Exercises in class - 1

for the lecture “Statistical Physics”, Master course “Computational Science”, year 2007/08

Exercise 1

In a famous German lottery each week 6 balls are drawn from an urn containing 49 balls, labeled with the numbers from 1 to 49. (We'll forget about the “Zusatzzahl”!) No number can be drawn twice, since the balls which have been picked are not put back into the urn.

1. What are the chances of getting n numbers right? Give:

   a) The exact formula.

   b) The numerical answer for $n = 0 \ldots 6$.

2. Let's assume that if you get all 6 numbers right, you'll get 1 000 000 Euro. If you get 5 right you'll get 10 000 Euro, if you get 4 right you'll get 100 Euro and if you get 3 right you'll get 10 Euro. What's the expected gain per game? Is it worthwhile to play if one game costs 1 Euro?

Exercise 2

What is the probability that among n randomly chosen people at least two have their birthday on the same day? (We know that birth rates are not quite constant throughout the year, and years are not equal in length. But for the purpose of this Exercise think of all years having 365 days and the chance for anyone being born on any of these days is the same.) Hence, how many people have to be together such that the chance of at least two of them having their birthday on the same day is at least 50%?

Exercise 3

Think of some random experiment whose outcomes can be classified as “success” and “failure” (for instance “hitting bulls eye” when playing darts, or “throwing a double with two dice”. Let the success probability be p. The probability for failure is then $q = 1 - p$.

1. What is the probability $B(k; n, p)$ of having exactly k successes when performing the experiment n times? Think about the chances of “hitting” k times and failing $n - k$ times, and then observe that it does not play any role in which order the k successes occur!

2. Let $\boldsymbol{X}$ be the random variable which measures the number of successes. Show that the expectation value of $\boldsymbol{X}$ is given by

$$\langle \boldsymbol{X} \rangle \; = \; np \; .$$

3. Show that the variance of $\boldsymbol{X}$ is given by

$$\sigma^2_{\boldsymbol{X}} \; = \; \langle \boldsymbol{X}^2 \rangle - \langle \boldsymbol{X} \rangle^2 \; = \; npq \; = \; np(1-p) \; .$$

Hints:

1. *Recall the Binomial formula* $(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$.

2. *There's something which is called "the geometric series". It is* $S = 1 + q + q^2 + q^3 + q^4 + \cdots + q^n$. *It can be calculated by the following neat trick:*

$$\begin{aligned} S &= 1 + q \; + q^2 + q^3 + q^4 + \cdots + q^n & (1) \\ qS &= q + q^2 + q^3 + q^4 + q^5 + \cdots + q^{n+1} & (2) \end{aligned}$$

If we now subtract line (1) from line (2), almost everything cancels, and we find

$$S(q-1) \; = \; qS - S = q^{n+1} - 1 \qquad \Longrightarrow \qquad S = \frac{1-q^{n+1}}{1-q} \; . \tag{3}$$

If $|q| < 1$, *we know that* $\lim_{q\to\infty} q^n = 0$. *Hence, using this and Eqn. (3) we get the following formula for the* infinite geometric series

$$\sum_{n=0}^{\infty} q^n \; = \; \frac{1}{1-q} \qquad \textit{if } \; |q| < 1 \; . \tag{4}$$

3. *The derivative of the general exponential function is the following:*

$$\frac{d}{dx} q^x \; = \; \frac{d}{dx} \mathrm{e}^{\ln(q)\,x} \; = \; \ln(q) \mathrm{e}^{\ln(q)\,x} \; = \; \ln(q)\, q^x \tag{5}$$

4. *The trick of parameter-differentiation is sometimes* very *useful: It consists of writing one variable as the derivative of some expression with respect to some other variable. This can be useful if one is summing over this variable, but not over the variable with respect to which one differentiates:*

$$\sum_{k=0}^{\infty} k \, \mathrm{e}^{-ak} = \sum_{k=0}^{\infty} \Big(- \frac{\partial}{\partial a} \Big) \, \mathrm{e}^{-ak} = - \frac{\partial}{\partial a} \sum_{k=0}^{\infty} \mathrm{e}^{-ak} \overset{(4)}{=} - \frac{\partial}{\partial a} \frac{1}{1-\mathrm{e}^{-a}} = \frac{\mathrm{e}^{-a}}{\big(1-\mathrm{e}^{-a}\big)^2} \; . \tag{6}$$

This trick is so useful that one sometimes "invents" an additional parameter a with respect to which one differentiate—and at the end of the game one sets it equal to 1*!*

Exercise 4

Let $\boldsymbol{X}$ be a random variable with probability distribution $w_{\boldsymbol{X}}(x)$, expectation value $\mu = \langle \boldsymbol{X} \rangle$, and variance $\sigma^2 = \langle \boldsymbol{X}^2 \rangle - \langle \boldsymbol{X} \rangle^2$. We want to prove *Chebyshev's inequality*

$$P\big(|\boldsymbol{X} - \mu| \geq \varepsilon\big) \;\leq\; \frac{\sigma^2}{\varepsilon^2} \,.$$

In words: "The probability that a particular outcome of the random experiment is farther away from the expectation value than some distance ε is smaller than the variance divided by ε^2." This gives us some feeling of what the variance actually measures: The "spread" of a distribution around its mean, or, how likely it is to get far away from the mean.
Proceed for the proof as follows:

1. Define the *unit step function* $\Theta(x)$ according to

$$\Theta(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad x \geq 0 \end{cases} \,.$$

   Plot the graph of $\Theta(x)$ to see what it looks like.

2. Now convince yourself (and the corrector), for instance by drawing a *clear* picture of the graph of the left hand side and the right hand side, that the following inequality holds:

$$\Theta\big(|x - \mu| - \varepsilon\big) \;\leq\; \frac{(x-\mu)^2}{\varepsilon^2}$$

3. Show now that you obtain Chebyshev's inequality by averaging this inequality over the probability distribution $w_{\boldsymbol{X}}(x)$ of the random variable $\boldsymbol{X}$. Recall that this means "putting angular brackets around", and that this formally means $\langle \cdots \rangle = \sum_x (w_{\boldsymbol{X}}(x) \cdots)$.

Comment: Chebyshev's inequality is of great generality, but because of that its upper bound for the probability $P(|\boldsymbol{X} - \mu| \geq \varepsilon)$ is unfortunately not particularly good.