Lab 2

# Readings:

The source of the exercises proposed for this meeting is Davidson-Pilon (2016), Chapter 1, a mandatory reading for this lab. The freely available version of the chapter can be found on Github (2017).

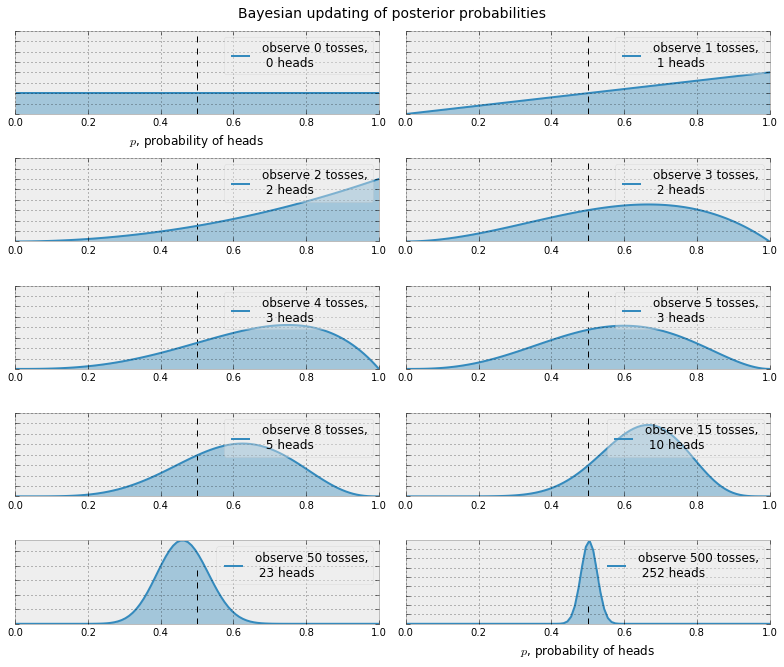
Frequentist (static analysis based on descriptive statistics – frequencies)

vs. Bayesian logics (updated results adjusted based on continuing data observation)

Code from Davidson-Pilon (2016), Chapter 1 (see it on Github, 2017):

|  |
| --- |
| %matplotlib inline  **from** **IPython.core.pylabtools** **import** figsize  **import** **numpy** **as** **np**  **from** **matplotlib** **import** pyplot **as** plt  figsize(11, 9)  **import** **scipy.stats** **as** **stats**  dist = stats.beta  n\_trials = [0, 1, 2, 3, 4, 5, 8, 15, 50, 500]  data = stats.bernoulli.rvs(0.5, size=n\_trials[-1])  x = np.linspace(0, 1, 100)  *# For the already prepared, I'm using Binomial's conj. prior.*  **for** k, N **in** enumerate(n\_trials):  sx = plt.subplot(len(n\_trials) / 2, 2, k + 1)  plt.xlabel("$p$, probability of heads") \  **if** k **in** [0, len(n\_trials) - 1] **else** **None**  plt.setp(sx.get\_yticklabels(), visible=**False**)  heads = data[:N].sum()  y = dist.pdf(x, 1 + heads, 1 + N - heads)  plt.plot(x, y, label="observe **%d** tosses,**\n** **%d** heads" % (N, heads))  plt.fill\_between(x, 0, y, color="#348ABD", alpha=0.4)  plt.vlines(0.5, 0, 4, color="k", linestyles="--", lw=1)  leg = plt.legend()  leg.get\_frame().set\_alpha(0.4)  plt.autoscale(tight=**True**)  plt.suptitle("Bayesian updating of posterior probabilities",  y=1.02,  fontsize=14)  plt.tight\_layout() |

As you will see on the following output, while the data are gathered, the distribution changes. The more observations are included, the distribution narrows and gets around the expected one (0.5).



Source: Davidson-Pilon (2016)/ Github (2017).

Go further with testing the code in Chapter 1. Please take a look at the differences between distributions which are call posterior based on data update.

**Questions (only one answer is correct):**

**Q1. Which of the following sentences is true?**

1. **Posterior distributions are included in the frequentist logic.**
2. **Posterior distributions are dependent on updated data.**
3. **Posterior distributions are always equal to a priori distribution.**

**Q2. An example of a continuous case is:**

1. **The probability that someone`s birthday is on October 10.**
2. **The probability to rain daily.**
3. **The frequency of individuals by country in a sample of European data.**

**Q3. Which of the following sentences is true?**

1. **If the mean values of two or more groups are different does not relate to variation in each group (homogeneity).**
2. **Goodness of fit tests allow us to see criticize a model after creating inferences.**
3. **Probabilistic programming does not include statistical analysis.**

**Programming (the exercises are based on the model and the code presented in the lecture):**

1. **Run the programs presented in the lecture.**
2. **What is the mean of given that we know is less than 45. That is, suppose we have been given new information that the change in behaviour occurred prior to day 45. What is the expected value of $\lambda\_1$ now? (You do not need to redo the PyMC part. Just consider all instances where tau\_samples < 45.)**
3. **Extend the previous model to consider two switchpoints.**

# Cited works:

Davidson-Pilon. (2016). *Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference*. Crawfordsville, United States: Addison-Wesley.

Github. (2017). Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/Chapter1\_Introduction/Ch1\_Introduction\_PyMC2.ipynb. Retrieved from https://github.com/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/blob/master/Chapter1\_Introduction/Ch1\_Introduction\_PyMC2.ipynb