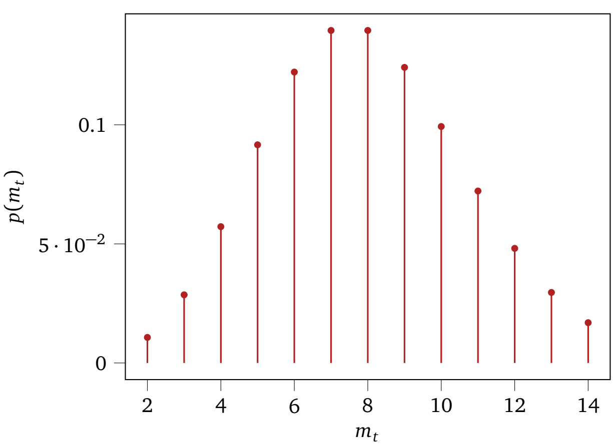
Exercise Sheet #4

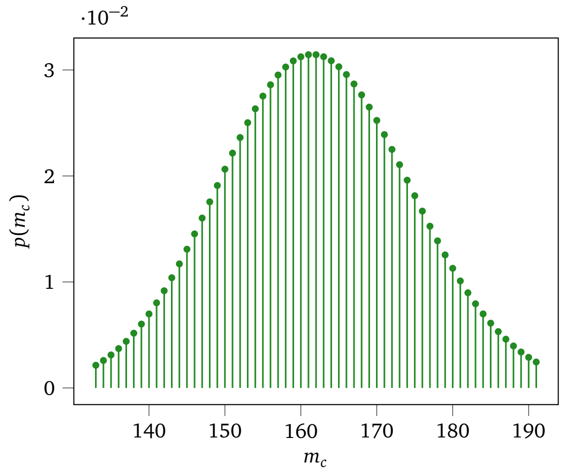
due on Tuesday, 01 December 2020, 12 noon sharp

1. **EXAMple** Im the lecture we encountered (not for the first time) the **binomial distribution**

$$p(m \mid f, n) = \binom{n}{m} f^{m} \cdot (1-f)^{n-m},$$

which provides the probability distribution for the number m of "successes" in a total of n independent Bernoulli trials, in which the positive outcome has probability f (and the negative one $(1-f)$). Given that you have observed m successes out of n trials, which value for f is the *most likely* one?





2. This exercise is a "pedestrian" approach to confidence estimation, to prepare a formal discussion in the lecture. On 18 Nov 2020, BioNTech and Pfizer announced that their experimental Covid vaccine is "95% effective". In such studies, the *vaccine efficacy E* is defined as the relative reduction in risk of contracting the disease:

$$E := \frac{\Delta p}{p_c} =: \frac{p_c - p_t}{p_c} = 1 - \frac{p_t}{p_c}$$

where p_t is the probability to contract the virus after treatment and p_c is the probability in the control group, i.e. for people who have not been treated. The study involved 44 000 participants, half of which where in the treatment and control groups, respectively. Of those, $m_t = 8$ and $m_c = 162$ people contracted the virus in the two groups, respectively.

   (a) Re-use the result from Exercies 4.1 above to estimate the *most likely* values $\hat{p}_c, \hat{p}_t$ of p_c and p_t, respectively. Plug these numbers into the formula for E above and see if they confirm the headline effectiveness.

   (b) Since the trial only contained a limited number of participants, the obvious question is how *confident* we can be that the 95% *point-estimate* is correct. If we assume that the estimates $\hat{p}_c, \hat{p}_t$ are correct, the two plots above show the binomial distributions over outcomes in the two groups (in each case, the plot covers the range from the 1% quartile to the 99% quartile). We see that "even if the estimates are correct, it wouldn't have been too unlikely to see $m_c = 135$ or $m_c = 190$, and $m_t = 2$ or $m_t = 14$". Use these rough bounds to construct an ad-hoc *confidence range* in which the true effect E may lie. (You may find that your answer corresponds to another, earlier news release about the same vaccine). **MORE CONTENT OVERLEAF!**

On 23 November 2020, the British pharma company AstraZeneca published a press release about *their* vaccine AZD1222. It contains the quote

> One dosing regimen ($n = 2,741$) showed vaccine efficacy of 90% when AZD1222 was given as a half dose, followed by a full dose at least one month apart, and another dosing regimen ($n = 8,895$) showed 62% efficacy when given as two full doses at least one month apart. The combined analysis from both dosing regimens ($n = 11,636$) resulted in an average efficacy of 70%. All results were statistically significant ($p \leq 0.0001$).

(c) It may seem puzzling that the *higher* dosage seemingly had a *weaker* effect (the PI, Andrew Pollard, spins this as a good thing: "Excitingly, weve found that one of our dosing regimens may be around 90% effective and if this dosing regime is used, more people could be vaccinated with planned vaccine supply"). Your final, hardest task in this exercise is to read the press release closely (click on link above) and build your own hypothesis for how such a result *could* have arisen in a way that is *not* related to the dosage. (Hint: Read the *entire* press release, *closely*.) If the link should stop working, you can also find the press release as a pdf on Ilias.

Disclaimer: This exercise is based on a press release, not a scientific publication. The PR is missing some crucial numbers, which may be published over the coming days. The goal of this task is to do some advanced critical thinking with statistics. It is not intended to dismiss AstraZeneca's vaccine, or vaccines in general. Quite in the contrary: the studies discussed on this sheet provide strong evidence for the efficacy of vaccines.

3. **Practical Question:** You can find this week's sheet on Ilias as `Exercise_04.ipynb`