Statistical Models and Data Analysis

Summer term 2018

Problem Set 10 **25.6.2018**

Please submit these exercises by 2 pm on 2.7.2018. You can send me email at stemmler@bio.lmu.de if you have questions.

---

1. (hypothesis testing, generating functions) Imagine that you are recording from two neurons that are synaptically connected. When you stimulate the first neuron, the second neuron will (sometimes) spike. You hypothesize that every stimulus has a likelihood of $p = 0.5$ of eliciting a postsynaptic spike, and that there are no history effects (i.e., the likelihood p does not change, regardless of whether the previous stimulus was successful or not.)

You record the number of times you need to stimulate until you get the first postsynaptic spike. This will be a random variable, which we will call X. The data read

number of stimuli	1	2	3	4	5	6
observed frequency	33	12	5	2	0	1

(a) Find the probability density function of X, given the hypothesis outlined above.

(b) For a discrete probability distribution,

$$G_X(t) = \sum P(X = x)t^x \tag{1}$$

is called the generating function. Show that, in our case, the generating function is

$$G_X(t) = \frac{pt}{1 - (1 - p)t}.$$

(c) From Eq. (1), without plugging in any specific distribution, show that

$$\begin{aligned} \mathbb{E}(X) =& G'_X(1) \\ \mathrm{var}(X) =& G''_X(1) + G'_X(1) - \left(G'_X(1)\right)^2 \end{aligned}$$

Hence, for the specific probability model at hand, compute the mean and variance of X as a function of p. With $p = 0.5$, compare the model mean and variance to the unbiased estimates of the mean and variance from the data.

(d) Compute the expected frequencies for $X = 1, 2, 3$ and $X \geq 4$, given the probability model.

(e) Compute the probability of the observed data, given the model.

(f) In the lecture, the $t-$statistic was covered. Another common statistic is χ^2, which is defined as follows: if the expected frequencies are E_i and the observed frequencies are O_i, then

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Compute the χ^2 value for the data and the given probability model. Lump all data for $X \geq 4$ together first.

(g) Look up the χ^2 distribution with three degrees of freedom in a table. At the 5% level of statistical significance, can one reject the probability model? What values of p would be consistent with the data at the 5% level?