Real-World Problems for Secondary School Mathematics Students

Case Studies

Juergen Maasz and John O'Donoghue (Eds.)

SensePublishers

Real-World Problems for Secondary School Mathematics Students

Real-World Problems for Secondary School Mathematics Students

Case Studies

Edited by

Juergen Maasz
University of Linz, Austria

John O'Donoghue
University of Limerick, Ireland

SENSE PUBLISHERS
ROTTERDAM/BOSTON/TAIPEI

A C.I.P. record for this book is available from the Library of Congress.

ISBN: 978-94-6091-541-3 (paperback)
ISBN: 978-94-6091-542-0 (hardback)
ISBN: 978-94-6091-543-7 (e-book)

Published by: Sense Publishers,
P.O. Box 21858,
3001 AW Rotterdam,
The Netherlands
www.sensepublishers.com

Printed on acid-free paper

Image – The Living Bridge, University of Limerick.
© Patrick Johnson, 2008.

"The Living Bridge – An Droichead Beo"
The Living Bridge is the longest pedestrian bridge in Ireland and links both sides of the University of Limerick's campus across the river Shannon. The bridge is constructed of 6 equal spans and follows a 350 metre long curved alignment on a 300 metre radius.

TABLE OF CONTENTS

PREFACE

We should start by pointing out that this is not a mathematics text book – this is an ideas book. This is a book full of ideas for teaching real world problems to older students (15 years and older, Upper Secondary level). These contributions by no means exhaust all the possibilities for working with real world problems in mathematics classrooms but taken as a whole they do provide a rich resource for mathematics teachers that is readily available in a single volume. While many papers offer specific well worked out lesson type ideas, others concentrate on the teacher knowledge needed to introduce real world applications of mathematics into the classroom. We are confident that mathematics teachers who read the book will find a myriad of ways to introduce the material into their classrooms whether in ways suggested by the contributing authors or in their own ways, perhaps through mini-projects or extended projects or practical sessions or enquiry based learning. We are happy if they do!

Why did we collect and edit them for you, the mathematics teachers? In fact we did not collect them for you but rather for your students! They will enjoy working with them at school. Having fun learning mathematics is a novel idea for many students. Since many students do not enjoy mathematics at school, students often ask: "Why should we learn mathematics?" Solving real world problems is one (and not the only one!) good answer to this question. If your students enjoy learning mathematics by solving real world problems you will enjoy your job as a mathematics teacher more. So in a real sense the collection of examples in this book is for you too.

Using real world problems in mathematics classrooms places extra demands on teachers and students that need to be addressed. We need to consider at least two dimensions related to classroom teaching when we teach real world problems. One is the complexity (intensity or grade) of reality teachers think is appropriate to import into the classroom and the other is about the methods used to learn and work with real problems. Papers in this collection offer a practical perspective on each dimension, and more.

Solving real world problems often leads to a typical decision situation where you (we hope together with your students) will ask: Should we stop working on our problem now? Do we have enough information to solve the real world problem? These are not typical questions asked in mathematics lessons. What students should learn when they solve real world problems is that an exact calculation is not enough for a good solution. They should learn the whole process of modelling from the first step abstracting important information from the complex real world situation, to the next steps of the mathematical modelling process. For example, they should learn to write down equations to describe the situation; do calculations; interpret the results of calculation; improve the quality of the model; calculate again (several times if needed); and discuss the results with others. Last but not least, they should reflect on the solution process in order to learn for the future.

How real should real world problems be? More realistic problems are generally more complex and more complex problems demand more time to work them out. On the other hand a very simplified reality will not motivate students intrinsically to work for a solution (which is much better for a sustaining learning). Experience suggests starting with simple problems and simple open questions and moving to more complex problems. We think it is an impossible task for students without any experience of solving complex real problems to start by solving difficult real problems. It is better if you start with a simpler question and add complexity step by step.

The second dimension of classroom teaching is concerned with methods of teaching real world problems. We are convinced that learning and teaching is more successful if you use open methods like group work, project planning, enquiry learning, practical work, and reflection. A lot of real world problems have more than one correct solution, and may in fact have several that are good from different points of view. The different solutions need to be discussed and considered carefully and this is good for achieving general education aims like "Students should become critical citizens". Students are better prepared for life if they learn how to decide which solution is better in relation to the question and the people who are concerned.

Finally we would like to counter a typical "No, thank you" argument against teaching real world problems. Yes, you will need more time for this kind of teaching than you need for a typical lesson training students in mathematical skills and operations. Yes, you will need to prepare more intensively for these lessons and be prepared for lot of activity in your classroom. You will need to change your role from a typical teacher in the centre of the classroom knowing and telling everything to that of manager of the learning process who knows how to solve the problem. But you need help to get started! We hope you will use this book as your starter pack

We don't expect you to teach like this every day but only on occasions during the year. It should be one of your teaching approaches but not the only one. Try it and you will be happy because the results will be great for the students and for you!

ACKNOWLEDGEMENTS

We would like to thank all those who made this book possible especially the many authors who so generously contributed papers. This collaboration, sharing of insights, expertise and resources benefits all who engage in an enterprise such as this and offers potential benefits to many others who may have access to this volume. We are especially pleased to bring a wealth of material and expertise to an English speaking audience which might otherwise have remained unseen and untapped.

The editors would like also to record their thanks to their respective organizations who have supported this endeavour viz. the Institut fur Didaktik der Mathematik,

Johannes Kepler University, Linz, and the National Centre for Excellence in Mathematics and Science Teaching and Learning (NCE-MSTL), at the University of Limerick.

Juergen Maasz
University of Linz, Austria

John O'Donoghue
NCE-MSTL
University of Limerick and Linz
Autumn 2010

MANFRED BOROVCNIK AND RAMESH KAPADIA

1. MODELLING IN PROBABILITY AND STATISTICS

Key Ideas and Innovative Examples

This chapter explains why modelling in probability is a worthwhile goal to follow in teaching statistics. The approach will depend on the stage one aims at: secondary schools or introductory courses at university level in various applied disciplines which cover substantial content in probability and statistics as this field of mathematics is the key to understanding empirical research. It also depends on the depth to which one wants to explore the mathematical details. Such details may be handled more informally, supported by simulation of properties and animated visualizations to convey the concepts involved. In such a way, teaching can focus on the underlying *ideas* rather than technicalities and focus on applications.

There are various uses of probability. One is to model random phenomena. Such models have become more and more important as, for example, modern physicists build their theory completely on randomness; risk also occurs everywhere not only since the financial crisis of 2008. It is thus important to understand what probabilities really do mean and the assumptions behind the various distributions – the following sections deal with genuine probabilistic modelling.

Another use of probability is to prepare for statistical inference, which has become the standard method of generalising conclusions from limited data; the whole area of empirical research builds on a sound understanding of statistical conclusions going beyond the simple representation of data – sections 6 and 7 will cover ideas behind statistical inference and the role, probability plays therein.

We start with innovative examples of probabilistic modelling to whet the appetite of the reader. Several examples are analysed to illustrate the value of probabilistic models; the models are used to choose between several actions to improve the situation according to a goal criterion (eg., reduce cost). Part of this modelling approach is to search for crucial parameters, which strongly influence the result.

We then explain the usual approach towards probability – and the sparse role that modelling plays therein by typical examples, ending with a famous and rather controversial example which led to some heated exchanges between professionals. Indeed we look at this example (Nowitzki) in some depth as a leitmotiv for the whole chapter: readers may wish to focus on some aspects or omit sections which deal with technical details such as the complete solution.

In the third and fourth sections, basic properties of Bernoulli experiments are discussed in order to model and solve the Nowitzki task from the context of sports.

J. Maasz and J. O'Donoghue (eds.), Real-World Problems for Secondary School Mathematics Students: Case Studies, 1–43.

The approach uses fundamental properties of the models, which are not always highlighted as they should be in teaching probability.

In the fifth section, the fundamental underlying ideas for a number of probability distributions are developed; this stresses the crucial assumptions for any situation, in which the distribution might be applied. A key property is discussed for some important distributions: waiting times, for example, may or may not be dependent on time already spent waiting. If independent, this sheds a special light on the phenomenon, which is to be modelled. In a modelling approach, more concepts than usual have to be developed (with the focus on *informal* mathematical treatment) but the effort is worthwhile as these concepts allow students to gain a more direct insight to understand the inherent assumptions, which are required from the situation to be modelled.

In the sixth section, the statistical question – (is Nowitzki weaker in away than in home matches?) – is dealt with thoroughly. This gives rise to various ways to tackle this question within an inferential framework. We deal informally with the methods that comprise much of what students should know about the statistical comparison of two groups, which forms the core of any introductory course at university for all fields, in which data is used to enrich research.

While the assumptions should be checked in any case of application, such a crucial test for the assumptions might be difficult. It will be argued that the perspective of probabilistic and statistical applications is different and linking heuristic arguments might be more attractive in the case of statistical inference.

Probabilistic distributions are used to make a probability statement about an event, or to calculate expected values to make a decision between different options. Or, they may be used to describe the ‘internal structure’ of a situation by the model’s inherent structure and assumptions. Statistical applications focus on generalizing facts beyond available data. For that purpose they interpret the given data by probabilistic models. A typical question is whether the data is compatible with a specific hypothesized model. This model as well as the answer is interpreted within the context. For example, can we assume that a medical treatment is – according to some rules – better than a placebo treatment?

The final two sections resume the discussion of teaching probability and statistics, some inherent difficulties, and the significance of modelling. Conclusions are drawn and pedagogical suggestions are made.

INNOVATIVE EXAMPLES OF PROBABILISTIC MODELLING

As we shall see in a later section, a key assumption of independence is not justified in many exercises set in probability. Indeed the key question in any modelling is the extent to which underlying assumptions are or are not justified. Rather than the usual approach of mechanistic applications of probability, a more realistic picture of potential applications will be developed in this section by some selected innovative examples. They describe a situation from the ‘real world’ and state a target to improve or optimize. A spectrum of actions or interventions is open for use to improve a criterion such as reduction of costs. A probability distribution is chosen to model the situation, even though the inherent assumptions might not be perfectly fulfilled.

A solution is derived and analysed: how does it change due to changes in parameters involved in the model, how does it change due to violations of assumptions? Sensitive parameters are identified; this approach offers ways of making the best use of the derived solutions and corroborating the best actions to initiate further investigations to improve the available information. The examples deal with novel applications – blood samples, twin light bulbs, telephone call times, and spam mail. Simulation, spreadsheets and other methods are used and illustrate the wide range of ideas where probability helps to model and understand a wide and diverse range of situations.

Blood Samples Modelled with Binomial Probabilities

The following example uses the binomial model for answering a typical question, which might be studied. The 'outcome' might be improved even if the model has some drawbacks: the cost situation is improved and hints for action are drawn from the model used though the actual cost improvement cannot be directly read off the solution. Crucial parameters that strongly influence the solution are identified, for which one may strive to get more information in the future.

Example 1. Donations of blood have to be examined as to whether they are suitable for further processing or not. This is done in a special laboratory after the simple determination of the blood groups. Each donation is judged – independently of each other – 'contaminated' with a probability $p = 0.1$ and suitable with the complementary probability of $q = 0.9$.

a. Determine the distribution of the number of non-suitable donations if 3 are drawn at random.
b. 3 units of different donations are taken and mixed. Only then are they examined jointly as to whether they are suitable or not. If one of those mixed was already contaminated then the others will be contaminated and become useless. One unit has a value of € 50.-. Determine the loss for the various numbers of non-suitable units among those which are drawn and mixed.
   Remember: If exactly one is non-suitable then the two others are 'destroyed'.
c. Determine the distribution of the loss as a random variable.
d. Calculate the expected loss if 3 units of blood are mixed in a 'pool'.
e. Testing of blood for suitability costs of € 25.- per unit tested; the price is independent of the quantity tested. By mixing 3 units in a pool, a sum of € 50.- is saved. With the expected loss from d., does it pay to mix 3 different units, or should one apply the test for suitability separately to each of the blood units?

A solution to this example is presented in a spreadsheet (Figure 1); the criterion for decision is based on the comparison of potential loss and benefit by pooling; pooling is favourable if and only if:

$$\text{Expected loss by pooling} < \text{Reduction of cost of lab testing} \qquad (1)$$

All blood donations are assumed to be contaminated with a probability of 0.1, *independently* of each other. That means we model the selection of blood donations, which should be combined in a pool and analysed in the laboratory jointly, as if it were ‘coin tossing with a success probability of $p = 0.1$’. While the benefit is a fixed number, loss has to be modelled by expected values. Comparison (1) yields 25.65 < 50, hence mixing 3 blood units and testing jointly, saves costs.

Modelling of pooling and calculation of expected cost

	0 / 1		Coding: Single unit is suitable = 0, is NOT suitable = 1
	n	3	Number of blood units to combine to a pool
	1 - p	0.9	Probability that a single unit is suitable for further processing
	p	0.1	Probability that a single unit is NOT suitable for further processing
a.	X ~ Bin(n, p)		Model for number X of NOT suitable single units in the pool: Draw with replacement
b.	Y		Loss as a function of the number of not suitable units in the pool
		50	Cost of units destroyed by pool = loss
c.	Distribution of Y		P(Y = y_i) - has to be determined
d.	E(Y)		Average cost of units destroyed by pooling = loss
e.	Comparison		To compare reduction of cost in testing pooled units and expected loss by pooling
		25	Cost of testing a unit in the laboratory

a.		**b.**		**c.**		**d.**	**e.**	
Not suitable	Probability	Destroyed	Loss Y	Distribution of loss		Expected loss	Cost of testing	
X = i	P(X = i)	by pool		Y = y_i	P(Y = y_i)	$y_i \cdot p_i$	Single	25
0	0.729	0	0	0	0.730	0	Size of pool	3
1	0.243	2	100	50	0.027	1.35	separate	75
2	0.027	1	50	100	0.243	24.3	pooled	25
3	0.001	0	0			25.65	Reduction	50

These figures have to be compared.

Figure 1. Spreadsheet with a solution to Example 1.

Example 2. Explore the parameters in Example 1.

a. For a probability of $p = 0.1$ of being contaminated what is a sensible recommendation for the number of blood units being mixed?
b. How does the recommendation in a. depend on the probability of suitable blood donations?

While a spreadsheet is effective in assisting to solve Example 1, it becomes vital for tackling the open questions in Example 2. Such an analysis of input parameters gives much more rationale to the questions posed. Of course, there are doubts on whether the suitability of blood donations might be modelled by a Bernoulli process. Even if so, how is one to get trustworthy information about the ‘success’ probability for suitable blood units? As this is a crucial input, it should be varied and the consequences studied. Also, just to mix a fixed number of blood units gives no clue why such a size of a pool should be chosen. It also gives no feeling about how the size makes sense relative to the expected monetary gain or loss connected to the size of a pool, which is examined jointly. In a spreadsheet, a slide control for the input parameters p and size n is easily constructed and allows an effective interactive investigation of the consequences (or the laboratory cost and the value of a unit). From a spreadsheet as in Figure 2, one may read off ($q = 1 - p$):

$$\text{Expected net saving } (q = 0.9, n = 4) = 26.22 \tag{2}$$

This yields an even better expected net saving as in (1). Interactively changing the size n of the pool shows that $n = 4$ yields the highest expected value in (2) so that to combine 4 is 'optimal' if the proportion of suitable blood units is as high as $q = 0.9$. With a smaller probability of good blood units, the best size of the pool drastically declines; with $q = 0.8$, size 2 is the most favourable in cost reduction; the cost reduction is as low as 9 € per pool of 2 as compared to 26.22 € per each 4 units combined to a pool. Reduction of cost per unit is 4.5 with $q = 0.8$ and 6.6 € with $q = 0.9$. There is much more to explore, which will be left to the reader.

Modelling of pooling - exploring the effect of parameter changes

Lab cost of testing	Value of a unit	Proportion q of suitable units	Size of pool n
25	50	0.90	4

k suitable	P(k)	reduction of test cost by pooling	cost of destroyed units by pooling	net saving	$x_i \cdot p_i$	Expected net saving
0	0.0001	75	0	75	0.0075	
1	0.0036	75	-50	25	0.0900	
2	0.0486	75	-100	-25	-1.2150	
3	0.2916	75	-150	-75	-21.8700	
4	0.6561	75	0	75	49.2075	
5	#ZAHL!	75	0	75	0.0000	lines
10	#ZAHL!	75	0	75	0.0000	hidden
						26.22

Figure 2. Spreadsheet to Example 2 – with slide controls for an interactive search.

It is merely an assumption that the units are independent and have the same probability of being contaminated. Nevertheless, the crucial parameter is still the size of such a probability as other measurements (of success) are highly sensitive to it. If it were a bit higher than given in the details of the example, pooling would not lead to decreasing cost of testing. If it were considerably smaller, then pools of even bigger size than suggested would lead to considerable saving of money. The monotonic influence becomes clear even if the exact size of decreasing cost cannot be given. Closer monitoring of the quality of the incoming blood donations is wise to clarify circumstances under which the required assumptions are less justified than usual.

Lifetime of Bulbs Modelled with Normal Distribution

Example 3. Bulbs are used in a tunnel to brighten it and increase the security of traffic; the bulbs have to be replaced from time to time. The first question is when a single bulb has to be replaced. The second is, whether it is possible to find a time when all bulbs may be replaced jointly. The reason for such an action is that one has to send a special repair team into the tunnel and block traffic for the time of replacement. While cost maybe reduced by a complete replacement, the lifetime of the still functioning lamps is lost. The time for replacement is, however, mainly determined by security arguments: with what percentage of bulbs still working is the tunnel still sufficiently lit? Here we assume that the tunnel is no longer secure if over two thirds of the bulbs have failed.

Two systems are compared for their relative costs: Single bulbs and twin bulbs, which consist of two single bulbs – the second is switched on when the first fails. Which system is less costly to install? Lifetime of bulbs in hours is modelled by a normal distribution with mean lifetime μ=1900 and standard deviation σ=200.

a. What is the probability that a single bulb fails within the 2300 hours in service?
b. Determine the time when two thirds of single bulbs have failed.
c. Determine an adequate model for the lifetime of twin bulbs and – based upon it – the time when two thirds of the twin bulbs have failed.
   Remark. If independence of failure is assumed then lifetime of twin bulbs is also normally distributed with parameters: mean = sum of single means, variance = sum of single variances.
d. Assume that at least one third of the lamps in the system have to function for security reasons. The cost of one twin bulb is 2.5 €, a single lamp costs 1 €. The cost of replacing all lamps in the tunnel is 1000 €. For the whole tunnel, 2000 units have to be used to light it sufficiently at the beginning. Relative to such conditions, is it cheaper to install single or twin bulbs?

The solution can be read-off from a spreadsheet like Figure 3. The probability in a. to fail before 2300 hours may be found by standardizing this value by the parameters to (2300–1900)/200 = 2 and calculating the value of the standard normal $\Phi(2)$. Parts b. and c. require us to calculate the ⅔ quantile of the normal distribution, which traditionally is done by using probability tables, or which may be directly solved by a standard function. As a feature of spreadsheets, such a quantile may be determined by a sliding control, which allows us to control x in the formula $P(X \leq x) =$ ∞ until the probability ⅔ is reached in the box.

For twin bulbs, first the parameters have to be calculated: $\mu_T = 1900 + 1900$ and $\sigma_T = \sqrt{(200^2+200^2)}$. Such a relation between parameters of single and twin bulbs may be corroborated by simulation – a proof is quite complex.

For the final comparison of costs in d., the calculation yields that single bulbs will be exchanged after 1986.1, twin bulbs after 3921.8 hours. The overall costs of the two systems have to be related to unit time.

Of course, the lifetime of the bulbs is not really normally distributed. The estimation of the parameters (mean and standard deviation of lifetime) is surrounded with imprecision. As a perfect model, normal distribution might not serve. However, perceived as a scenario in order to investigate "what happens if …?" it helps to analyse the situation and detect crucial parameters. Thus, it might yield clues for the decision on which system to install. On the basis of such a scenario, we have a clear goal, namely to compare the expected costs per unit time E[C] /h of the two systems (see Figure 3):

$$\text{E}[C]\,/h \text{ (single bulbs)} - \text{E}[C]\,/h \text{ (twin bulbs)} = 1.51 - 1.53 < 0 \qquad (3)$$

The comparison of expected costs gives 1.51 € per unit time for single as opposed to 1.53 € for twin bulbs. A crucial component is the price of twin bulbs: a reduction from 2.5 to 2.4 € per twin bulb (which amounts to a reduction of 4%)

changes the decision in favour of twin bulbs. Hence this approach allows students to investigate assumptions, the actual situation and relative costs. It encourages them to use their own prior knowledge in the situation by varying parameters and costs. This gives practice in applying the normal distribution and then investigating consequences from the situation, some of which may be hard to quantify.

Single bulbs

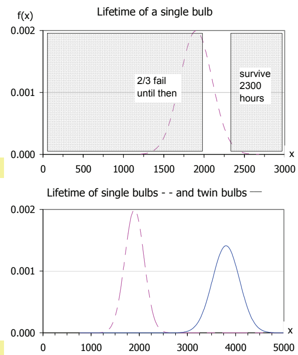
μ	σ	x	$z = (x-\mu)/\sigma$	$\Phi(z)$	
1900	200	2300	2	0.9772	a.

p	z_p	$x_p = \mu + z_p \cdot \sigma$	
0.6667	0.4307	1986.1	b.

Twin bulbs

μ_T	σ_T	x	$z = (x-\mu)/\sigma$	$\Phi(z)$
3800	282.84	4600	2.83	0.9977

p	z_p	$x_p = \mu + z_p \cdot \sigma$	
0.6667	0.4307	3921.8	c.



Comparison of cost per unit time

	Devices		Cost			Time		
	€ per unit	Number	Light system	Exchange	total	working	Cost per unit time	
Single	1.0	2000	2000	1000	3000	1986.1	1.51	} d.
Twin bulbs	2.5	2000	5000	1000	6000	3921.8	1.53	
Variant	2.4	2000	4800	1000	5800	3921.8	1.48	

Figure 3. Spreadsheet with a solution to Example 3.

In fact, there is a systematic error in the normal distribution as for a lifetime no negative values are possible. However, for the model used, the probability for values less than 0 amounts to 10^{-21}, which is negligible for all practical purposes.

We will illustrate key ideas behind distributions in section 5; hazard rate is one of them, which does not belong to the standard repertoire of concepts in probability at lower levels. Hazard is a tendency to fail. However, it is different from propensity (to fail), which relates the tendency to fail for *new* items only. To look at an analogy: it is folklore that the tendency to fail (die) for human beings differs with age. The older we get, the higher the *age-related* tendency to die.

Again, such a relation is complex to prove. However, a simulation study should be analysed accordingly and the percentage of failures can be calculated in two ways: one based on all items in the experiment, the other based only on those, which are still in function at a specific point of time. This enhances the idea of an age-related (conditional) risk to fail as compared to a general risk to fail, which implicitly is related to all items, which are exposed to failure in the experiment.

The normal distribution may be shown to follow an *increasing* hazard. Such an increasing risk to fail with increasing age is reasonable with lifetime of bulbs – as engineers know. Thus, even if the normal model may not be interpreted directly by relative frequencies of failures, it may count as a reasonable model for the situation. The exact amount of net saving by installing the optimal system cannot be quantified but the size of it may well be read from the scenario. These are the sort of modelling issues to discuss with students in order to enhance their understanding and interest.

Call Times and Cost – the Exponential and Poisson Distributions Combined

Example 4. The length Y of a telephone call is a random variable and can be modelled as exponentially distributed with parameter $\lambda = 0.5$, which corresponds to a mean duration of a call of $1/\lambda = 2$. The cost k of a call is a function of its duration y and is given by a fixed amount of 10 for $y \leq 5$ and amounts to $2y$ for $y > 5$.

a. Determine the expected cost of a telephone call.
b. Calculate the standard deviation (s.d.) of the cost of a telephone call.
c. Determine an interval for the cost of a single call with probability of 99%.
   The number of calls during a month will be modelled by a Poisson distribution with $\mu = 100$; this coincides with an expected count of calls of 100. Determine
d. the probability that the number of phone calls does not exceed 130 per month;
e. a reasonable upper constraint for the maximum cost per month and a reasonable number depicting the risk that such a constraint does not hold – describe how you would proceed to determine such a limit more accurately.

Cost of single calls and bills of 130 calls - a simulation study

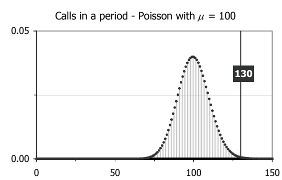
λ = 0.5

Nr.	Random numbers from (0, 1)	Exponential - length of call	Cost of call	Length of 130 calls	Bill for period with 130 calls	empty column!	Many bills simulated *	Bill > 1387.5 ?	* data tables in EXCEL
	Z_i	X_i	$c(X_i)$				1329.84	0	
1	0.525	1.487	10.00	240.69	1329.84		1329.86	0	
2	0.233	0.530	10.00				1325.67	0	lines
713	0.419	1.086	10.00				1337.23	0	hidden
714	0.202	0.452	10.00				1307.75	0	

a.	Single calls	E = Expected cost of single call			10.25	mean	} of simulated data on cost
b.		σ = S.d.			1.41	s.d.	
		Risk is	0.010	that single call cost more than:	18.25	cost	} 99% quantile of simulated data
c.		Risk is	0.010	that length of call longer than:	9.13	length	
e.	Bill	Risk is	0.020	that bill (with 130 calls) costs more:	1387.50	percentage rank in simulated bills	
		Risk is	0.010	that bill for the period is higher than	1393.78	99% quantile of simulated bills.	

The simulation study is based on only 714 random numbers.
It fluctuates still, but would stabilize if more data were generated.

d. Risk that number of calls > 130: 0.002 Risk seems 'negligible'



While for the single calls the cost cannot be approximated by a normal (as exponential distributions are heavily skewed), the cost of 130 calls (as a sum) may be approximated by the normal distribution, though it still is slightly skewed (CLT!)

	single call	**	130 calls	130
expected cost	E	10.3283	$E_n = n \cdot E$	1342.68
s.d.	σ	1.5871	$\sigma_n = \sqrt{n} \cdot \sigma$	18.10

** tricky integrals or simulation

	p quantile	0.99	
standard normal	z_p	2.326	
e. cost of 130 calls	$x_p = E_{130} + z_p \cdot \sigma_{130}$	1384.78	99% of 130 calls cost less

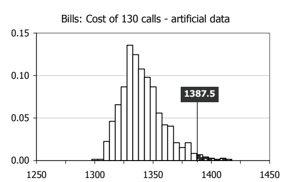


Figure 4. Spreadsheet with a solution to Example 4.

The models suggested are quite reasonable. However, the analytic difficulties are considerable – even at university level. A solution to part e. may be found from a simulation scenario of the assumptions. The message of such examples is that not all models can be handled rigorously. The key idea here is to understand the assumptions implicit in the model and judge whether they form a plausible framework for the real situation to be modelled and interpret the result accordingly.

In section 5, the model of the exponential distribution will be related to the idea of 'pure' random length and to the memory-less property according to which the further duration of a call has the same distribution regardless how long it already goes on. If such a condition is rejected for the phoning behaviour of a person the result derived here would not be relevant. The number of phone calls is modelled here by a Poisson distribution. The key idea behind this model is that events (phone calls) occur completely randomly in time with a specific rate per unit time; see section 5 for more details. The simulation is based on the following mathematical relation between distributions (which itself may be investigated by simulation):

If Z is a random number in the interval (0, 1) then
$Y = -\ln(1-Z)/\lambda$ is exponentially distributed with parameter λ (4)

A further key idea used in this example is the normal approximation for the cost of the bill of 130 calls for a period as it is the sum of the 'independent' single call cost. The histogram of simulated bills 'confirms' that to some extent but shows still some skewness, which originates from the highly skewed exponential distribution. To apply the approximation, mean and s.d. of the sum of independent and identically distributed (called *iid*, a property analogue to (9) later) variables has to be known. An estimate of single-call cost may again be gained from a simulation study as the integrals here are not easy to solve. From single calls X_i to bills of n calls, the following mathematical relation is needed:

$$X_i \overset{iid}{\sim} X\,,\; T_n := X_1 + \ldots + X_n\,, \text{ then } E(T_n) = n \cdot E(X)\,,\; \sigma(T_n) = \sqrt{n} \cdot \sigma(X)\,. \quad (5)$$

The first part is intuitive – if $E(X)$ is interpreted as fair prize of a game, then T_n describes the win of n repetitions of it; the fair prize of it should be n times the prize of a single game. The second part is harder and has to be investigated further. As it is fundamental, a simulation study can be used to clarify matters.

Spam Mail – Revising Probabilities with Bayes' Formula

Example 5. Mails may be 'ham' or 'spam'. To recognize this and build a filter into the mail system, one might scan for words that are contained in hams and in spams; e.g., 30% of all spams contain the word 'free', which occurs in hams with a frequency of only 1%. Such a word might therefore be used to discriminate between ham and spam mails. Assume that the basic frequency of spams in a specific mailbox is 10 (30)%

a. If a mail arrives with the word ‘free’ in it, what is the probability that it is spam?
b. If a message passes such a mail filter, what is its probability that it is actually spam?
c. Suggest improvements of such a simple spam filter.

The easiest way to find the solution is to re-read all the probabilities as expected frequencies of a suitable number of mails. If we base our thought on 1000 mails, we expect 100 (300) to be spam. Of these, we expect 30 (90) to contain the word ‘free’. The imaginary data – natural frequencies in the jargon of Gigerenzer (2002) are in Table 1, from which it is easy to derive answers to the questions:

If the mail contains ‘free’, its conditional probability to be spam is 0.7692 (30/39 with 10% spam overall) and 0.9278 (90/97 with 30% spam). The filter, however, has limited power to discriminate between spam and ham as a mail, which does not contain the word ‘free’ still has a probability to be spam of 0.0728 (70/961) or 0.2326 (210/903) depending on the overall rate of spam mails.

Table 1. Probabilities derived from fictional numbers – natural frequencies

Spam overall		*10%*			*30%*	
	‘free’	~~*‘free’*~~	*all*	*‘free’*	~~*‘free’*~~	*all*
Ham	9	891	900	7	693	700
Spam	30	70	100	90	210	300
all	39	961	1000	97	903	1000

This shows that the filter is not suitable for practical use. Furthermore, the conditional probabilities are different for each person. The direction for further investigation seems clear: to find more words that separate between ham and spam mails and let the filter learn from the user who classifies several mails into these categories. The table with natural (expected) frequencies – sometimes also called the statistical village behaviour – delivers the same probabilities as the Bayesian formula but is easier to understand and is accepted much better by laypeople.

For further development, the inherent key concept of conditional probability should be made explicit. Updating of probabilities according to new incoming information is a basic activity. It has a wide-spread range of application such as in medical diagnoses or before court when indicative knowledge has to be evaluated. The Bayesian formula to solve this example is

$$P(spam\,|\,'free') = \frac{P(spam)\cdot P('free'\,|\,spam)}{P(spam)\cdot P('free'\,|\,spam) + P(ham)\cdot P('free'\,|\,ham)} \qquad (6)$$

Gigerenzer (2002) gives several examples how badly that formula is used by practitioners who apply it quite frequently (not knowing the details, of course). There are many issues to clarify, amongst them the tendency to confuse $P(spam|'free')$ and $P('free'|spam)$. Starting with two-way tables is advisable, as advocated by Gigerenzer (2002); later the structure of the process should be made explicit to promote the idea of continuously updating knowledge by new evidence.

Summary

Overall, these four examples show the interplay between modelling and probability distributions, with simplifying assumptions made but then analysed in order for students to develop deeper understanding of the role, probability may play to make decisions more transparent. Probability models are not only part of glass bead games but may partially model situations from reality and contribute to compare alternatives for action rationally. In each case, there is an interesting and key question to explore within a real-life context. Simplifying assumptions are made in order to apply a probability distribution. The results are then explored in order to check the validity of the assumptions and find out the sensitivity of the parameters used. We have only given an outline of the process in each example; a longer time would be needed with students in the classroom or lecture hall.

THE USUAL APPROACH TOWARDS TEACHING PROBABILITY

The usual approach towards probability is dominated by games of chance, which per se is not wrong as the concepts stem from such a context, or from early insurance agreements, which are essentially loaded games of chance with the notable exception of symmetry arguments; probabilities that would otherwise arise from symmetry considerations in games are replaced by estimates for the basic (subjective) probabilities.

The axiomatic rules of probabilities are usually discussed cursorily and merely used as basic rules to obey when dealing with probabilities. Thus, no detailed proofs of simple properties are done and if done are simplified and backed up by didactical methods like tree diagrams, which may be applied as one deals primarily with finite or countably infinite probability spaces. The link from an axiomatically 'determined' probability to distributions is not really established[1] – so the many probability distributions develop their own 'life'. Part of their complexity arises from their multitude. Normally, only a few are dealt with in teaching 'paradigmatically'. At the secondary stage this is mainly the binomial and normal distributions; in introductory courses at universities hypergeometric, Poisson, or exponential distributions are also added.

The various distributions are dealt with rather mechanistically. A short description of an artificial problem is followed by the question for the probability of an event like 'the number of successes (in a binomial situation) does not exceed a specified value'. Hereby, the context plays a shallow role. Questions of modelling, e.g., what assumptions are necessary to apply the distribution in question, or, in what respect are such requirements fulfilled in the context, are rarely addressed. The following examples illustrate the usual 'approach'. Example 6 illustrates an attitude towards modelling, which is not rare.

Example 6. For the production of specific screws, it is known that 'on average' 5% are defective. If 10 screws are packed in a box, what is the probability that one finds two (or, not more than two) defective pieces in a box?

The screws could be electric bulbs, or electronic devices, etc; defective may be defined as 'lasting less than 2000 hours'. The silent assumption in all these examples is: 'model' the selected items as a random sample from the production. Sometimes the model is embodied by the paradigmatic situation of random selection from an urn with two sorts of marbles – marked 1 and 0 – predominantly with (sometimes without) replacement of the drawn marbles. The context is used as a fig leaf to tell the different stories for very similar tasks, namely to drill skills in calculating probabilities from the right distribution. Neither a true question to solve, nor alternative models, nor a discussion of validity of assumptions is involved. No clear view is given of why probabilistic modelling helps to improve one's understanding of the context. The full potential of probability to serve as a means of modelling is missed by such an attitude; see Borovcnik (2011).

Modelling from a Sporting Context – the Nowitzki Task

The following example shows that such restricted views on probability modelling are not bound to single teachers, textbooks, or researchers in educational statistics. The example is taken from a centrally organized final exam in a federal state in Germany but could be taken from anywhere else. The required assumptions to solve the problems posed are clarified. This is – at the same time – a fundamental topic in modelling. The question, as we shall see later, aroused fierce controversy.

Example 7 (Nowitzki task). The German professional basketball player Dirk Nowitzki plays in the American professional league NBC. In the season 2006–07 he achieves a success rate of 90.4% in free throws. (For the original task, which was administered in 2008, see Schulministerium NRW, n.d.[2])

Probabilistic part. Calculate the probability that he
a. scores exactly 8 points with 10 trials;
b. scores at the most 8 points with 10 trials;
c. is successful in free throws at the most 4 times in a series.

Statistical part. In home matches he scored 267 points with 288 free throws, in away matches the success rate was 231/263. A sports reporter [claimed that] Nowitzki has a considerably lower success rate away. At a significance level of 5%, analyse whether the number of scores in free throws away
a. lies significantly below the 'expected value' for home and away matches;
b. lies significantly below the 'expected value'[3] for home matches.

This example will be referred to as the Nowitzki task and will be used extensively below to illustrate the various modelling aspects both in probability and statistics. From the discussion it will become clearer what assumptions we have to rely on and how these are 'fulfilled' differently in probabilistic and statistical applications.

MODELLING THE NOWITZKI TASK

The Nowitzki task (Example 7) has a special history in Germany as it was an item for a centrally administered exam. The public debate provoked a harsh critique: its probabilistic part was criticized as unsolvable in the form it was posed; its statistical part was disputed as difficult and the ministerial solution was 'attacked' for blurring the fundamental difference between 'true' values of parameters of models and estimates thereof. The probabilistic task shows essentially the same features as Example 6; the context could be taken from anywhere, it is arbitrary. The statistical part, however, allows for more discussion on the fundamental problem of empirical research which has to deal with generalizing results from samples to populations. Various models are compared with respect to their quality to model the situation. Here we focus on modelling and then solving the probability part.

Basic Assumptions of Bernoulli Processes

This task is designed to be an exemplar of an application of the binomial distribution. It is worthwhile to repeat the basic features of the model involved. This distribution allows one to model experiments, with a fixed number of trials with two outcomes for each repetition of the experiment (trial); one is typically named 'success' and the other 'failure'. The basic assumption is that

– the probability p of success is the same for all trials, and
– single trials do not 'influence' each other, which means – probabilistically speaking – the trials are independent.

Such assumptions are usually subsumed under the name of Bernoulli experiments (Bernoulli process of experiments). If the results of the single trials are denoted by random variables $X_1, ..., X_n$ with

$$X_i = 1 \text{ (success) or } X_i = 0 \text{ (failure)} \tag{7}$$

the random variables have to be independent, ie.,

$$P(X_i = x_i, X_j = x_j) = P(X_i = x_i)P(X_j = x_j) \text{ if } i \neq j\,. \tag{8}$$

Such a product rule holds also for more than two variables X_i of the process. The assumption usually is denoted by:

$$X_i \overset{iid}{\sim} X \sim B(1, p)\,, \tag{9}$$

where the '*iid*' refers to *i*ndependent, *i*dentically *d*istributed random variables X_i. The model, however, is not uniquely determined by such a Bernoulli process. Still missing is information about the value of p.

From the perspective of modelling, the decision on which distribution to apply is only one step towards a model for the situation in question. Usually, the model consists of a *family* of distributions, which differ by the values of one (or more) parameters. The next step is to find values for the parameters to fix one of the distributions as the model to use for further consideration. How to support the modeller to choose such a family like the binomial or normal distributions is discussed from a modelling perspective in section 5. The step of modelling to get values for the parameters of an already chosen model is outlined here.

The process to get numbers for the parameters (p here) is quite similar for all models. In this section matters are discussed for Bernoulli processes. Some features arise from the possibility to model the data to come from a finite or an infinite population. The parameter p will be called the 'strength' of Nowitzki.

The Nowitzki task was meant to go beyond an application within games of chance. The probability of success of a single trial of Nowitzki is not determined by a combinatorial consideration (justified by the symmetry of an experiment). The key idea to get numbers for the parameters is to incorporate some 'knowledge' or information. The reader is reminded that the problem involves 10 trials and the task will be treated as an open task to explore rather than just an examination question. So one might form a hypothesis on the success rate subjectively, for example.

Model 1. The probability p could be determined by a hypothesis like Nowitzki is equally good as in the last two years when his success rate has been (e.g.) 0.910. Supposed that such a value holds also for the present season, the model is fixed. This value could well be a mere fiction – just to 'develop a scenario' and determine what would be its consequences. On the basis of this model, the random variable

$$T_n = X_1 + ... + X_n \sim B(n = 10, p = 0.910), \tag{10}$$

ie., the (overall) number of successes follows a binomial distribution with parameters 10 and 0.910.

Model 2. The probability p is *estimated* from the given data, i.e., $\hat{p} = 0.904$. In this case, the further calculations usually are based on

$$T_n \sim B(n = 10, p = 0.904). \tag{11}$$

This is somehow the 'best' model as the estimation procedure leading to $\hat{p} = 0.904$ fulfils certain optimality criteria like unbiasedness, minimum variance, efficiency, and asymptotic normality.

Model 2 (variant). The approach in model 2 misses the fact that the estimate $\hat{p}$ is not identical to the 'true'[4] value of p. There is some inaccuracy attached to the estimation of p. Thus, it might be better to derive a (95%) confidence interval for the unknown parameter p in a first approach leading to an interval

$$[p_L, p_U] = [0.8792, 0.9284] \tag{12}$$

and only then calculate the required probabilities in Example 7 with the worst case p_L and the best case p_U. This procedure leads to a confidence interval for the required probability reflecting the inaccuracy of the estimate $\hat{p}$.

Model 3. The value of p is *equated* (not estimated) to the success rate of the whole season, i.e., p:=0.904. This leads to the same model as in (11) but with a complete different connotation. The probability of success in a free throw may be – after the end of the season – viewed as factually known as the number of successes (498) divided by the number of trials (551). There is no need to view it any longer as unknown and treat the success rate of 0.904 as an estimate $\hat{p}$ of an unknown p, which results from an imagined infinite series of Bernoulli trials.

Investigating and Modelling the Unknown Value of p

Three different ways may be pursued to provide the required information for the unknown parameter. With the binomial distribution, one needs to have information about the value of p, which may be interpreted as success rate in the Bernoulli process in the background. The information to fix the parameter has different connotation as described here. In due consequence, the models applied inherit part of this meaning. The cases to differentiate are:

i. p is known
ii. p is estimated from data
iii. p is hypothesized from similar situations

With games of chance, symmetries of the involved random experiment allow one to derive a value for p; eg., ½ for head in tossing a 'fair' coin – case *i*. Most applications, however, lack such considerations and one has to evoke *ii*. or *iii*.

i. The assumption of equiprobability for all possible cases (of which one is called 'success') is – beyond games of chance – sometimes more a way of putting it, just to fix a model to work with. For coin tossing, this paves the way to avoid tedious data production (actually tossing the coin quite often) and work with such a model to derive some consequences on the basis 'what would be if we suppose the coin to be fair (symmetric)'. Normally, for coins, a closer scrutiny would not deviate too much from the presupposition of p = ½ and yield quite similar results.

With the basketball task, the value of p could be known from the past (model 1), which also refers to the further assumption that 'things remain the same'[5], i.e., there was no change from past to present. This is an assumption for which usually a substantial justification is lacking – those who do not rely too heavily on the information of the past might react more intelligently in the present.[6]

ii. Closer to the present situation in Example 7 is to use the latest data available (from the present season). The disadvantage of such a procedure is that the difference between the 'true' value of p and an estimate $\hat{p}$ might be blurred, and thereby forgetting that the estimate is inaccurate. One possibility to deal with this is the variant of model 2. The inaccuracy of estimates is best illustrated by confidence intervals. To vary the size of underlying samples where the data stem from, gives a clearer picture of the influence of this lack of information.

An important assumption for the data base from which the estimate is calculated, is: it has to be a random sample of the underlying Bernoulli process, which is essentially the same as the parent Bernoulli process. Clearly, the assumption of a sample to be random is rarely fulfilled and often is beyond scrutiny. Usually there are qualitative arguments to back up such an assumption. It is to be noted that the estimation of the probability is interwoven with the two key assumptions of Bernoulli experiments – the same success probability, and occurring independently in all trials. Otherwise, probabilities, such as a probability of success with a free throw, have no meaning.

iii. A hypothesis about the success probability could be corroborated by knowledge about the past as in *i*. However, the season is completed and as a matter of fact, the success rate of Nowitzki was 0.904. To apply the factual knowledge about all games yields a value for the unknown parameter as

$$p := 0.904, \tag{13}$$

which amounts to much more concrete information than the estimation procedure leading to $p \approx \hat{p} = 0.904$. The success probability in (13) may be justified and clarified by the following: As the season is completed, one knows *all* data. There will be no more. A new season will have irretrievably different constellations. The success rate in 2006–07 *is* – as a matter of fact – 0.904. There could well be the question as to how to interpret this number and whether it is possible to interpret it as a success *probability*.

The key question is whether it makes sense to interpret this 0.904 as a success probability. This interpretation is bound to the assumption that the data stem from an independent repetition of the same Bernoulli experiment. This requires – taken literally – that for each free throw of Nowitzki the conditions have been exactly the same, independently of each other and independent of the actual score and the previous course of the match, etc. With this point of view one might question whether the data really are compatible with the pre-requisites of a Bernoulli process. One could, e.g., inspect the number of 'runs' (points or failures in a series) and evaluate whether they are above or below the expected value for a Bernoulli process or not in order to check for the plausibility of its underlying assumptions.

The way in which information is used to get numerical values for the unknown parameter influences the character of the model, which is fixed by it. From a modelling perspective, this has deep consequences as any interpretation of results from the model has to take such 'restrictions' into consideration.

If the value of p is thought to be known – either by reference to a symmetric experiment, or by an unambiguous statement like 'from long-standing experience from the past we know p to be 0.910' in the wording of Example 7, the probabilistic part of it becomes trivial from a modelling perspective and a direct application of binomial probabilities is required. The solution may be found either by means of a hand-held calculator or a spreadsheet, or even by old-fashioned probability tables – the answer is straightforward and undisputed.

The discussion about the various ways to deal with information about the success rate p might lead to the didactical conclusion that such questions have to be excluded from a final exam, especially if it is put forward centrally. The information in such tasks has to be conveyed clearly, the models have to be precisely and explicitly determined by the very text (not the context) of the task. The question remains – under such circumstances – would it still be worthwhile to teach probability as it would be reduced to a mere mechanistic application of the formulae in such exams? What is interesting is how the process of modelling used allows for an answer of the problem and in what respect such a model misses important features of the situation involved.

To choose between various possible models and to critically appreciate the model finally chosen is a worthwhile standard to reach in studying about probability. In only rare cases is there one distinct answer to a problem in question.

The assumptions of a Bernoulli process are not well fulfilled in sports and in many other areas where such methods are 'blindly' applied. Such assumptions establish (more or less well) a scenario (as opposed to a model that fits very well to the real situation), which allows an inspection of the situation on the basis of an 'what would be – if we assume …' Then of course, situations have to be set out where such scenarios may deliver suitable orientations despite their lack of fit to the situation (for the idea of a scenario instead of a model, see Borovcnik, 2006).

If p is not known directly, there are various ways to fill in the gap of information – the scale ranges from hypotheses of differing credibility to estimates from statistical data of differing relevance (depending on the 'grade of randomness' of the sample). Clearly, a true value of p has to be distinguished from an estimate $\hat{p}$ of p. The whole of inferential statistics is based on a careful discrimination between true parameters and estimations thereof. However, again, issues are not as easy and clear-cut. What may be viewed as a true parameter in one model may be viewed as an estimate in another model – see the ideas developed subsequently.

If the option of a *statistical estimate* of the unknown parameter is chosen as in (11), then the data has to fulfil the assumption of a random sample – an independent repetition of the same basic experiment yielding each item of data. The accuracy linked to a specific sample may be best judged by a confidence interval as in (12). It might be tempting to reduce the length of such a confidence interval and to increase the precision of information about the unknown parameter by increasing the sample size. However, in practice, to obtain more data usually means a lower quality of data; ie., the data no longer fulfil their fundamental property of being a random sample, which involves a bias in the data with no obvious way to repair it.

If the option of hypothesizing values for the unknown parameter is chosen as in (10), or in (13), one might have trouble in justifying such a hypothesis. In some cases, however, good arguments might be given. For the statistical part of Example 7 when it comes to an evaluation whether Nowitzki is better in home than in away matches, a natural hypothesis emerges from the following modelling. Split the Bernoulli process for home and away matches by a different value for p as p_H and p_A. The assumption of equal strength (home and away) leads to the hypothesis

$$p_A = p_H \text{, or } p_A - p_H = 0 . \tag{14}$$

Analysis of the data is then done under the auspices 'as if the difference of the success probabilities home and away were zero'. However, it is not straightforward to derive the distribution for the test statistic $\hat{p}_A - \hat{p}_H$.

More about Assumptions – A Homogenizing Idea 'Behind' the Binomial Distribution

In the context of sports, it is dubious to interpret relative frequencies as a probability and – vice versa – it is difficult to justify estimating an unknown probability by relative frequencies. What is different in the sports context from games of chance where the idea of relative frequencies has emerged? It is the comparability of single trials, the non-dependence of single trials – that is the hinge for transferring the ideas from games to other contexts. For probabilistic considerations such a transfer seems to be more crucial than for a statistical purpose, which focuses on a summary viewpoint.

In theory, the estimation of the success parameter *p* improves by increasing the sample size. Here, this requires combining several seasons together. However, the longer the series – especially in sports – the less plausible the assumptions for a Bernoulli process. And, if relative frequencies are used to estimate the underlying probabilities, condition (9) of a Bernoulli process has to be met. Only then do the estimates gain in precision by increasing the sample size.

However, for Nowitzki's scores, the assumptions have to be questioned. People and the sport change over time, making assumptions of random, independent trials as the basic modelling approach less tenable. Take the value of 0.904 as an estimate for his current competency to make a point with a free throw – formalized as a probability *p*. This requires an unrealistic assumption of 'homogenizing': Nowitzki's capacity was constant for the whole season and independent of any accompanying circumstances, not even influenced by the fact that in one game everything is virtually decided with his team leading or trailing by a big gap close to the end of the match, or there is a draw in the match and this free throw – the last event in the match – will decide about winning or not.

For statements related to the *whole entity* of free throws, such a homogenization might be a suitable working hypothesis. Perhaps the deviations from the assumptions balance for a longer series. To apply results derived on such a basis for the whole entity to a sequence of 10 specific throws and calculate the probability of 8 points at the most, however, makes hardly any sense, and even less so if it deals with the last 10 of an all-deciding match.

To imbue *p* with a probabilistic sense, to apply the binomial distribution sensibly, one has to invent a scenario like the following: All free throws of the whole season have been recorded by video cameras. Now we randomly select 10 clips and ask: How often does Nowitzki make the point?

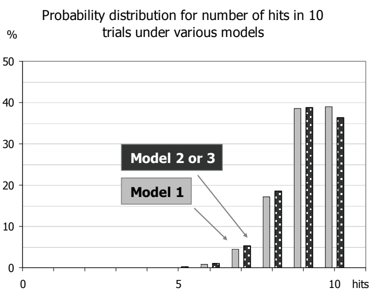
'SOLUTION' OF THE PROBABILISTIC PART OF THE NOWITZKI TASK

In this section, the solutions are derived from the various models and critically appraised. As the choice of model depends not only on assumptions but also on an estimation of unknown parameters, the question arises, which of the available models to choose – a fundamental issue in modelling. Several models are dealt with in the Nowitzki task and their relative merits are made clear. The methods of determining a good choice for the parameters also convey key features of pedagogy – some knowledge is taken from the context, some will be added by statistical estimation.

While for parts a. and b. the results seem straightforward, part c. gives 'new' insights. This task was fiercely rejected by some experts as unsolvable. However, by highlighting a fundamental property of Bernoulli series a solution of part c. is easier. If the chosen model is taken seriously, then the modeller is in the same situation as in any game of chance. In such games, the player can start at any time – it does not matter. The player can also eliminate any randomly chosen games without a general change in the result. That is the key idea involved.

Calculation of the Probability of Simple Events – Parts a. and b.

With model 1 and the assumption that $p = 0.910$, the distribution is given in (10). Using a spreadsheet gives the solution with this specific binomial distribution as set out in Table 2. Model 3 is handled in the same way.



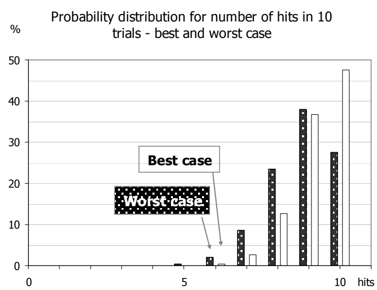


Figure 5. Probability distributions for the number of hits under the various models.

With model 2 and the estimate $p = 0.904$, the distribution is fixed in (11). Using model 2 (variant), one derives the confidence interval (12) for Nowitzki's 'strength' and uses the binomial distribution with parameters corresponding to worst and best cases for the playing strength. The distributions for the number of hits in 10 trials are depicted in Figure 5. While models 1 and 2 are similar, there is a huge difference between best and worst case in model 2 (variant).

Table 2. Probabilities calculated under the various models

	Model 1	*Model 2 (3)*[7]	*Model 2 (variant)*	
	$p = 0.910$	$\hat{p} = 0.904$	*Worst case* p_L	*Best case* p_U
$P(T_{10} = 8)$	0.1714	0.1854	0.2345	0.1273
$P(T_{10} \leq 8)$	0.2254	0.2492	0.3449	0.1573

It is remarkable, and worthy of pedagogical discussion in the classroom, that the solutions differ so much when the assumptions seem to be very similar. The inaccuracy as conveyed by the confidence interval (12) on p only reflects a margin of just under 5 percentage points. Nevertheless the variant of model 2 gives a margin of 0.1573 to 0.3449 for the probability in question b.[8] Thus, it is crucial to remember that one has only estimates of the unknown parameter and imperfect knowledge.

The Question 'Nowitzki Scores at Most Four Times in a Series'

Task c. was disputed in a public discussion, in which statisticians were also involved. It was claimed that it can not be solved without an explicit number of observations given. Suggestions to fix the task are reported and a correct solution using a key property of Bernoulli processes is given below. The following passage is taken from an open letter to the ministry of education (Davies et al, 2008):

> "The media reported that one part of the [Nowitzki task] was not solvable because the number of trials is missing. This – in fact – is true and therefore several interpretations of the task are admissible, which lead to differing solutions."

Davies (2009, p. 4) illustrates three ways to cope with the missing number of trials:

The first is to take $n = 10$ as it was used in the first part of the task. This suggestion comes out of 'student logic' but leads into an almost intractable combinatorial problem. One has to inspect all the $2^{10} = 1024$ series of 0's and 1's (for failures and successes) whether they have a single segment of five or more 1's in it (indicating the complement of the event in question), or not.

A second possibility is to take $n = 5$, which makes the problem very simple: The only sequence not favourable to the event in question is 1 1 1 1 1. Thus the probability for the complementary series, for which one is looking for here, equals

$$1 - p^5 \tag{15}$$

and the result is dependent on the chosen model (see Table 3).

Again it is surprising that the change from model 2 to its variant reveals such a high degree of imprecision implicit in the task as the required probability is known only within a range of 0.31 to 0.47, if one refers to an estimate of the unknown probability of Nowitzki's strength. But the reader is reminded that model 3 does not have this problem. Its result coincides with model 2 (without the variant) as the parameter is set to be known by (13).

Table 3. Probabilities of the series 'less than 5' with n = *5 under the various models*

	Model 1	*Model 2 (3)*	*Model 2 (variant)*	
	$p = 0.910$	$\hat{p} = 0.904$	*Worst case* p_L	*Best case* p_U
p^5	0.6240	0.6031	0.5253	0.6898
$1 - p^5$	0.3760	0.3969	0.4747	0.3102

The third and final remedy to fill in for the gap of the missing number of trials, which Davies (2009) offers, refers to an artificial new gaming situation.

> "We imagine a game where two players perform free throws. One of the players begins and continues to throw as long as his ball passes correctly through the basket and he scores. If Nowitzki starts this game what is his probability that he scores at the most four times in a series in his first try?"

Now, the possible sequences of this game are

$$0 \qquad 1\,0 \qquad 1\,1\,0 \qquad 1\,1\,1\,0 \qquad 1\,1\,1\,1\,0 \tag{16}$$

The solution emerging from (16) coincides exactly with that where the number of trials is fixed by 5, which numerically was the officially accepted solution (though it was derived without the assumption of $n = 5$ trials). Regarding the common factor $1 - p$ in the single probabilities involved in (16), we get the solution by:

$$(1 + p + p^2 + p^3 + p^4)(1 - p) = 1 - p^5 . \tag{17}$$

The third attempt to solve task c. without the missing number of trials yields the solution. However, it implies an artificial new gaming situation, which makes things unnecessarily complicated. In fact, the task is solvable *without* supplementing the missing number of trials and without this artificial game. One only has to remind oneself of what really amounts to a Bernoulli process, what properties are fundamental to such processes. The property in question will – once recalled – lead to a deeper understanding of what Bernoulli processes are. The next section will illustrate this idea. If one agrees with the assumption (9) of a Bernoulli process for the trials of Nowitzki then part c. of the probabilistic task is trivial.

> If the conditions are always the same throughout then it does not matter when one starts to collect the data. Mathematically speaking: If $X_1, X_2, \ldots$ is a Bernoulli process with relation (9), then the following two sub processes have essentially the same probabilistic features, i.e., they also follow property (9):

– random start of data collection i_0:

$$X_{i_0}, X_{i_0+1}, \ldots \overset{iid}{\sim} X \sim B(1, p)\,; \tag{18a}$$

– random selection i_0, i_1, … of all data:

$$X_{i_0}, X_{i_1}, \ldots \overset{iid}{\sim} X \sim B(1, p) \tag{18b}$$

This is a fundamental property of Bernoulli processes in particular and of random samples in general. One may start with the data collection whenever one chooses, therefore, (18a) applies. One can also eliminate some data if the elimination is undertaken randomly as in (18b). While this key property of Bernoulli processes should be explained intuitively to students, it could also be supported by simulation studies to address 'intuitive resistance' from students.

Statisticians coin the term *iid* variables. One needs to explain to students that each single reading comes from a process that – at each stage (for each single reading) – has a distribution *i*ndependent of the other stages and which follows an *i*dentical (i.e., the same) *d*istribution throughout (hence *iid*); this deceptively complex idea takes time for students to absorb. It has already been mentioned that in sports such an assumption is doubtful but the example was put forward with this assumption for modelling, which therefore will not be challenged at this stage.

If it does not matter when we (randomly) start the data collection, we just go to the sports hall and wait for the next free throws. We note whether

– Nowitzki does not score a point more than four times in a series – event A, or,
– he succeeds in scoring more than four times in a series – event $\overline{A}$

Clearly it holds: $$P(\overline{A}) = p^5 \cdot 1 \text{ and } P(A) = 1 - p^5 . \qquad (19)$$

The term p^5 in (19) stands for the first five 1's in the complementary event; this probability has to be multiplied by 1 as from the sixth trial onwards the outcome does not matter (and therefore is the certain event). The result coincides with solution (16). However, there is no need to develop this imaginary game with an opponent as Davies (2009) does in his attempt to 'fix' the task. The task is easily solved using the fundamental property (18a) of Bernoulli processes.

If one just goes to a match (maybe one is late, it would not matter!) and observes whether Nowitzki scores more than four times in a series right from the start, then everything beyond the fifth trial is redundant and the solution coincides with solution (15) where the number of observations is fixed with $n = 5$.

The Solution is Dependent on the Number of Trials Observed!

If the number n of trials is *pre-determined*, the probability to have at the most four successes in a series changes. If one observes the player only up to four times, he cannot have more than four successes, whence it holds: $P(A) = 1$. The longer one observes the player, the more is the chance to finally see him score *more* than four times in a series. It holds:

$$P(A \mid n) \to 0, \quad n \to \infty . \qquad (20)$$

KEY IDEA BEHIND VARIOUS DISTRIBUTIONS

In this section, we explain the underlying key pedagogical ideas of the following seven distributions.

a. Binomial distribution: repeated, independent trials, called Bernoulli process;
b. Hypergeometric distribution: repeated dependent trials;

c. Poisson distribution: completely random events in time – Poisson process;
d. Geometric distribution: waiting times in the Bernoulli process;
e. Exponential distribution: Poisson and memory-less waiting;
f. Weibull distribution: conditional failure rates or hazards;
g. Normal distribution: the hypothesis of independent elementary errors;

For students' understanding and also for good modelling reasons it is of advantage to have a key idea behind each of the distributions. Otherwise, it is hard to justify a specific distribution as a suitable model for the phenomenon under scrutiny. Why a Poisson distribution, or why a normal? The key idea of a distribution should convey a direct way of judging whether such a distribution could model the phenomenon in question. It allows one to check the necessary assumptions in the data generating process and whether they are plausible.

Such a fundamental idea behind a specific distribution is sometimes hidden; it is difficult to recognise it from discrete probabilities or density functions, which might also have complicated mathematical terms. Other concepts related to a random variable might help to reveal to students 'the' idea behind a distribution.

For example, a feature like a memory-less property is important for the phenomenon, which is described by a distribution. However, this property is a mathematical consequence of the distribution but cannot directly be recognized from its shape or mathematical term. In the context of waiting time, the memory-less property means that the ongoing waiting time until the 'event' occurs, has the same distribution throughout – regardless of the time already waiting for this event.

Or, technical units might show (as human beings do) a phenomenon of wearing-out, i.e., the future lifetime has, amongst others, an expected value decreasing by the age of the unit (or, the human being). To describe such behaviour, further mathematical concepts have to be introduced like the so-called hazard (see below). In technical applications continuous service of units might postpone wearing-out. For human beings, insurance companies charge higher premiums for a life insurance policy to older people.

While further mathematical concepts might be – at first sight – an obstacle for teaching, they help to shed light on key ideas for a distribution that enhance 'internal mechanisms' lurking in the background, and also help to understand the phenomena better. On the contrary, the usual examination as to whether a specific distribution is an adequate model for a situation is performed by a *statistical* test on whether the data is compatible with what is to be 'expected' from a random sample of this model or not. Such tests focus on the 'external' phenomenon of frequencies as observed in data.

a. Binomial Distribution – Repeated Independent Trials

A Bernoulli process may be represented by drawing balls from an urn where there is a fixed proportion p of balls marked by a 1 (success) and the rest marked by a 0 (failure). If one draws a ball repeatedly from that urn n times always replacing the drawn ball, the number of successes follows a binomial distribution with parameters n and p.

There are characteristic features inherent to the depicted situation: repeated experiments with the same success probability p, independent trials (mixing the balls before each draw), so that the success probability remains the same throughout. One could spin a wheel repeatedly with a sector marked as 1 and another marked as 0. This distribution was discussed at length with the Nowitzki task. The binomial distribution is intimately related to the Bernoulli process (9), which may also be analysed from the perspective of continuously observing its outcomes, until the first event occurs – see the geometric distribution below.

b. Hypergeometric Distribution – Repeated Dependent Trials

This distribution, too, is best explained by the artificial but paradigmatic context of drawing balls from an urn with a fixed number of marked (by 1's) and non-marked (the 0's) balls as in the binomial situation; however, now the drawn balls are not replaced. Under this assumption, the number of marked balls among the n drawn follows a hypergeometric distribution.

The characteristics are repeated experiments with the same success probability p, but *dependent* trials, so that the success probability remains the same only if one does not know the history of the process, otherwise there is a distinct dependence.

The context of drawing balls explains also that – under special circumstances – the hypergeometric may be well approximated by the binomial distribution: if the number n of balls drawn from the urn is small compared to the number N of all balls in the urn, then the dependence between successive draws is weaker and the conditions (9) of a Bernoulli process are nearly met.

c. Poisson Distribution – Pure Random Events in Time

It is customary to introduce the Poisson distribution as the – approximate – distribution of rare events in a Bernoulli process (p small); it is also advantageous to refer this distribution to the Poisson process even if this is lengthy and more complex. The process of generating 'events' (e.g., emitted radioactive particles), which occur in the course of time (or in space), should intuitively obey some laws that may compare to the Bernoulli process:

- The start of the observations is not relevant for the probability to observe any event; see the fundamental property (18a) – this leads to A1 in (21) below.
- If one observes the process in non-overlapping intervals, the pertinent random variables have to be independent, which corresponds to the independence of the single observations X_i in the Bernoulli process – this leads to A4.
- The main difference in the processes lies in the fundamental frequency pulsing: in the Bernoulli process, there is definitely a new experiment X_i at 'time' i whereas in the Poisson process time is continuously flowing with no apparent performing of an experiment (leading to an event or not) – events just occur.

– It remains to fix the probability of an event. As there is no distinct experiment with the outcome of the event (or its non-occurrence), we can speak only of an *intensity* λ of the process to bear events. This intensity has to be related to unit time; its mathematical treatment in A2 involves infinitesimal concepts.
– Paradoxically, a further requirement has to be demanded: even if two or more events may occur in an interval of time, which is not too small, such a probability of coincidences should become negligible if the length of the observation interval becomes small – that leads to A3 below.

Mathematically speaking (compare, e.g., the classic text of Meyer, 1970, p. 166), a Poisson process has to meet the following conditions (the random variable X_t counts the number of events in the interval $(0, t)$):

A1 If Y_t counts the events in $(t_0, t_0 + t)$ then $Y_t \sim X_t$

A2 $$P(X_{\Delta t} = 1) = \lambda \cdot \Delta t + o(\Delta t) \tag{21}$$

A3 $$P(X_{\Delta t} \geq 2) = o(\Delta t)$$

A4 X_t and Y_t are independent random variables if they count events in non-overlapping time intervals.

Assumptions (21) represent *pure randomness*; they imply that such a process has no preference for any time sequence, has no coincidences as they would occur by ‘intention’, and shows no dependencies on other events observed. The assumptions may also be represented locally by a grid square as is done in Example 8.

The main difference of the Poisson to the Bernoulli process lies in the fact that there is no definite unit of time, linked to trials 1, 2, 3, etc., which may lead to the event (1) or not. Here, the events just occur at a specific point of time but one cannot trace when an ‘experiment’ is performed. The success probability p of the Bernoulli process associated with single experiments becomes an intensity λ per unit time. The independence of trials becomes now an independence of counting events in mutually exclusive intervals in the sense of A4.

The Poisson process will have a further analogue to the Bernoulli process in terms of waiting for the first event – the Geometric and the Exponential distribution (which describe waiting times in pertinent situations) both have similar properties (see below). We present one example here to illustrate a modelling approach to the Poisson. This shows how discussions can be initiated with students on the theoretical ideas presented above, and help students to understand how and when to apply the Poisson distribution.

Example 8. Are the bomb attacks of London during World War II the result of a planned bombardment, or may they be explained by pure random hitting? To compare the data to the scenario of a Poisson process, the area of South London is divided into square grids of ¼ square kilometres each. The statistics in Table 4 shows e.g., 93 squares with 2 impacts each, which amounts to 186 bomb hits. In sum, 537 impacts have been observed.

Table 4. Number of squares in South London with various numbers of bomb hits – Comparison to the frequencies under the assumption of a Poisson process with λ = 0.9323

No. of hits in a square	0	1	2	3	4	5 and more	all
No. of grid squares with such a no. of hits	229.	211.	93.	35.	7.	1.	576.
Expected numbers under Poisson process	226.74	211.39	98.54	30.62	7.14	1.57	

If targeting is completely random, it follows the rules of a Poisson process (21) and the number of 'events' per grid square follows then a Poisson distribution. The parameter λ is estimated from the data to fix the model by

$$\lambda = \tfrac{537}{576} = 0.9323 \text{ hits per grid square.} \tag{22}$$

As seen from Table 4, the fit of the Poisson distribution to the data is extremely good. Feller (1968, pp 161) highlights the basic property of the Poisson distribution as modelling pure randomness and contrasts it to wide-spread misconceptions:

> "[The outcome] indicates perfect randomness and homogeneity of the area; we have here an instructive illustration of the established fact that to the untrained eye randomness appears as regularity or tendency to cluster."

In any case of an application, one might inspect whether the process of data generation fulfils such conditions – which could justify or rule out this distribution as a candidate for modelling. The set of conditions, however, also structures thinking about phenomena, which may be modelled by a Poisson distribution. All phenomena following internal rules, which come close to the basic requirements of a Poisson process, are open to such a modelling.

d. Geometric Distribution – Memory-Less Waiting for an Event

Here, a Bernoulli process with success parameter p is observed. In contrast to the binomial distribution, the number of trials is not fixed. Instead, one counts the number of trials until the first event (which corresponds to the event 'marked' by a 1) occurs. The resulting distribution 'obeys' the following 'memory-less' property:

$$P(T > k_0 + k \mid T > k_0) \;=\; P(T > k)\,. \tag{23}$$

This feature implies that the remaining waiting time for the first event is independent of the time k_0 one has already waited for it – waiting gives no bonus. Such a characterization of the Bernoulli process helps in clarifying some basic misconceptions. The following example can be used to motivate students on the underlying features.

Example 9. Young children remember long waiting times for the six on a die to come. As waiting times of 12 and longer still have a probability of 0.1346, see also Figure 6, this induces them to 'think' that a six has less probability than the other numbers on the die for which such a 'painful' experience is not internalised.

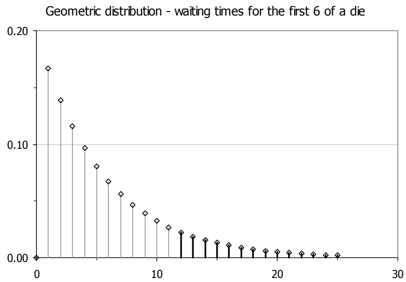


Figure 6. Waiting for the first six of a die – Bernoulli process with p = *1/6.*

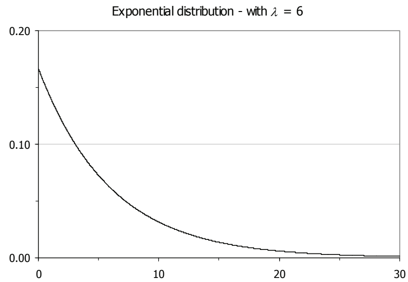


Figure 7. Exponential distribution has the same shape as the geometric distribution.

e. Exponential Distribution – Memory-Less Waiting for Events in Time

The exponential distribution is connected to two key ideas: one links it to the Poisson process; the other uses the concept of conditional failure rate. In a Poisson process, if one is waiting for the next event to occur and the data are subsequent waiting times between the events, then the exponential distribution is the model of choice. This is due to a mathematical theorem (see Meyer 1970, p. 191). It can also be illustrated by simulation studies. An important feature of the exponential distribution is its *memory-less property*:

$$P(t_0 < T \leq t_0 + \Delta t \mid T > t_0) = \frac{P(t_0 < T \leq t_0 + \Delta t)}{P(T > t_0)} = P(0 < T \leq \Delta t)\,. \qquad (24)$$

Due to the memory-less property, the conditional probability to fail within Δt units of time for a device that has reached age t_0 is the same as within the first Δt units for a new device. This implies that the future lifetime (or, waiting time) is independent of age reached (or, the time already spent in waiting), i.e., t_0, which amounts to a further characterization of 'pure' randomness. Exponential and geometric distributions share the memory-less property. This explains why the models have the same shape. If this conditional failure probability is calculated per unit time and the time length Δt is made smaller, one gets the conditional failure rate, or hazard $h(t)$:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t_0 < T \leq t_0 + \Delta t \mid T > t_0)}{\Delta t}\,. \qquad (25)$$

A hazard (rate) is just a different description of a distribution. Now it is possible to express the other key idea behind the exponential distribution, namely that its related conditional failure rate (or, hazard) is *constant* over the lifetime. If a (technical) unit's lifetime is analysed and the internal structure supports that the remaining lifetime is independent of the unit's age, then it may be argued that an exponential distribution is the model of choice. While such a property might seem paradoxal (old units are equally good as new units), it is in fact well fulfilled for electronic devices for a long part of their ordinary lifetime. Mechanical units, on the contrary, do show a wearing effect, so that their conditional lifetime gets worse with age. Similarly with human beings, with the exception of infant mortality when – in youngest ages – humans' lifetime as a probability distribution improves.

f. Weibull Distribution – Age-Related Hazards

Lifetimes are an important issue in technical applications (reliability issues and quality assurance), waiting times are important in describing the behaviour of systems. There are some families of distributions, which may serve as suitable models. The drawback with these is that they require more mathematics to describe their density functions. Furthermore, the shape of their density gives no clue why they should yield a good model for a problem to be analysed.

To view lifetimes (or waiting times) from the perspective of units that have reached some specific age already (have waited some specific time) sheds much more light on such phenomena than to analyse the behaviour of new items (with no time spent in waiting in the system). One would, of course, simulate such models first, explore the simulated data, and draw preliminary conclusions before one starts to delve deeper into mathematical issues. It may pay to learn – informally – about hazards and use this concept instead of probability densities to study probability models. Hazards will directly enhance the basic assumptions, which have to be fulfilled in case of applications.

With the key idea of hazard or conditional failure rate, the discussion can relate to infant mortality (decreasing), purely random failures due to exponential lifetime (constant) and wearing-out effects (increasing). The power function is the simplest model to describe all these different types of hazard:

$$h(t) = \beta \left(\tfrac{t}{\alpha}\right)^{\beta-1}, \; \alpha, \beta > 0 \,. \tag{26}$$

The parameter α is interpreted as the scale of time while β influences the shape and thus the quality of the change of hazard over lifetime.

g. Normal Distribution – the Hypothesis of Independent Elementary Errors

Any random variable that might be split into a sum of other (hidden) variables is – according to the central limit theorem (CLT) – approximately normally distributed. This explains the key underlying idea and ubiquity of the normal distribution. In the history of probability, the CLT prompted the 'hypothesis of elementary errors' (Gauss and earlier) where any measurement error was hypothesized to be the result (sum) of other, elementary errors. This supported the use of the normal distribution for modelling measurement errors in astronomy and geodesy.

A generalization to the normal 'law' of distribution by Quételet and Galton is straightforward: it is an expression of God's will (or Nature) that any biometric measurement of human beings and animals is normally distributed as it emerges from a superposition of elementary 'errors of nature' (Borovcnik, 2006)[9]. An interesting article about the history and myth of the normal law is Goertzel (n.d.).

The mathematics was first proved by de Moivre and Laplace; the single summands X_i had then been restricted to a Bernoulli process (9). In this way, the binomial distribution is approximately normally distributed and the approximation is good enough if there are enough elements in the sum:

$$T_n = X_1 + X_2 + ... + X_n . \tag{27}$$

To illustrate matters, the Galton board or an electronic quincunx (see, e.g., Pierce, R., n.d.) may be used in teaching. Such a board has several rows of pegs arranged in a shape similar to Pascal's triangle. Marbles are dropped from the top and then bounce their way down. At the bottom they are collected in little bins. Each time the marble hits one of the pegs, it may bounce either left or right. If the board is set up symmetrically the chances of bouncing either way are equal and the marbles in the bins follow the 'bell shaped' curve of the normal distribution. If it is inclined, a skewed distribution emerges, which normalizes, too, if enough rows are taken.

Theoretically, one has to standardize the value of the sum T_n according to

$$U_n = \frac{T_n - E(T_n)}{\sqrt{\mathrm{var}(T_n)}} \tag{28}$$

and the CLT in its crudest form becomes:

$$\text{If } X_i \overset{iid}{\sim} X \text{ are an } iid \text{ process with finite variance } var(X) < \infty ,$$
$$\text{then it holds } \lim_{n \to \infty} P(U_n \leq u) = \Phi(u) \tag{29}$$

Here $\Phi(u)$ stands for the cumulative distribution function of the standard normal distribution, i.e., with parameters 0 and 1.

Despite such mathematical intricacies, the result is so important that is has to be motivated in teaching. The method of simulation again is suitable not only to clarify the limiting behaviour of the sum (the distribution of its standardized form converges to the normal distribution), but also to get an orientation about the speed of convergence. Furthermore, this convergence behaviour is highly influenced by the shape of the distribution of the single X_i's.

A scenario of simulating 1000 different samples of size $n = 20$ and then $n = 40$ from two different distributions (see Figure 8) may be seen from Figure 9. The graph shows the frequency distributions of the *mean* of the single items of data

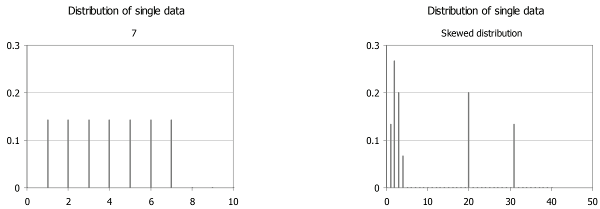


Figure 8. A symmetric and a skewed distribution for the single summands in the scenario.

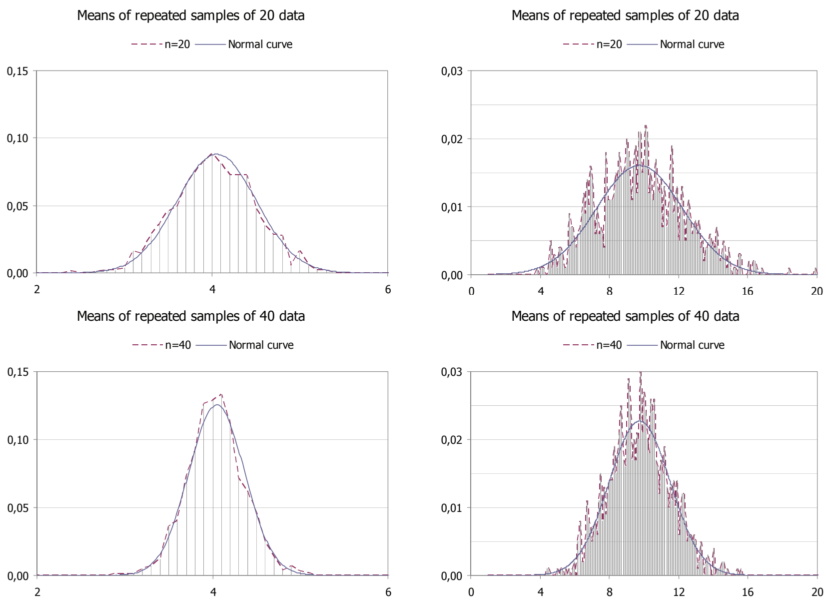


Figure 9. Scenario of 1000 samples: distribution of the mean compared to the normal curve – left drawing from an equi-distribution, right drawing from a skewed distribution.

instead of the sum (27) – this is necessary to preserve scales as the sums are simply diverging. For the limit, the calculation of the mean still does not suffice as the mean converges weakly to one number (the expected value of X_i if all have the same distribution) – thus in the limit there would be no distribution at all.

The simulation scenarios in Figure 9 illustrate that the calculated means of the repeated samples have a frequency distribution, which comes quite close to a normal distribution for only 20 summands. If the items of data are skewed, the approximation is slightly worse but with calculated means of 40 items of data in each sample the fit is sufficiently good again.

The influence of the single summands (like those in Figure 8) on the convergence behaviour of the sum may be studied interactively: With a spreadsheet with slide controls for the number of values in the equi-distribution for the single data one could easily see that more values give a faster convergence to the fit. With a slide control for the skewness, e.g., to move the two highest values further away from the bulk of the data, one may illustrate a negative effect on convergence as the fit would become worse this way. By changes of the slide controls the effect on the distribution for an item of data X_i is seen from the bar graphs in Figure 8 and the effect on the 'normalizing' of the distribution of the mean of the repeatedly drawn samples may be studied from Figure 9 interactively.

Distributions Connected to the Normal Distribution

There are quite a few distributions, which are intimately connected to the normal distribution. The main usage of these is to describe the theoretical behaviour of certain test statistics based on a sample from a normal distribution. Amongst them are the t, the χ^2 and F distribution. They are mainly used for coping with the mathematical problems of statistical inference and not for modelling phenomena. The χ^2 distribution is somehow an exception to this, as the so-called Maxwell and Rayleigh distribution (the square root of a χ^2) are also used by physicists to model velocity of particles (like molecules) in two or three dimensional space (with 2 or 3 degrees of freedom), see also Meyer (1970, pp. 220).

SOLUTIONS TO THE STATISTICAL PART OF THE NOWITZKI TASK

This section returns to the statistical questions of the Nowitzki task. Is Nowitzki weaker away than home? The usual way to answer such questions is a statistical test of significance. Such a modelling approach includes several steps to transfer the question from the context into a statistical framework, in which a null hypothesis reflects the situation of 'no differences' and alternative distributions depict situations of various degree of difference. As always in empirical research, there is no unique way to arrive at a conclusion.

The chosen model might fit more for the one expert, and less for another one. Already the two questions posed in the formulation of the example (away weaker than home, or, away weaker than in all matches) give rise to disputes. The logic of a statistical test makes things not easier as one always has to refer to fictional situations in the sense 'what would be if …' Errors of type I and II, or p values give rise to many misinterpretations by students. And there are many different test statistics which use information, which could discriminate between the null and alternative hypotheses differently (not only in the sense of less precise and more precise but simply different with no direct way for a comparison). Moreover, one has to estimate parameters, or use other information to fix the hypotheses.

If a Bernoulli process with success probability p is observed n times, then for the expected value of the number of successes T_n it holds:

$$T_n = X_1 + ... + X_n : \qquad E(T_n) = n\,p\,. \tag{30}$$

Here, the value of p usually is not known and is called the true value for the (success) probability.

Solutions to the First Statistical Part – Nowitzki Away Weaker than Home & Away?

Part a. of the statistical questions in Example 7 is ill-posed[10] insofar as the comparison of away matches against all matches does not reflect what one really should be interested in. Is Nowitzki away weaker than in home matches? A reference to comparison including *all* matches blurs the differences. Therefore, this question is omitted here. We will only discuss the inherent problems. With the three different success probabilities for home, away and all matches, it holds:

$$p_0 = \frac{n_H}{n} \cdot p_H + \frac{n_A}{n} \cdot p_A \tag{31}$$

If the season is regarded as a self-contained entity, all success rates are known. If they are perceived as probabilities, the next question to discuss is whether there is one common process or two or more with different probabilities; a covariate like 'location' of the free throw (home, away) would explain the differences. If p_0 is set as known, the problem may be handled in this way: 'May the away free throws be modelled by a Bernoulli process with p_0 from the overall strength?'

If the data is seen as a sample from an infinite (Bernoulli) process, p_0 has to estimated from it, however, there are drawbacks in question a. and its modelling. Firstly, by common sense, no one would compare away scores to *all* scores in order to find differences between the two groups of trials *away* and *home*. Secondly, as the overall strength is estimated, it could also be estimated by the separate scores of away and home matches using equation (31): $\hat{p}_A$ and $\hat{p}_H$ are combined to an estimate $\hat{p}_0$ of p_0. And the test would be performed by the data on away matches, which coincide with $231 = n_A \cdot \hat{p}_A$.

Confusing here is that p_A is dealt with as unknown (a test is performed whether it is lower than the overall strength), an estimate of it is used to get an estimate of the overall strength, and it is used as known data to perform the test.

Solution to the Second Statistical Part – Nowitzki Weaker Away Than at Home?

In this subsection, the away scores are compared to the home matches only (part b). Various steps are required to transform a question from the context to the statistical level. It is illustrated how these steps lead to hypotheses at the theoretical level, which correspond and 'answer' the question at the level of context.

Three different Bernoulli processes are considered: home, away, and all (home and away combined). After the end of the season, the related success probabilities are (factually) known from the statistics (see Table 5). Or, one could at least estimate some of these probabilities from the data.

Table 5. Playing strength as success probabilities from hits and trials of the season

Matches	*Hits*	*Trials*	*Strength 'known' or estimated*
Home	$T_H = 267$	$n_H = 288$	$p_H = \frac{267}{288} = 0.927$
Away	$T_A = 231$	$n_A = 263$	$p_A = \frac{231}{263} = 0.878$
All	$T = 498$	$n = 551$	$p_0 = \frac{498}{551} = 0.904$

The basis for reference to compare the away results is the success in home matches. For home matches the success probability is estimated as in model 2 or 'known' from the completed season as in model 3

$$p_H = 0.927. \tag{32}$$

Formally, the question from the context can be transferred to a (statistical) test problem in several steps of choosing the statistical model and hypotheses:

$$T_A \sim \mathrm{B}(n_A, \pi)\,, \tag{33}$$

ie., the number of hits (successes) in away matches is binomially distributed with unknown parameter π (despite reservations, this model is used subsequently). As null hypothesis

$$H_0 : \pi = p_H\,, \tag{34}$$

will be chosen. This corresponds to 'no difference of away to home matches' from the context with p_H designating the success probability in home matches. For the alternative, a one-sided hypothesis

$$H_1 : \pi < p_H \tag{35}$$

is suggested. As in sports in general, the advantage of the home team is strong folklore, a one-sided[11] hypotheses makes more sense than a two-sided alternative of $\pi \neq p_H$. The information about p_H comes from the data, therefore it will be estimated by 0.927 to form the basis of the 'model' for the away matches. No other information about the 'strength' in home matches is available. Thus, the reference distribution for the number of successes in away matches is the following:

$$T_A\big|_{H_0} \sim \mathrm{B}(n_A, 0.927)\,. \tag{36}$$

Relation (36) corresponds to the probabilistic modelling of the null effect that 'away matches do not differ from home matches'. The alternative is chosen to be one-sided as in (35). The question is whether the observed score of $T_A = 231$ in away matches amounts to an event, which is significantly too low for a Bernoulli process with 0.927 as success probability, which is in the background of (36). Under this assumption, the expected value of successes in away matches is $263 \cdot 0.927 = 243.8$. The p value of the observed number of 231 successes is now as small as 0.0033! Consistently, away matches differ significantly from home matches (at the 5% level of significance).

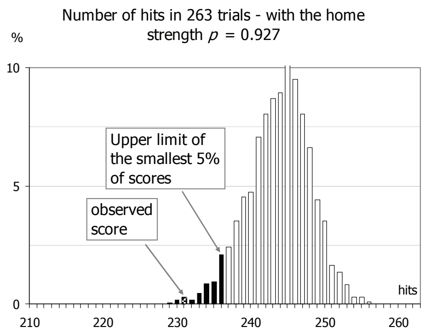


Figure 10. Results of 2000 fictitious seasons with n_A = 263 trials – based on an assumed strength of p = 0.927 (corresponding to home matches).

In Figure 10, the scores of Nowitzki in 2000 fictitious seasons are analysed. The scenario is based on his home strength with 263 free throws (the number of trials in away matches in the season 2006–07) ie., on the distribution of the null hypothesis in (36). From the bar chart it is easily seen that the observed score of 231 is far out in the distribution; it belongs to the smallest results of these fictitious seasons. In fact, the p value of the observation is 0.3%.

A simulation study gives concrete data; the lower 5% quantile of the artificial data separates the 5% extremely low values from the rest. It is easy to understand that if actual scores are smaller than this threshold, they may be judged as not ‘compatible’ with the underlying assumptions (of the simulated data, especially the strength of 0.927). To handle a rejection limit (‘critical value’) from simulated data is easier than to derive a 5% quantile from an ‘abstract’ probability distribution.

Validity of Assumptions – Contrasting Probabilistic and Statistical Point of Views

The *scenario* of a Bernoulli process is more compelling for evaluating the question of whether Nowitzki is weaker in away matches than for the calculation of single probabilities of specific short sequences. In fact, whole blocks of trials are compared. This is not to ask for the probability for a number of successes in short periods of the process but to ask whether there are differences in large blocks *on the whole*. Homogenization means a balancing-out of short-term dependencies, or of fluctuation of the scoring probability over the season due to changes in the form of the player, or due to social conditions like quarrels in the team over an unlucky loss. In this way, the homogenization idea seems more convincing.

> The compensation of effects across single elements of a universe is essentially a fundamental constituent of a *statistical* point of view.

For a statistical evaluation of the problem from the context, the scenario of a Bernoulli process – even though it does not apply really well – might allow for relevant results. For *whole* blocks of data, which are to be compared against each other, a homogenization argument is much more compelling as the violations of the assumptions might balance out 'equally'. The situation seems to be different from the probabilistic part of the Nowitzki problem where it was doomed to failure to find suitable situations for which this scenario could reasonably be applied.

Alternative Solutions to – Nowitzki Weaker Away than at Home?

Some alternatives to deal with this question are discussed; they avoid the 'confusion' arising from the different treatment of the parameters (some are estimated from the data and some are not). Not all of them are in the school curricula.

Table 6. Number of hits and trials

Matches	*Hits*	*Failures*	*Trials*
Home	267	21	288
Away	231	32	263
All matches	498	53	551

Fisher's exact test. is based on the hypergeometric distribution. It is remarkable that it relies on less assumptions than the Bernoulli series and it uses nearly all information about the difference between away and home matches. The (one-sided) test problem is now depicted by the pair of hypotheses:

$$H_0\text{: } p_A - p_H = 0 \text{ against } H_1\text{: } p_A - p_H < 0 \tag{37}$$

'No difference' between away and home matches is modelled by an urn problem relative to the data in Table 6: All $N = 551$ trials are represented by balls in an urn; $A = 498$ are white and depict the hits, 53 are black and model the failures. The balls are well mixed and then one draws $n = 263$ balls for the away matches (without replacement). The test is based on the number of white balls N_W among the drawn balls. Here, a hypergeometric distribution serves as reference distribution:

$$N_W\big|_{H_0} \sim Hyp(N = 551, A = 498, n = 263)\,, \tag{38}$$

Small values of N_W indicate that the alternative hypothesis H_1 might hold. The observation of 231 white balls has a p value of 3.6%. Therefore, at a level of 5% (one-sided), Nowitzki is significantly weaker away than in home matches.

With this approach neither the success probability for away nor for home matches were estimated. The hypergeometric distribution needs not to be tackled with all the mathematical details. It is sufficient to describe the situation structurally and get (estimations of) the probabilities by a simulation study.

Test for the difference of two proportions. This test treats the two proportions for the success in away and home matches in the same manner as both are estimated from the data, which are modelled as separate Bernoulli processes according to (9):

$$\hat{p}_A \text{ estimates the away strength } p_A\text{; } \hat{p}_H \text{ the home strength } p_H \tag{39}$$

The (one-sided) test problem is again depicted by (37). However, the test statistic now is directly based on the estimated difference in success rates $\hat{p}_A - \hat{p}_H$. By the central limit theorem, this difference (as a random variable) – normalized by its standard error – is approximately normally distributed; it holds:

$$U := \frac{\hat{p}_A - \hat{p}_H}{\sqrt{\frac{\hat{p}_A \cdot (1-\hat{p}_A)}{n_A} + \frac{\hat{p}_H \cdot (1-\hat{p}_H)}{n_H}}} \overset{approx}{\sim} N(0,1)\,. \tag{40}$$

There is now a direct way to derive rejection values to decide whether the observed difference between the success rates of the two Bernoulli processes is significant or not: The estimated value of the difference of the success rates of – 0.04876 gives rise to an U of –1.9257 which amounts to a p value of 2.7%. In this setting a simulation of the conditions under the null hypothesis is not straightforward.

If one tests *mean* values from two different samples for difference then one can use the so-called Welch test. Both situations – testing two proportions or two means for significant differences – are too complex for introductory probability courses at the university. One could motivate the distributions and focus on the problem of judging the difference between the two samples. However, the simpler Fisher test may be seen as the better alternative for proportions.

Some Conclusions on the Statistical Modelling of the Nowitzki Task

Inferential statistics means evaluating hypotheses by data (and mathematical techniques). Is the strength of Nowitzki away equal to $p = 0.927$? An answer to this question depends on whether we search for deviations from this hypothesized value in both directions (two-tailed) or only in the direction of lower values (one-tailed). From the context, the focus may well be on lower values of p for the alternative as the advantage of the home team is a well-known matter in sport.

The null hypothesis forms the reference basis for the given data. For its formulation, further knowledge is required, either from the context, or from the data. Such knowledge should never be mixed with data, which is used in the subsequent test procedure; this is a crucial problem of task a., which asks to compare the away scores to the score in all matches. A value for all matches – if estimated from the data – also contains the away matches. However, this should be avoided, not only for methodological reasons but by common sense too.

If the season is seen as self-contained, the value of $p = 0.927$ *is* known. A test of 0.927 against alternative values of the strength less than 0.927 corresponds to the question 'Is Nowitzki away weaker than home?' 0.927 might as well be seen as an estimate of a larger imaginary season. An evaluation of its accuracy (as in section 3) is usually not pursued. A drawback might be seen in the unequal treatment of the scores: home scores are used to estimate a parameter p_H while the away scores are treated as random variable. Note that in this test situation no 'overlap' occurs between data used to fix the null hypothesis and data used to perform the test.

The alternative tests discussed here treat the home and away probabilities in a symmetric manner: both are assumed as unknown; either both are estimated from the data, or estimation is avoided for both. These tests express a correspondence between the question from the context and their test statistics differently. They view the situation as a two-way-sample. Such a view paves the way to more general types of questions in empirical research, which will be dealt with below.

STATISTICAL ASPECTS OF PROBABILITY MODELLING

Data have to be seen in a setting of model and context. If two groups are compared – be it a treatment group receiving some medical treatment and a control group receiving only placebo (a pretended treatment), two success rates might be judged for difference as in the Nowitzki task. Is treatment more effective than placebo? The assumption of Bernoulli processes 'remains' in the background (at least if we measure success only on a 0–1 scale). However, such an assumption requires a heuristic argument like the homogenization of data in larger blocks.

The independence assumption for the Bernoulli model is not really open to scrutiny as it leads to methodological problems (a null hypothesis can not be statistically confirmed). The idea of searching for covariates serves as a strategy to make the two groups as equal as they could be. Data may be interpreted sensibly and used for statistical inference – in order to generalize findings from the concrete data – only by carefully checking whether the groups *are* homogenous. Only then, do the models lead to relevant conclusions beyond concrete data. If success is measured on a continuous scale, the mathematics becomes more complicated but the general gist of this heuristic still applies.

Dealing with the Inherent Assumptions

A further example illustrates the role of confounders.

Example 10. Table 7 shows the proportions of girl births in three hospitals. Can they be interpreted as estimates of the same probability for a female birth? Assume such a proportion equals 0.489 worldwide; with hospital B (and 358 births) the proportion of girls would lie between 0.437 and 0.541 (with a probability of

Table 7. Proportion of girls among new-borns

Hospital	*Births*	*Proportion of girl births*
A	514	0.492
B	358	0.450
C		0.508
'World stats'		0.489

approximately 95%). The observed value of 0.450 is quite close to the lower end of this interval. This consideration sheds some doubt on it that the data have been 'produced' by mere randomness, ie., by a Bernoulli process with $p = 0.489$.

It may well be that there are three different processes hidden in the background and the data do not emerge from one and the same source. Such 'phenomena' are quite frequent in practice. However, it is not as simple as that one could go to hospital *B* if one wants to give birth to a boy? One may explain the big variation of girl births between hospitals by reference to a so-called *covariate*: one speculation refers to location; hospital *A* in Germany, *C* in Turkey, and *B* in China. In cases where covariates are not open to scrutiny as information is missing about them, they might blur the results – in such cases these variables are called confounders.

For the probabilistic part of Nowitzki, it might be better to search for confounders (whether he is in good form, or had a quarrel with his trainer or with his partner) in order to derive at a probability that he will at the most score 8 times out of 10 trials instead of modelling the problem by a Bernoulli process with the seasonal strength as success probability. Such an approach cannot be part of a formal examination but should feature in classroom discussion.

Empirical Research – Generalizing Results from Limited Data

Data always has to be interpreted by the use of models *and* by the knowledge of context, which influences not only the thinking about potential confounders but also guides the evaluation of the practical relevance of conclusions drawn.

A homogenization idea was used to support a probabilistic model. For the Bernoulli process the differing success probabilities should ease out, the afflictions of independency should play less of a role when series of data are observed, which form a greater entity – as is done from a statistical perspective. This may be the case for the series of away matches as a block and – in comparison to it – for the home matches as a block. One might also be tempted to reject such a homogenization idea for the statistical part of the Nowitzki task. However, a slight twist of the context, leaving the data unchanged, brings us in the midst of empirical research; see the data in Table 8, which are identical to Table 5.

Table 8. Success probabilities for treatment and control group

Group	*Success*	*Size*	*Success probability*
Treatment	267	$n_T = 288$	$p_T = 0.927$
Control	231	$n_C = 263$	$p_C = 0.878$
All	498	$N = 551$	$p_0 = 0.904$

This is a two-sample problem, we are faced with judging a (medical) treatment for effectiveness (only with a 0, 1 outcome, not with a continuous response). The Nowitzki question reads now as: Was the actual treatment more effective than the placebo treatment applied to the persons in the control group?

How can we justify the statistical inference point of view here? We have to model success in the two groups by a different Bernoulli process. This modelling includes the same success probability throughout, for all people included in the treatment group as well the independence of success between different people.

Usually, such a random model is introduced by the *design of the study*. Of course, the people are not selected randomly from a larger population but are chosen by convenience – they are primarily patients of the doctors who are involved in the study. However, they are *randomly attributed* to one of the groups, i.e., a random experiment like coin tossing decides whether they are treated by the medical treatment under scrutiny or they receive a placebo, which looks the same from outside but has no expected treatment effect – except the person's psychological expectation that it could affect. Neither the patient, nor the doctors, nor persons who measure the effect of the treatment, should know to which group a person is attributed – the golden standard of empirical research is the so-called double-blind randomized treatment and control group design. The random attribution of persons to either group should make the two groups as comparable as they could be – it should balance all known covariates and all unknown confounders, which might interfere with the effect of treatment.

Despite all precautions, patients would differ by age, gender, stage of the disease, etc. Thus, they do not have a common success probability that the treatment is effective. All what one can say is that one has undertaken the usual precautions and one hopes that the groups are now *homogenous* enough to apply the model in a manner of a scenario: 'what does the data tell us if we think that the groups meet a ceteris paribus condition'. A homogenization argument is generally applied to justify drawing conclusions out of empirical data. It is backed by random attribution of persons to the groups, which are to be compared. The goal of randomizing is to get two homogenous groups that differ only with respect to what has really been administered to them: medication or placebo.

CONCLUSIONS

The two main roles for probability are to serve as a genuine tool for modelling and to prepare and understand statistical inference.

- Probability provides an important set of concepts in modelling phenomena from the real world. Uncertainty or risk, which combines uncertainty with impact (win or loss, as measured by utility) is either implicitly inherent to reality or emerges of our partial knowledge about it.
- Probability is the key to understand much empirical research and how to generalize findings from samples to populations. Random samples play an eminent role in that process. The Bernoulli process is a special case of random sampling. Moreover, inferential statistical methods draw heavily on a sound understanding of conditional probabilities.

The present trend in teaching is towards simpler concepts focusing on a (barely) adequate understanding thereof. In line with this, a problem posed to the students has to be clear-cut, with no ambiguities involved – neither about the context nor about the questions. Such a trend runs counter to any sensible modelling approach.

The discussion about the huge role intuitions play in the perception of randomness was initiated by Fischbein (1975). Kapadia and Borovcnik (1991) focused their deliberations on 'chance encounters' towards the interplay between intuitions and mathematical concepts, which might influence and enhance mutually. Various, psychologically impregnated approaches have been seen in the pertinent research. Kahneman and Tversky (1972) showed the persistent bias of popular heuristics people use in random situations; Falk and Konold (1992) entangle with causal strategies and the so-called outcome approach, a tendency to re-formulate probability statements into a direct, clear prediction. Borovcnik and Peard (1996) have described some specifities, which are peculiar to probability and not to other mathematical concepts, which might account for the special position of probability within the historic development of mathematics. The research on understanding probability is still ongoing, as may be seen from Borovcnik and Kapadia (2009). Lysø (2008) makes some suggestions to take up the challenge of intuitions right from the beginning of teaching.

All these endeavours to understand probability more deeply, however, seem to have had limited success. On the contrary, the more the educational community became aware about the difficulties, the more it tried to suggest cutting out critical passages, which means that probability is slowly but silently disappearing from the content being taught. It is somehow a solution that resembles that of the mathematicians when they teach probability courses at university: they hurry to reach sound mathematical concepts and leave all ambiguities behind. The approach of modelling offers a striking opportunity to counterbalance the trend.

Arguments that probability should be reduced in curricula at schools and at universities in favour of more data-handling and statistical inference might be met by the examples of this chapter; they connect approaches towards context and applications like that of Kapadia and Andersson (1987). Probability serves to model reality, to impose a specific structure upon it. In such an approach, key ideas to understand probability distributions turn out to be a fundamental tool to convey the implicit specific thinking about the models used and the reality modelled hereby. Contrary to the current trend, the position of probability within mathematics curricula should be reinforced instead of being reduced.

We may have to develop innovative ways to deal with the mathematics involved. To learn more about the mathematics of probability might not serve the purpose as we may see from studies in understanding probabilistic concepts by Díaz and Batanero (2009). The perspective of modelling seems more promising: a modeller never understands all mathematical relations between the concepts. However, a modeller 'knows' about the inherent assumptions of the models and the restrictions they impose upon a real situation.

Indirectly, modelling was also supported by Chaput, Girard, and Henry (2008, p. 6). They suggest the use of simulation to construct mental images of randomness.

Real applications are suggested by various authors to overcome the magic ingredients in randomness, e.g. Garuti, Orlandoni, and Ricci (2008, p. 5).

Personal conceptions about probability are characterized by an overlap between objective and subjective conceptions. In teaching, subjective views are usually precluded; Carranza and Kuzniak (2008, p. 3) note the resulting consequences:

> "Thus the concept […] is truncated: the frequentist definition is the only one approach taught, while the students are confronted with frequentist and Bayesian problem situations."

In modelling real problems, the two aspects of probability are always present; it is not possible to reduce to one of these aspects as the situation might lose sense. Modelling thus might lead to a more balanced way in teaching probability. From the perspective of modelling, the overlap between probabilistic and deterministic reasoning is a further source of complications as Ottaviani (2008, p. 1) stresses that probability and statistics belong to a line of thought which is essentially different from deterministic reasoning:

> "It is not enough to show random phenomena. […] it is necessary to draw the distinction between what is random and what is chaos."

Simulation or interactive animations may be used to reduce the need for mathematical sophistication. The idea of a scenario helps to explore a real situation as shown in section 1. The case of taking out an insurance policy for a car is analysed to some detail in Borovcnik (2006). There is a need for a reference concept wider than the frequentist approach. The perspective of modelling will help to explore issues.

In modelling, it is rarely the case that one already knows all relevant facts from the context. Several models have to be used in parallel until one may compare the results and their inherent assumptions. A final overview of the results might help to solve some of the questions posed but raises some new questions. Modelling is an iterative cycle, which leads to more insights step by step. Of course, such a modelling approach is not easy to teach, and it is not easy for the students to acquire the flexibility in applying the basic concepts to explore various contexts.

Examinations are a further hindrance. What can and should be examined and how should the results of such an exam be marked? Problems are multiplied by the need for centrally set examinations. Such examination procedures are intended to solve the goal of comparability of results of final exams throughout a country. They also form the basis for interventions in the school system: if a high percentage fail such an exam in a class, the teacher might be blamed, while if such a low achievement is found in a greater region, the exam papers have to be revised etc. However, higher-order attainment is hard to assess.

While control over the result of schooling via central examinations 'guarantees' standards, such a procedure also has a levelling effect in the end. The difficult questions about a comparison of different probability models and evaluating the relative advantages of these models, and giving justifications for the choice of one or two of these models – genuine modelling aspects involve ambiguity – might not leave enough scope in the future classroom of probability.

As a consequence of such trends, teaching will focus even more on developing basic competencies. From applying probabilistic models in the sense of modelling contextual problems, only remnants may remain – mainly in the sense of mechanistic ‘application’ of rules or ready-made models.

To use one single model at a time does not clarify what a modelling approach can achieve. When one model is used finally, there is still much room for further modelling activities like tuning the model’s parameters to improve a specific outcome, which corresponds to one’s benefit in the context. A wider perspective on modelling presents much more potential for students to really understand probability.

NOTES

1. Kolmogorov’s axioms are rarely well-connected to the concept of distribution functions.
2. All texts from German are translated by the authors.
3. ’What is to be expected?’ would be less misleading than ‘significantly below the expected value’.
4. A ‘true’ value is nothing more than a façon de parler.
5. This is quite similiar to the ‘ceteris paribus’ condition in economic models.
6. In fact, the challenge is to detect a break between past and present; see the recent financial crisis.
7. Model 3 yields identical solutions to model 2. However, its connotation is completely different.
8. The final probability needs not be monotonically related to the input probability *p* as in this case.
9. Quetelet coined the idea of ‘l’homme moyen’. Small errors superimposing to the (ideal value of) l’homme moyen ‘lead directly’ to the normal distribution.
10. Beyond common sense issues, the ill-posed comparison of scores in away matches against all matches has several – statistical and methodological drawbacks.
11. The use of a one-sided alternative has to be considered very carefully. An unjustified use could lead to a rejection of the null hypothesis and to a statistical ‘proof’ of this pre-assumption.

REFERENCES

Borovcnik, M. (2006). Probabilistic and statistical thinking. In M. Bosch (Ed.), *European research in mathematics education IV* (pp. 484–506). Barcelona: ERME. Online: ermeweb.free.fr/CERME4/

Borovcnik, M. (2011). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, C. Reading, & A. Rossman, (Eds). *Teaching statistics in school mathematics. Challenges for teaching and teacher education: A joint ICMI/IASE study*. New York: Springer.

Borovcnik, M., & Kapadia, R. (2009). Special issue on “Research and Developments in Probability Education”. *International Electronic Journal of Mathematics Education*, *4*(3).

Borovcnik, M., & Peard, R. (1996). Probability. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 239–288). Dordrecht: Kluwer.

Carranza, P., & Kuzniak, A. (2008). Duality of probability and statistics teaching in French education. In C. Batanero, G. Burrill, C. Reading, & A. Rossman.

Chaput, B., Girard, J. C., & Henry, M. (2008). Modeling and simulations in statistics education. In C. Batanero, G. Burrill, C. Reading, & A. Rossman. *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education*. Monterrey: ICMI and IASE. Online: www.stat.auckland.ac.nz/~iase/publications

Davies, P. L. (2009). Einige grundsätzliche Überlegungen zu zwei Abituraufgaben (Some basic considerations to two tasks of the final exam). *Stochastik in der Schule*, *29*(2), 2–7.

Davies, L., Dette, H., Diepenbrock, F. R., & Krämer, W. (2008). Ministerium bei der Erstellung von Mathe-Aufgaben im Zentralabitur überfordert? (Ministry of Education overcharged with preparing the exam paper in mathematics for the centrally administered final exam?) *Bildungsklick*. Online: http://bildungsklick.de/a/61216/ministerium-bei-der-erstellung-von-mathe-aufgaben-im-zentralabitur-ueberfordert/

Díaz, C., & Batanero, C. (2009). University Students' knowledge and biases in conditional probability reasoning. *International Electronic Journal of Mathematics Education 4*(3), 131–162. Online: www.iejme.com/

Falk, R., & Konold, C. (1992). The psychology of learning probability. In F. Sheldon & G. Sheldon (Eds.). *Statistics for the twenty-first century*, MAA Notes 26 (pp 151–164). Washington DC: The Mathematical Association of America.

Feller, W. (1968). *An introduction to probability theory and its applications* (Vol. 1, 3rd ed.). New York: J Wiley.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: D. Reidel.

Garuti, R., Orlandoni, A., & Ricci, R. (2008). Which probability do we have to meet? A case study about statistical and classical approach to probability in students' behaviour. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (2008). *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education.* Monterrey: ICMI and IASE. Online: www.stat.auckland.ac.nz/~iase/publications

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.

Girard, J. C. (2008). The Interplay of probability and statistics in teaching and in training the teachers in France. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (2008) *Joint ICMI/IASE study: teaching statistics in school mathematics. Challenges for teaching and teacher education*. Monterrey: ICMI and IASE. Online: www.stat.auckland.ac.nz/~iase/publications

Goertzel, T. (n.d.). *The myth of the Bell curve*. Online: crab.rutgers.edu/~goertzel/normalcurve.htm

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology*, *3*, 430–454.

Kapadia, R., & Andersson, G. (1987). *Statistics explained. Basic concepts and methods*. Chichester: Ellis Horwood.

Kapadia, R., & Borovcnik, M. (1991). *Chance Encounters: Probability in education*. Dordrecht: Kluwer.

Lysø, K. (2008). Strengths and limitations of informal conceptions in introductory probability courses for future lower secondary teachers. In *Eleventh International Congress on Mathematics Education, Topic Study Group 13 "Research and development in the teaching and learning of probability"*. Monterrey, México. Online: tsg.icme11.org/tsg/show/14

Meyer, P. L. (1970). *Introductory probability and statistical applications reading*. Massachusetts: Addison-Wesley.

Ottaviani, M. G. (2008). The interplay of probability and statistics in teaching and in training the teachers. In C. Batanero, G. Burrill, C. Reading, & A. Rossman.

Pierce, R. (n.d.), Quincunx. Rod. In *Math is Fun – Maths Resources*. Online: www.mathsisfun.com/data/quincunx.html

Schulministerium NRW. (n.d.). Zentralabitur NRW. Online: www.standardsicherung.nrw.de/abitur-gost/fach.php?fach=2. Ex 7: http://www.standardsicherung.nrw.de/abitur-gost/getfile.php?file=1800

Manfred Borovcnik
Institute of Statistics
Alps-Adria University Klagenfurt
Austria

Ramesh Kapadia
Institute of Education
University of London
London
WC1H OAL
United Kingdom

ASTRID BRINKMANN AND KLAUS BRINKMANN

2. PROBLEMS FOR THE SECONDARY MATHEMATICS CLASSROOMS ON THE TOPIC OF FUTURE ENERGY ISSUES

INTRODUCTION

The students' interest and motivation in mathematics classroom towards the subject as a whole may be increased by using and applying mathematics. "The application of mathematics in contexts which have relevance and interest is an important means of developing students' understanding and appreciation of the subject and of those contexts." (National Curriculum Council 1989, para. F1.4). Such contexts might be, for example, environmental issues that are of general interest to everyone. Hudson (1995) states "it seems quite clear that the consideration of environmental issues is desirable, necessary and also very relevant to the motivation of effective learning in the mathematics classroom". One of the most important environmental impacts is that of energy conversion systems. Unfortunately this theme is hardly treated in mathematics education.

Dealing with this subject may not only offer advantages for the mathematics classroom, but also provide a valuable contribution to the education of our children. The younger generation especially, would be more conflicted with the environmental consequences of the extensive usage of fossil fuels, and thus a sustainable change from our momentary existing power supply system to a system based on renewable energy conversion has to be achieved. The decentralised character of this future kind of energy supply surely requires more personal effort of everyone and thus it is indispensable for young people to become familiar with renewable energies.

However, at the beginning of the 21th century there was a great lack of suitable school mathematical problems concerning environmental issues, especially strongly connected with future energy issues. An added problem is that the development of such mathematical problems requires the co-operation of experts in future energy matters, with their specialist knowledge, and mathematics educators with their pedagogical content knowledge.

The authors working in such a collaboration have developed a special didactical concept to open the field of future energy issues for students, as well as for their teachers, and this is presented below. On the basis of this didactical concept we have created several series of problems for the secondary mathematics classroom on the topics of rational usage of energy, photovoltaic, thermal solar energy, biomass, traffic, transport, wind energy and hydro power. The collection of worked out problems, with an extensive solution to each problem, has been published in a book in the

J. Maasz and J. O'Donoghue (eds.), Real-World Problems for Secondary School Mathematics Students: Case Studies, 45–66.

German language (Brinkmann & Brinkmann, 2005). Further problems dealing with so-called energy hybrid systems i.e., combinations of several energy types, will be developed (see Brinkmann & Brinkmann, 2009).

Some problem examples are presented in paragraph 3 of this article.

DIDACTICAL CONCEPT

The cornerstones of the didactical concept developed by the authors in order to promote renewable energy issues in mathematics classrooms are:

- The problems are chosen in such a way that the mathematical contents needed to solve them are part of mathematics school curricula.
- Ideally every problem should concentrate on a special mathematical topic such that it can be integrated into an existing teaching unit; as project-oriented problems referring to several mathematical topics are seldom picked up by teachers.
- The problems should be of a greater extent than usual text problems, in order to enable the students and also their teachers to concern themselves in a more intensive way with the subject.
- The problems should not require special knowledge of teachers concerning future energy issues and especially physical matters. For this reason all nonmathematical information and explanations concerning the problem's foundations are included in separate text frames.
- In this way information about future energy issues is provided for both teachers and students, helping them to concentrate on the topic. Thus, a basis for inter-disciplinary discussion, argumentation and interpretation is given.

EXAMPLES OF MATHEMATICAL PROBLEMS

The Problem of CO_2 Emission

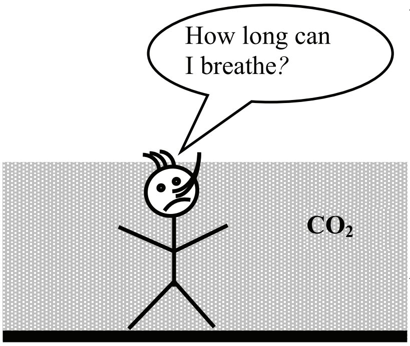
This is an inter-disciplinary problem linked to the subjects of mathematics as well as chemistry, physics, biology, geography, and social sciences. Nevertheless, it may be treated in lower secondary classrooms. With respect to mathematics the *conversion of quantities* is practised, knowledge of *rule of three* and *percentage calculation* is required. The *amount of* CO_2 *produced annually in Germany* especially by transport and traffic is illustrated vividly so that students become aware of it.

Information:

In Germany, each inhabitant produces an annual average of nearly 13 t of CO_2 (Carbon dioxide). Combustion processes (for example from power plants or vehicle combustion motors) are responsible for this emission into the atmosphere.

Assume now that this CO_2 would build up a gaseous layer which stays directly above the ground.

a) What height would this CO_2-layer reach in Germany after one year?



Hints:

- *Knowledge from chemical lessons is helpful for your calculations. There you learned that amounts of material could be measured with the help of the unit 'mole'.* 1 mole of CO_2 weighs 44 g and takes a volume of 22.4 l*, under normal standard conditions (pressure 1013 hPa and temperature 0°C). With these values you can calculate approximately.*
- *You will find the surface area and the number of inhabitants of Germany in a lexicon.*

Help: Find the answers of the following partial questions in the given order.

i) *How many tons of CO_2 are produced in total in Germany every year?*
ii) *What volume in* l *(litres) takes this amount of CO_2? (Regard the Hint!)*
iii) *How many m^3 of CO_2 are therefore produced annually in Germany? Express this in km^3!*
iv) Assume, the CO_2 produced annually in Germany forms a low layer of gas directly above the ground, what height would it have?

Information:

- In Germany the amount of waste is nearly 1 t for each inhabitant (private households as well as industry) every year, the average amount of CO_2 produced per inhabitant is therefore 13 times of this.
- The CO_2, which is produced during combustion processes and emitted into the atmosphere, distributes itself in the air. One part will be absorbed by the plants with the help of the photosynthesis, a much greater part goes into solution in the oceans' waters. But the potential for CO_2 absorption is limited.
- In the 1990s, 20% of the total CO_2-emissions in Germany came from the combustion engines engaged in traffic activities alone.

b) What height would the CO_2-layer over Germany have, if this layer results only from the annual emissions from individual vehicles? How many km^3 of CO_2 is this?

Usable Solar Energy

This problem deals with the heating of water in private households using *solar energy*. It can be treated in lessons involving the topic of *percent calculation* and *the rule of three*. It requires the *understanding and usage of data representations*.

Information:

In private households the warm water required can be partly heated up by solar thermal collectors. They convert the solar radiation energy in thermal energy. This helps us to decrease the usage of fossil fuels which leads to environmental problems.

Private households need heated water preferably in the temperature region 45°–55°C. Permanently in our region, the usable thermal energy from the sun is not sufficient to bring water to this temperature because of seasonal behavior. Thus, an input of supplementary energy is necessary.

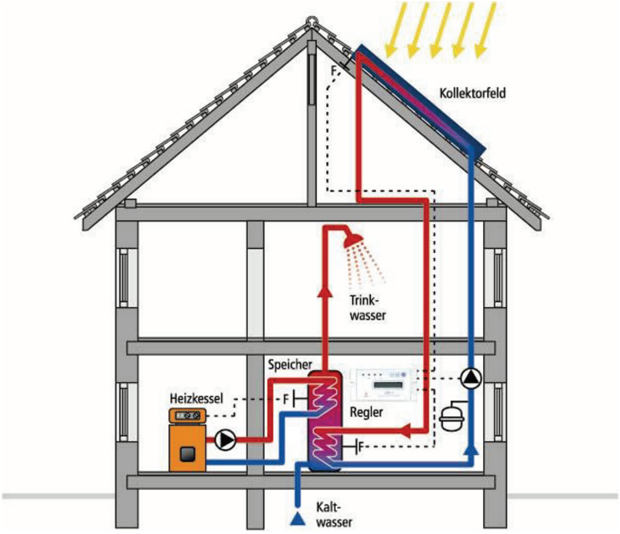


Figure 1. A solar thermal energy plant (source: DGS LV Berlin BRB).

The following figure (Figure 2) shows how much of the needed energy for heating up water to a temperature of 45°C in private households can be covered respectively by solar thermal energy and how much supplementary energy is needed.

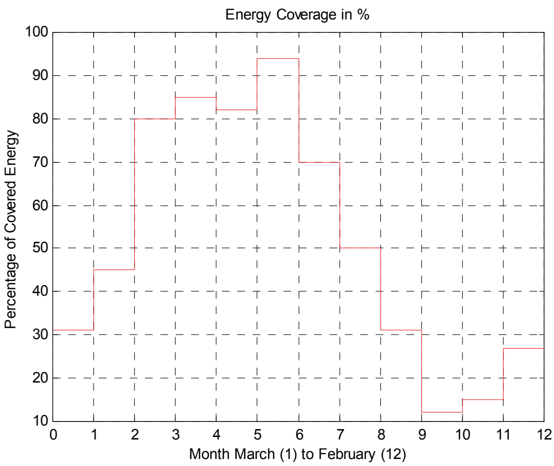


Figure 2. Usable solar energy and additional energy.

The following problems refer to the values shown by Figure 2.

a) What percent of the thermal energy needed for one year can be provided by solar thermal energy?

Information:

- Energy is measured by the unit kWh. An average household in central Europe consumes nearly 4000 kWh energy for the heating of water per year.
- 1 *l* fossil oil provides approximately 10 kWh thermal energy. The combustion of 1 *l* oil produces nearly 68 *l* CO_2.
- The great amount of CO_2 worldwide produced at present by the combustion of fossil fuels damages the environment.

b) How many kWh may be provided in one year by solar thermal energy?

c) How many litres of oil have to be bought to supply the supplementary thermal energy needed for one year for a private household?

d) How many litres of oil would be needed without solar thermal energy?

e) How many litres of CO_2 could be saved by an average household in Germany during one year by using Solar Collectors?

The Problem of Planning Solar Collector Systems

This problem deals with calculations for planning *solar collector systems* by using *linear functions*. The understanding and usage of *graphical representations* is performed.

Information:

- In private households, the heating of warm water can partly be done by solar collector systems. Solar collector systems convert radiation energy in thermal energy. This process is called solar heat.
- The usage of solar heat helps to save fossil fuels like natural gas, fuel oil or coal that damages the environment.
- Energy is measured by the unit kWh. An average household in central Europe consumes nearly 4000 kWh thermal energy per year for the heating warm water.
- Private households need heated water preferably in the temperature region 45°–55°C. Permanently at our latitude, the usable thermal energy from the sun is not sufficient to bring water to this temperature because of the seasonal behavior. Thus, an input of supplementary energy is necessary.
- The warm water requirement per day per person can be reckoned at about 50 *l*. By energy-conscious living this value can be easily reduced to 30 *l* per person and day.
- 1 *l* of fossil oil provides approximately 10 kWh of thermal energy. The combustion of 1 *l* of oil produces nearly 68 *l* of CO_2.

The diagram in Figure 3 provides data for planning a solar collector system for a private household. It shows the dependence of the collector area needed, for example on the part of Germany where the house is situated, on the number of persons living in the respective household, on the desired amount of warm water per day per person, as well as the desired output of thermal energy needed by solar thermal energy (in per cents).

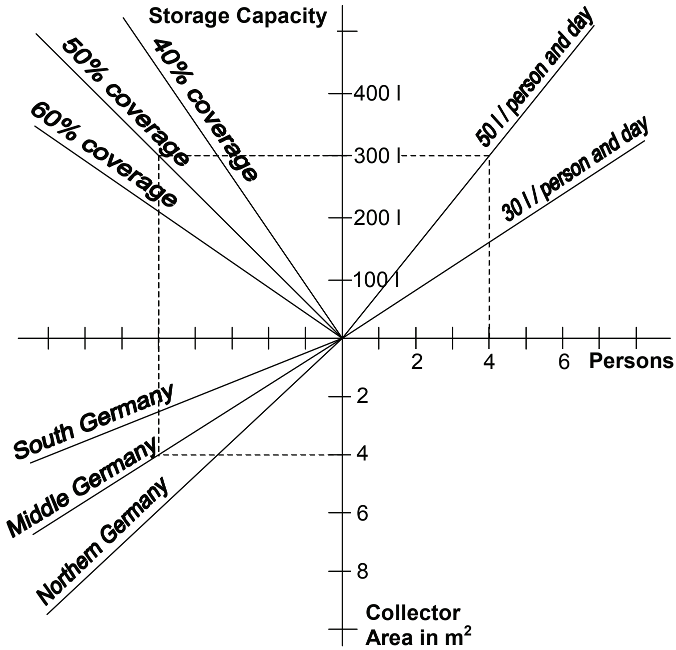


Figure 3. Dimensioning diagram.

Example: In a household in central Germany with 4 persons and a consumption of 50 l of warm water per day for each one, a collector area of 4 m^2 is needed for a reservoir of 300 l and an energy coverage of 50%.

a) What would be the collector area needed for the household you are living in? What assumptions do you need to make first? What would be the minimal possible collector area, what the maximal one?
b) A collector area of 6 m^2 that provides 50% of the produced thermal energy is installed on a house in southern Germany. How many persons could be supplied with warm water in this household?
c) Describe using a linear function the dependence of the storage capacity on the number of persons in a private household. Assume first a consumption of 50 l of warm water per day per person, and second a consumption of 30 l. Compare the two function terms and their graphical representation.
d) Show a graphical representation of the dependence of the collector area on a chosen storage capacity assuming a thermal energy output of 50% for a house in central Germany.

Sun Collectors

This problem can be integrated in lessons about *quadratic parabola* and uses their *focus property* as application in the field of *sun collectors*.

Information:

- Direct solar radiation may be concentrated in a focus by means of parabolic sun collectors (Figure 4). These use the focus property of quadratic parabola.
- sun collectors are figures with rotational symmetry, they evolve by rotation of a quadratic parabola. Their inner surface is covered with a reflective mirror surface; that is why they are named parabolic mirrors.
- Sun beams may be assumed to be parallel. Thus, if they fall on such a collector, parallel to its axis of rotation, the beams are reflected so that they all cross the focus of the parabola. The thermal energy radiation may be focused this way in one point.
- The temperature of a heating medium, which is lead through this point, becomes very high, relative to the environment. This is used for heating purposes, but also for the production of electric energy.

Figure 4. Parabolic sun collectors (source: DLR).

a) A parabolic mirror was constructed by rotation of the parabola $y = 0.125x^2$. Determine its focal length (x and y are measured in meters).
b) A parabolic mirror has a focal length of 8 m. Which quadratic parabola was used for its construction?
c) Has the parabolic mirror with $y = 0.0625x^2$ a greater or a smaller focal length than that one in b)? Generalize your result.
d) A parabolic mirror shall be constructed with a width of 2.40 m and a focal length of 1.25 m. How great is its arch, i.e., how much does the vertex lay deeper than the border?
e) In Figure 5 you see a parabolic mirror, the EuroDish with a diameter of 8.5 m. Determine from the figure, neglecting errors resulting from projection sight, its approximate focal length and the associated quadratic parabola.

Figure 5. EuroDish system (source: Wikimedia Commons).

Information:

Other focussing sun collectors are figures with length-symmetry, they evolve by shifting a quadratic parabola along the direction of one axis. They are named parabolic trough solar collectors (Figure 6).

Figure 6. Parabolic trough solar collectors in Almería, Spain and California, USA (source: DLR and FVEE/PSA/DLR).

f) The underlying function of a parabolic trough solar collector is given by $y = 0.35x^2$ (1 unit ≙ 1 m). Where has the heating pipe to be installed?

Photovoltaic Plant and Series Connected Efficiencies

The aim of this problem is to make students familiar with the *principle of series connected efficiencies*, as they occur in complex energy conversion devices. As an example, an off-grid *photovoltaic plant* for the conversion of solar energy to AC-current as a self-sufficient energy supply is considered. The problem can be treated in a teaching unit on the topic of *fractions*.

Figure 7 shows the components of an interconnected energy conversion system to build up a self-sufficient electrical energy supply. This kind of supply system is of special interest for developing countries, and also for buildings in rural off-grid areas (Figure 8).

Figure 7 shows in schematic form the production of electrical energy from solar radiation with the help of a solar generator for off-grid applications. In order to guarantee a gap-free energy supply for times without sufficient solar radiation, a battery as an additional storage device is included.

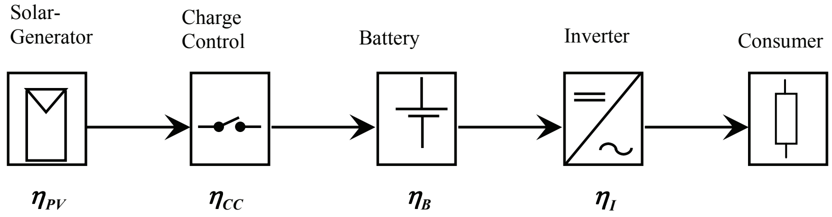


Figure 7. Off-Grid photovoltaic plant.

Figure 8. Illustration of an off-grid photovoltaic plant on the mountain hut "Starkenburger Hütte" (Source: Wikipedia).

Information:

The components of an off-grid photovoltaic (PV) plant are 1) a solar generator, 2) a charge control, 3) an accumulator and 4) an inverter (optional for AC-applications).
The solar generator converts the energy of the solar radiation into electrical energy as direct current (DC). The electricity is passed to a battery via a charge control. From there it can be transformed directly, or later after storage in the battery, to alternating current (AC), when it is needed by most of the electric devices.
Unfortunately, it is not possible to use the radiation energy without losses. Every component of the conversion chain produces losses, so only a fraction of the energy input for each component would be the energy input for the following component. The efficiency η of a component is defined by $\eta = \frac{\text{Power going out}}{\text{Power coming in}}$.
Power is the energy converted in 1 second. It is measured by the unit W or kW (1 kW = 1000 W). For comparison standard electric bulbs need a power of 40 W, 60 W or 100 W, a hair blower consumes up to 2 kW.

Assume in the tasks a), b) and c) that all the electric current is first stored in the battery before it reaches the consumer.

a) Consider that the momentary radiation on the solar generator would be 20 kW. Calculate the out going power for every component of the chain, if:

$$\eta_{PV} = \frac{3}{25},\ \eta_{CC} = \frac{19}{20},\ \eta_B = \frac{4}{5} \text{ and } \eta_I = \frac{23}{25}.$$

b) What is the total system efficiency $\eta_{total} = \frac{\text{gained power for the consumer}}{\text{insolated power}}$?

How can you calculate η_{total} by using only the values η_{PV}, η_{CC}, η_B and η_I? Give a formula for this calculation.

c) Transform the efficiency values given in a) into decimal numbers and percents. Check your result obtained in a) with these numbers.

d) How do the battery efficiency and the total system efficiency change, if only $1/3$ of the electric power delivered by the charge control would be stored in the battery and the rest of $2/3$ goes directly to the inverter? What is your conclusion from this?

Wind Energy Converter

This problem deals with *wind energy converters*. It can be treated in lessons on *geometry, especially calculations of circles* or in lessons on *quadratic parabola*. The *conversion of quantities* is practised.

Information:

The nominal power of a wind energy converter depends upon the rotor area A with the diameter D as shown in Figure 9 below.

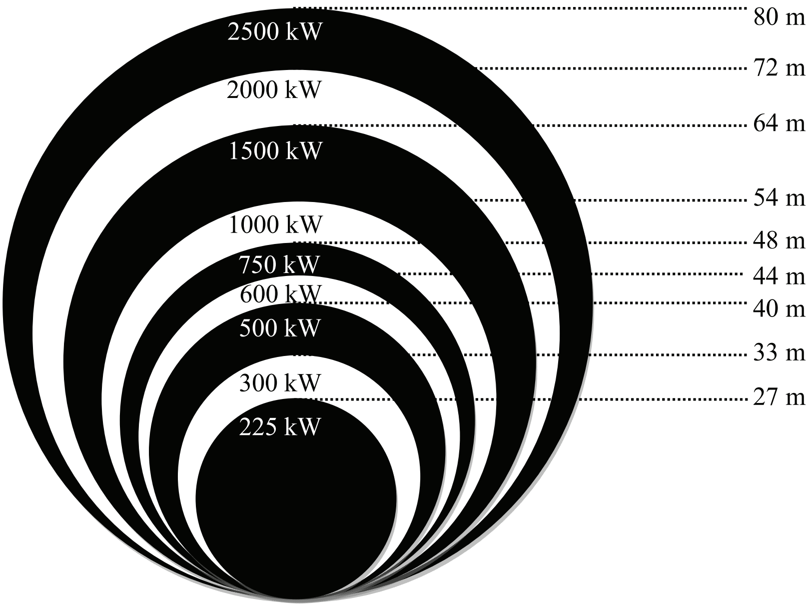


Figure 9. Nominal power related to the rotor area.

a) Interpret the meaning of Figure 9.
b) Show the dependence of the nominal power of the wind energy converter on the rotor diameter D and respectively on the rotor area A by graphs in co-ordinate systems.
c) Find the formula which gives the nominal power of the wind energy converter as a function of the rotor area and of the rotor diameter respectively.
d) What rotor area would you expect to need for a wind energy converter with a nominal power of 3 MW? Give reason for your answer. (Note: 1 MW = 1000 kW.)
   What length should the rotor blades have for this case?

Information:

- The energy that is converted in one hour [h] by the power of one kilowatt [kW] is 1 kWh.
- In central Europe, wind energy converters produce their nominal power on average for 2000 hours a year when wind energy conditions are sufficient.

e) Calculate the average amount of energy in kWh, which would be produced by a wind energy converter with a nominal power of 1.5 MW during one year in middle Europe.

Information:

- An average household in central Europe consumes nearly 4000 kWh electrical energy per year.

f) In theory, how many average private households in central Europe could be supplied with electrical energy by a 1.5 MW wind energy converter? Why do you think, that this could only be a theoretical calculation?
g) Assume the nominal power of a 600 kW energy converter would be reached at a wind speed of 15 m/s, measured at the hub height. How many km/h is this? How fast are the movements of the tips of the blades, if the rotation speed is 15/min. Give the solutions in m/s and km/h, respectively. Compare the result with the wind speed.

Wind Energy Development

This problem requires the *usage and interpretation of data and statistics* which is done in the context of *wind energy development in Germany*.

Information:

At the end of 1990 the installed wind energy converters in Germany had a total nominal power of 56 MW. At the end of 2000 this amount increased to a total of 6113 MW.
Power is the energy converted in a time unit; it is measured by the unit Watt [W]. 10^6 W are one Megawatt [MW].

The following table shows the development of the new installed wind power in Germany in the years 1991–2000.

Table 1. Development of new installed wind energy in Germany

Year	*Number of new installed wind energy converters*	*Total of new installed nominal power*	*Total of nominal power*
1991	300	48	
1992	405	74	
1993	608	155	
1994	834	309	
1995	911	505	
1996	804	426	
1997	849	534	
1998	1010	733	
1999	1676	1568	
2000	1495	1665	

a) Fill in the missing data in the 4th column.
b) Show the development of the annual new installed nominal power and of the total annual nominal power in graphical representations.

c) Considering only the data of the years 1991–1998, what development in respect to the installation of new wind power in Germany could be expected in your opinion? Give a well-founded answer!
Compare your answer with the real data given for 1999 and 2000 and comment on it.
What is your projection for 2005? Why?

d) Calculate using the data in Table 1 for each of the years 1991–2000, the average size of new installed wind energy converters in kW. (Note: 1 MW = 1000 kW.)
Show the respective development graphically and comment on it.
Can you offer a projection for the average size of a wind energy converter that will be installed 2010?

e) Comment on the graphical representation in Figure 10.
Also take into account political and economical statements and possible arguments.

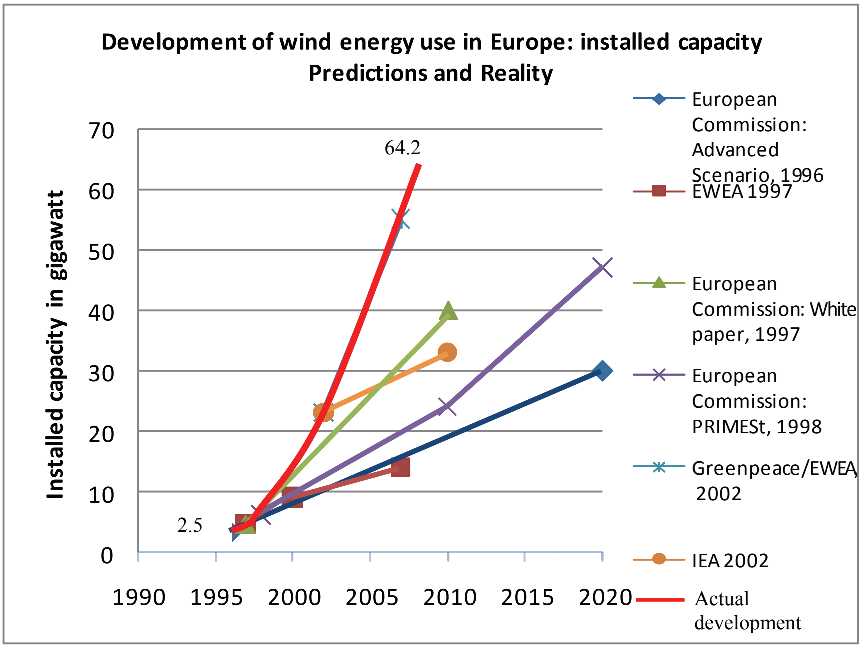


Figure 10. Development of wind energy use in Europe.

Betz' Law and Differentiation

This problem deals with the *efficiency of a wind energy converter*; it can be treated in lessons on *differentiation and the determination of local extreme values*.

Information:

2% of the radiated solar energy is converted to kinetic energy of air molecules. In combination with the earth's rotation, this results in a wind production. The kinetic energy of an air mass Δm is $E = 1/2 \cdot \Delta m \cdot v^2$, in which v denotes the velocity of the air mass. The kinetic energy can be written as $E = 1/2 \cdot \rho \cdot \Delta V \cdot v^2$ given the density of air $\rho = 1,2\text{g/l}$ and the relation $\Delta m = \rho \cdot \Delta V$ with the volume element ΔV. The power is defined as the ratio of energy to time as $P = E/\Delta t$.

a) A wind volume ΔV flows through a rotor area A and needs the time Δt to travel the distance Δs. Therefore the speed is $v = \Delta s/\Delta t$. Determine the general formula for the volume element which passes the rotor area A during the time interval Δt as a function of the wind speed.

b) Give the formula for the amount of wind power P_{wind}, which passes through the rotor area A as a function of the wind velocity.
Show that the power increases with the third power of the wind velocity.

Information:

A rotor of a wind energy converter with area A slows down the incoming wind speed from v_1 in front of the rotor to the lower speed v_2 behind the rotor (Figure 11). The wind speed in the rotor area itself can be shown to be the average of v_1 and v_2 i.e., $\bar{v} = (v_1 + v_2)/2$. The converted power is then given by:

$$P_c = P_1 - P_2 = \frac{1}{2} \cdot \rho \cdot \frac{\Delta V}{\Delta t} \cdot \left(v_1^2 - v_2^2\right).$$

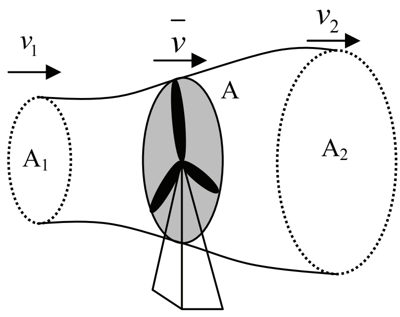


Figure 11. Wind flow through the rotor area.

c) Express the formula for the determination of P_c as a function of A, v_1 and v_2.

d) Describe the converted power P_c in c) as a function of the variable $x = v_2/v_1$ and v_1.

Information:

The efficiency of a wind energy converter (power coefficient) is defined as the ratio of the converted power to the wind power input as $c_P = P_c / P_{wind}$.

e) Express the power coefficient as a function of the variable $x = v_2/v_1$. Draw the graph of this function as a function of x . Note that $x \in [0,1]$, why?

f) Determine the value $x_{\max}$ which corresponds to the maximum value of the power coefficient, the so-called Betz' efficiency. This is the value for $x = v_2/v_1$ which gives the best energy conversion.

Biomass and Reduction of CO_2 Emissions

This problem deals with *fossil fuels* and *biomass*, especially with the production of *CO_2 emissions* and possibilities for their reduction. The *conversion of quantities* is practised, and knowledge of *rule of three* and *percentage calculation* is required.

Information:

In Germany for example, an average private household consumes nearly 18000 kWh of energy annually. 80% of this amount is for heating purposes and 20% for electrical energy. The energy demand for heating and hot water is mainly covered by the use of fossil fuels like natural gas, fuel oil or coal.

Assume that the calorific values of gas, oil and coal can be converted to useable heating energy with a boiler efficiency of 85%. This means that 15% is lost in each case.

a) The following typical specific calorific values are given

Natural gas: 9.8 kWh/m³
Fuel oil: 11.7 kWh/kg
Coal: 8.25 kWh/kg
(That is, the combustion of 1 m^3 natural gas supplies 9.8 kWh, 1 kg fuel oil supplies 11.7 kWh and 1 kg coal supplies 8.25 kWh.)
What amount of these fuels annually is necessary for a private household in each case?

b) The specific CO_2-emissions are approximately:

Natural gas: 2.38 kg CO_2/kg
Fuel oil: 3.18 kg CO_2/kg
Coal: 2.90 kg CO_2/kg
The density of natural gas is nearly 0.77 kg/m³.
How many m³ of CO_2 each year for a private household in Germany does it take in each case?
Hint: Amounts of material could be measured with the help of the unit 'mole'. 1 mole of CO_2 weights 44 g and has a volume of approximately 22.4 *l*.

Information:

Wood is essentially made with CO_2 taken from the atmosphere and water. The bound CO_2 is discharged by burning the wood and is used again in the building of plants. This is the so-called CO_2-circuit.
Spruce wood has a specific calorific value of nearly 5.61 kWh/kg. The specific CO_2-emissions are approximately 1.73 kg CO_2/kg.

c) How many kg of spruce wood would be needed annually for a private household instead of gas, oil or coal? (Assume again a boiler efficiency of 85%).
How many m³ of fossil CO_2-emissions could be saved in this case?

Information:

Spruce wood as piled up split firewood has a storage density of 310 kg/m³.

d) How much space has to be set aside in an average household for a fuel storage room, which contains a whole year's supply of wood?
Compare this with your own room!
e) Discuss the need for saving heat energy with the help of heat reduction.

Automobile Energy Consumption

This problem can be treated in lessons on *trigonometry*. Its solution requires knowledge of the *rule of three*. The problem makes clear the *dependence of an automobile's energy consumption* on the distance-height-profile, the moved mass and the velocity.

Tim and Lisa make a journey through Europe. Just before the frontier to Luxembourg their fuel tank is empty. Fortunately they have a reserve tank filled with 5 *l* fuel. "Let's hope it will be enough to reach the first filling station in Luxembourg. There, the fuel is cheaper than here" Tim says. "It would be good if we had an exact description of the route, than we would be able to calculate our range", answers Lisa.

Information:

– In order to drive, the resisting forces have to be overcome. Therefore a sufficient *driving force* F_{drive} is needed. For an average standard car, the law for this force (in N) is given by the following formula:
$F_{drive} = (0.2 + 9.81 \cdot \sin\alpha) \cdot m + 0.3 \cdot v^2$ for $F_{drive} \geq 0$,
where m the moving mass (in kg) is the mass of the vehicle, passengers and packages; v is the velocity (in m/s), and α is the angle relative to the horizontal line. α is positive for uphill direction and negative in the downhill case (Figure 12).
– The energy E (in Nm) which is necessary for driving, can be calculated in cases of a constant driving force by: $E = F_{drive} \cdot s$, with s as the actual distance driven (in m).
– The primary energy consisting of the fuel amounts to about 9 kWh for each 1 *l* of fuel. (kWh is the symbol for the energy unit 'kilowatt-hours'; it is 1 kWh = 3 600 000 Nm).

– The efficiency of standard combustion engines in cars for average driving conditions is between 10% and 20% nowadays; this means only 10%–20% of the primary energy in the fuel is available to generate the driving forces.

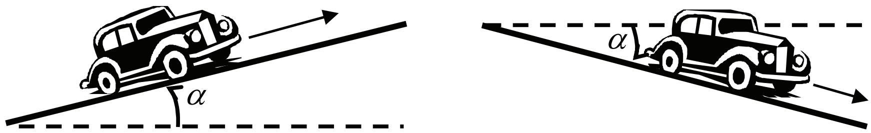


Figure 12. Definition of the angle α.

The distance which Tim and Lisa have to drive to the first filling station in Luxembourg can be approximately given by a graphical representation like that one given in Figure 13. (Attention: think of the different scaling of the co-ordinate axis). The technical data sheet of their vehicle gives the unladen weight of their car as about 950 kg. Tim and Lisa together weigh c. 130 kg, and their packages nearly 170 kg. The efficiency of the engine can be assumed to be c. 16%.

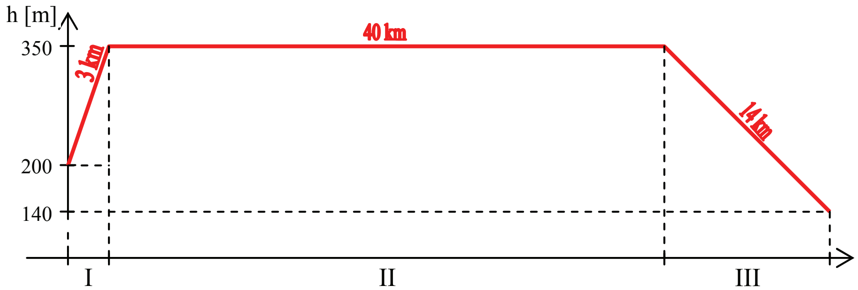


Figure 13. Distance-Height-Diagram to the next filling station (h is the height above mean sea level).

a) Can Tim and Lisa take the risk of not searching for a filling station before the frontier? Assume at first, the speed they drive is 100 km/h.
b) Would Tim and Lisa have less trouble, if they had only 50 kg packages instead of 170 kg?
c) Would the answer to a) change if Tim and Lisa chose their speed to be only 50 km/h?

Help for a):

i) Note, the speed in the formula for F_{drive} has to be measured in m/s.
ii) The value for $\sin\alpha$ can be calculated with the information given in Figure 13.
iii) Determine for each section the force F_{drive} and the energy needed. The distance s has to be measured in m. Convert the energy from Nm in kWh.

iv) Determine the total energy which is needed for the whole distance as a sum over the three different sections.
v) How many kWh of energy to drive are given by the 5 l reserve fuel? Consider the efficiency of the motor.

Automobiles: Forces, Energy and Power

This is a problem that can be treated in higher secondary mathematics education, in the context of *differential and integral calculus*. This problem shows the *dependence of an automobile's power and energy consumption* on the distance-height-profile, the moved mass and the velocity.

Kay has a new electric vehicle, of which he is very proud. He wants to drive his girlfriend Ann from the disco to her home. Ann jokes: "You will never get over the hill to my home with this car!" "I bet that I will", says Kay.

The distance-height-characteristic of the street from the Disco (D) to the house (H), in which Ann lives, is shown in Figure 14, and it can be described by the following function:

$$h(x) = \frac{1}{1\ 000} \cdot (-4x^2 + 104x + 300) \ \text{ for } \ x \in [0;\ 20],$$

where x and h are measured in kilometres (km).

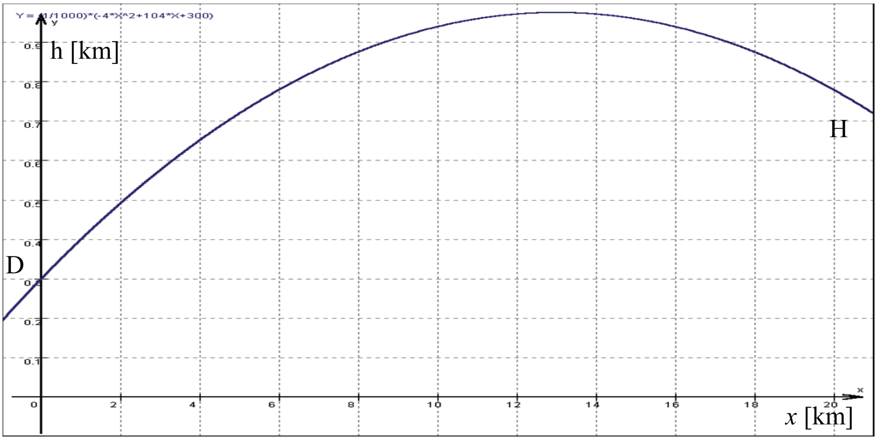


Figure 14. Distance-Height-Diagram between D and H.

a) Show, that it is possible to calculate the real distance s depending of a given height function h over the interval $[x_1, x_2]$ with the help of the following formula:

$$s = \int_{x_1}^{x_2} \sqrt{1 + (h'(x))^2}\, dx\,.$$

Help: Consider the right angle triangle as shown in Figure 15.

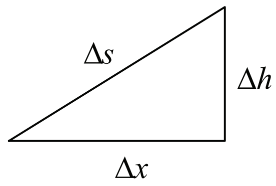


Figure 15. Geometrical representation.

b) How long is the real distance which Kay has to drive from the Disco to the house where Ann is living?
Help: Show that

$$\int \sqrt{1+x^2}\,dx = \frac{1}{2}(x\sqrt{1+x^2} + \ln(x+\sqrt{1+x^2})) + \text{const.}$$

with the help of the derivative of $\frac{1}{2}(x\sqrt{1+x^2} + \ln(x+\sqrt{1+x^2}))$.

(Hint: This partial result is helpful for solving the problem f).)

c) α *means the angle of the tangent to the curve* h *at the point* x_0 *on the* x*-axis. Prove that*

$$\sin\alpha = \frac{h'(x_0)}{\sqrt{1+(h'(x_0))^2}}.$$

(Note that $h'(x_0) = \tan\alpha$ *.)*

d) *Assume Kay wants to drive at a constant speed of 110* $\frac{\text{km}}{\text{h}}$.

Determine the driving power necessary at the top of the hill (maximum of h *) and at the points with* $h(x) = 0.4$ *and* $h(x) = 0.8$.

For this purpose you need the following data: Kay's electric vehicle has an empty weight of 620 kg, Kay and Ann together weigh nearly 130 kg.

Information:

– In order to drive, the resisting forces have to be overcome. Therefore a sufficient *driving force* F_{drive} is needed. For an average standard car, the law for this force (in N) is given by the following formula:

$$F_{drive} = (0.2 + 9.81\cdot\sin\alpha)\cdot m + 0.3\cdot v^2 \text{ for } F_{drive} \geq 0,$$

where m the moving mass (in kg), is the mass of the vehicle, passengers and packages, v is the velocity (in $\frac{\text{m}}{\text{s}}$), and α is the angle relative to the horizontal line. α is positive for uphill direction and negative in the downhill case (Figure 16).

– The driving power P (in $\frac{\text{Nm}}{\text{s}}$), which is needed to hold the constant speed, can be calculated using the product of the driving force (in N) and the velocity (in $\frac{\text{m}}{\text{s}}$):

$P = F_{drive} \cdot v$.

The power P is measured with the unit [kW], with: $1\text{ kW} = 1\,000\ \frac{\text{Nm}}{\text{s}}$.

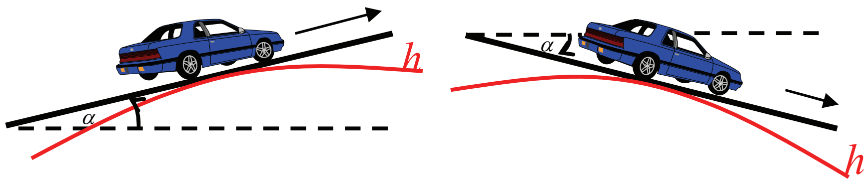


Figure 16. Angle α dependant on the function of $h(x)$.

e) Kay's electric vehicle has a nominal power of 25 kW. Is it possible for him to bring Ann home?

f) Determine the driving energy which has to be consumed for the route from the disco to Ann's home. Assume that Kay drives uphill with a speed of 80 $\frac{\text{km}}{\text{h}}$ and downhill with a speed of 110 $\frac{\text{km}}{\text{h}}$. (Attention: Because $h(x)$ as well as x are expressed in km in the function equation of h, the resulting energy in the following equation is obtained in $\text{N} \cdot \text{km} = 1\,000\text{ Nm}$.)

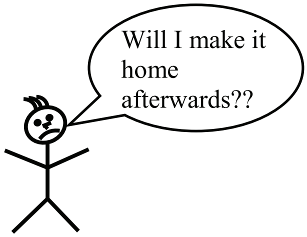
Information:

– The pure driving energy E (in Nm), which is necessary for driving, can be calculated as:

$E = \int_0^{s_0} F_{drive} ds = \int_0^{x_0} F_{drive} \sqrt{1 + (h'(x))^2}\, dx$, where s is the actual distance (in m) driven.

– The energy E is usually measured in kilowatt-hour [kWh]. $1\text{ kWh} = 3\,600\,000\text{ Nm}$.

g) The actual charged electrical energy in the batteries of Kay's vehicle is 6 kWh. The driving efficiency of his electrical vehicle is nearly 70%; this means only 70% of the stored energy can be used for driving.
Is the charging status of Kay's batteries sufficient to bring Ann to her home, under the assumptions in f)?



CLASSROOM IMPLEMENTATION

The problems on environmental issues developed by the authors must be seen as an offer for teaching material. In each case, the students' abilities have to be considered. In lower achieving classes it might be advisable not to present every problem in full length. In addition, lower achievers need a lot of help for solving complex problems that require several calculation steps. The help given in some problems, like in example 1 above, addresses such students.

The problems should be presented to higher achievers without much help included. It might even be of benefit not to present the given hints from the beginning. Students would thus have to find out, which quantities are yet needed in order to solve the problem. The problem would become more open and the students would be more involved in modelling processes.

As the intention of the authors is also an informal one, in order to give more insight in the field of future energy issues, the mathematical models/formulas are mostly given in the problem texts. Students are generally not expected to find out by themselves the often complex contexts; these are already presented, thus guaranteeing realistic situation descriptions. The emphasis in the modelling activities lies rather in the argumentation and interpretation processes demanded, recognising that mathematical solutions lead to a deeper understanding of the studied contents.

In the context of an evaluation of lessons dealing with problems like those presented in this paper, students amongst others were asked to express what they have mainly learned. The given answers can be divided in three groups: mathematical concepts, contents concerning renewable energy topics, as well as convenient problem solving strategies. As regards the last point, students stressed especially that they had learned that it is necessary to read the texts very carefully, and also to consider the figures and tables very carefully. Almost all students expressed, that they would like to work on much more problems of this kind in mathematical classes, as the problems are interesting, relevant for life, and are more fun than pure mathematics.

Classroom experiences show that students react in different ways to the problem topics. While some are horrified by recognizing for example that the worldwide oil reserves are already running low during their life time, others are unmoved by this fact, as twenty or forty years in the future is not a time they worry about. In school lessons there are again and again situations in where students drift away in political and social discussions related to the problem contexts. Although desirable, this would sometimes lead to too much time loss for mathematical education itself. Cooperation with teachers of other school subjects would be profitable if possible.

OUTLOOK AND FINAL REMARKS

In order to integrate future energy issues into curricula of public schools, several initiatives have already been started in Germany, supported and in co-operation with the 'Deutsche Gesellschaft für Sonnenergie e.V. (DGS)', the German section of the ISES (International Solar Energy Society).

There exists a European project, named "SolarSchools Forum", that aims to integrate future energy issues into curricula of public schools. In the context of this project the German society for solar energy DGS highlights the teaching material the authors created (http://www.dgs.de/747.0.html). Most of this material is only available in German language. This article is a contribution towards making these materials accessible in English also. (The English publications up to now (see e.g. Brinkmann & Brinkmann, 2007) present only edited versions of some of the problems.)

Although the education in the field of future energy issues is of general interest, the project that we presented in this paper seems to be the only major activity focusing especially on *mathematics* lessons. The amount of problems should thus be increased, especially with problems which deal with a combination of different renewable energy converters, like hybrid systems, to give an insight into the complexity of system technology.

Additionally, the sample mathematical problems on renewable energy conversion and usage have to be permanently adjusted to actual and new developments because of the dynamic evolution of the technology in this field.

REFERENCES

Brinkmann, A., & Brinkmann, K. (2005). *Mathematikaufgaben zum Themenbereich Rationelle Energienutzung und Erneuerbare Energien*. Hildesheim, Berlin: Franzbecker.

Brinkmann, A., & Brinkmann, K. L. (2007). Integration of future energy issues in the secondary mathematics classroom. In C. Haines, P. Galbraith, W. Blum, & S. Chan (Eds.), *Mathematical modelling (ICTMA 12): Education, engineering and economics* (pp. 304–313). Chichester: Horwood Publishing.

Brinkmann, A., & Brinkmann, K. (2009). Energie-Hybridsysteme – Mit Mathematik Fotovoltaik und Windkraft effizient kombinieren. In A. Brinkmann & R Oldenburg, (Eds.). *Schriftenreihe der ISTRON-Gruppe. Materialien für einen realitätsbezogenen Mathematikunterricht*, Band 14 (pp. 39–48), Hildesheim, Berlin: Franzbecker.

Hudson, B. (1995). Environmental issues in the secondary mathematics classroom. *Zentralblatt für Didaktik der Mathematik*, *27*(1), 13–18.

National Curriculum Council. (1989). *Mathematics non-statutory guidance*. York: National Curriculum Council.

Astrid Brinkmann
Institute of Mathematics Education
University of Münster, Germany

Klaus Brinkmann
Umwelt-Campus Birkenfeld
University of Applied Science Trier, Germany