25$, where p = probability that a treated rod will corrode more than an untreated.

13.2 In exercise 12.3, compute the power function for the test. What is the power if the true value of μ is $\mu = 3.8$? If $\mu = 4.3$?

13.3 in exercise 12.4, compute the power function of the test. What is the power if the true value is 1051 centigrade? If it is 1053 centigrade?

13.4 Eight persons measure their heights (cm) morning and evening. Result:

individual	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
morning	172	168	180	181	160	163	165	177
evening	172	167	177	179	159	161	166	175

The difference between morning and evening height is assumed to be from $N(\mu, \sigma)$. Test the hypothesis $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ with error risk 5%.

14.1 SIFO made opinion polls for the political parties in May and June 2016. In May, 498 individuals out of 1937 said they would vote for “Moderaterna” (M). In June, 531 individuals out of 1937 said they would vote for M. Test with error risk 5% if there has been a shift in the support for M between the two months

a) employing a confidence interval for the difference in proportions in favour of M.

b) employing a χ^2-test.

14.2 In a genetic experiment the offsprings were studied when two types of guinea pigs were interbred. Out of 87 offsprings 43 were red, 10 black and 34 white.

According to the genetic model, the probabilities for these colours are 9/16, 3/16 and 4/16. Should that hypothesis be rejected at 5% risk level?

14.3 During 81 days the number of cars that passed a certain point on a road in a ten minute period were counted. The result was:

#cars	0	1	2	3	4	5	6
#days	14	12	25	16	10	3	1

Test, with 5% risk level, if the number of cars passing can be regarded as coming from a Poisson distribution.

14.4 Out of each of three populations of humans, P_1, P_2 and P_3, a random sample was selected. One wanted to test if the gender distribution was the same in the three populations. The result was

	males	females
P_1	46	54
P_2	78	72
P_3	143	107

Test, at a 5% risk level, if the gender distribution is the same in the three populations.

14.5 From 500 accidents on a country road data have been collected according to this table:

injuries	safety belt	no safety belt
none or light	101	143
severe	58	198

Test, with error risk 5% if safety belts have an impact on the severity of injuries.