# Chapter 3

More about Inference

Jussi Ahola

Introduction

- In chapter 3 the Bayes' theorem is applied to inference problems.

- The feasibility of the Bayesian inference is demonstrated through simple examples.

- Presentation outline
    - Example 1: Unstable particles
    - Example 2: Bent coin
    - Example 3: Legal evidence
    - Lesson's learned
    - Home exercises

Unstable particles problem

- Unstable particles are emitted from a source and decay at a distance x, a real number that has an exponential probability distribution with characteristic length λ. Decay events can only be observed if they occur in a window extending from $x = 1$ cm to $x = 20$ cm. N decays are observed at locations $\{x_1, x_2, ... , x_N\}$. What is λ?

x cm

1 cm
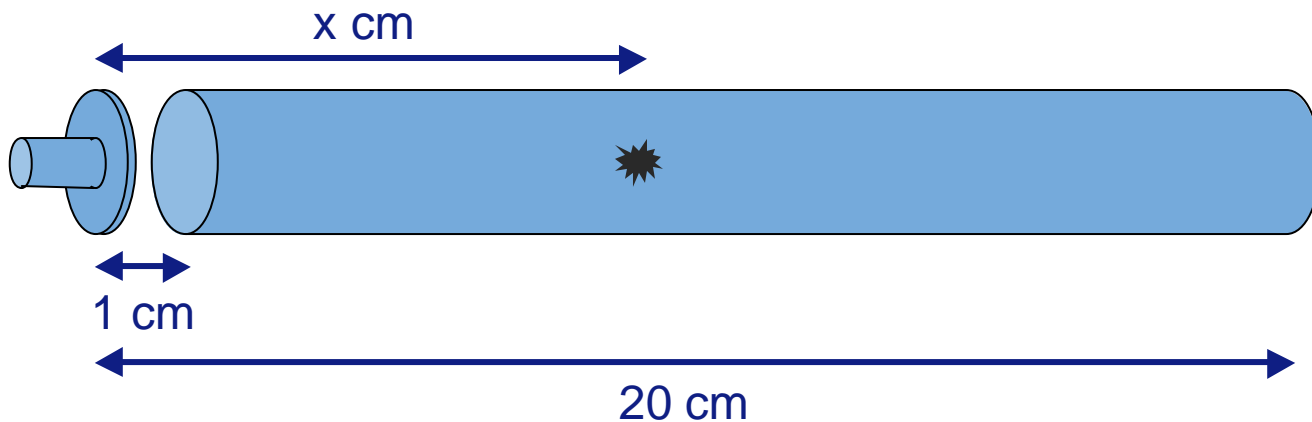
20 cm

Traditional solution

- Constructing estimators of λ.

  - λ is the mean of unconstrained exponential distribution. -> Sample mean $\overline{x}$ reasonable starting point for obtaining estimator $\hat{\lambda}$.

    - $\hat{\lambda} = \overline{x} - 1$ is an appropriate for $\lambda \ll 20$ cm.

    - Promising estimators for $\lambda \ll 20$ cm could be found.

  - No obvious estimator that would work under all conditions.

- Fitting the model to the data, or a processed version of the data.

  - No satisfactory approach based on fitting the density $P(x|\lambda)$ to a histogram derived from data could be found.

- *What is the general solution to this problem and others like it?*

Probabilistic solution

- Find the (posterior) probability of λ given the data.

  - The probability of one data point, given λ is

$$P(x\mid\lambda)=\begin{cases}\dfrac{1}{\lambda}\dfrac{e^{-\frac{x}{\lambda}}}{Z(\lambda)} & 1<x<20, \\ 0 & \text{otherwise}\end{cases} \quad\text{where } Z(\lambda)=\int_{1}^{20}\frac{1}{\lambda}e^{-\frac{x}{\lambda}}dx=\left(e^{-\frac{1}{\lambda}}-e^{-\frac{20}{\lambda}}\right)$$

  - Using the Bayes' theorem the posterior is:

$$P(\lambda\mid\{x_1,x_2,...,x_N\})=\frac{P(\{x\}\mid\lambda)P(\lambda)}{P(\{x\})}\propto\frac{1}{[\lambda Z(\lambda)]^N}e^{\frac{-\sum_1^N x_n}{\lambda}}P(\lambda)$$

- The posterior probability distribution represents the unique and complete solution to the problem.

  - There is no need to invent "estimators" nor do we need to invent criteria for comparing alternative estimators with each other.

Graphical interpretation
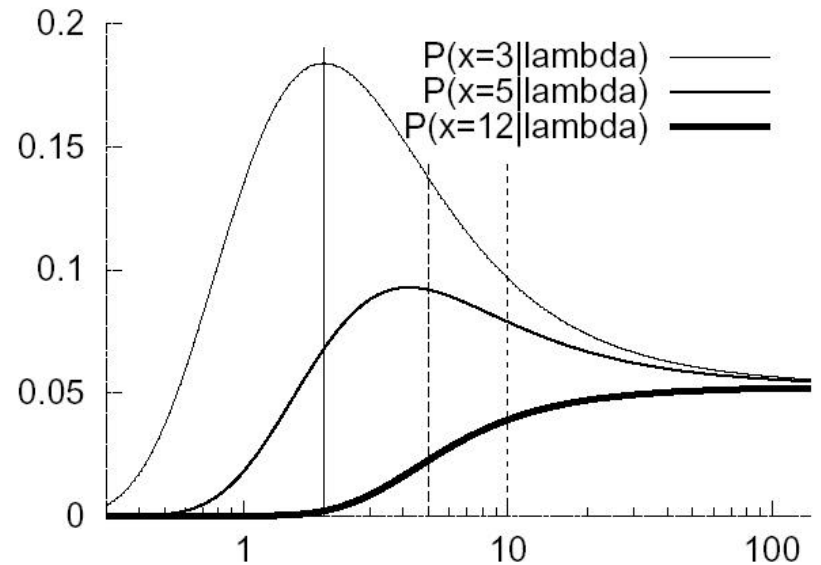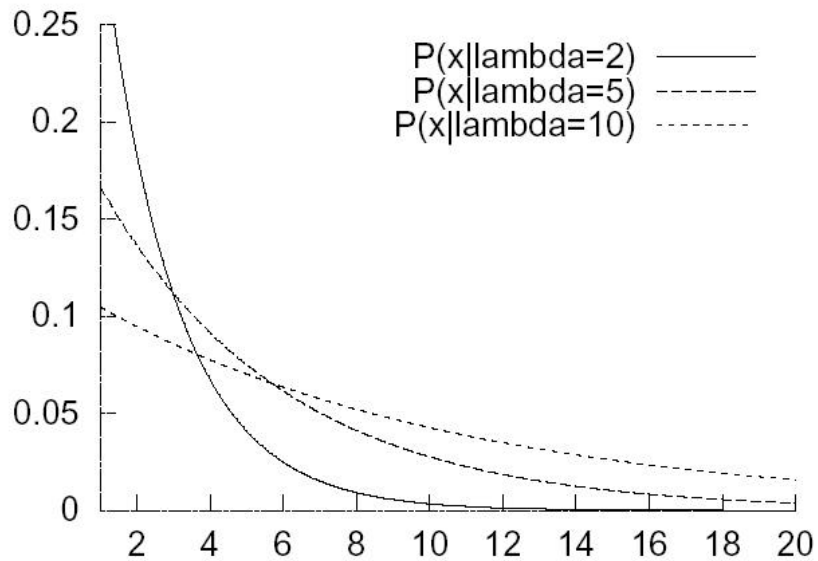
Example

- For a data set consisting of several points, e.g., the six points $\{x\}_{n=1}^{N} = \{1.5,2,3,4,5,12\}$, the likelihood function $P(\{x\}|\lambda)$ is the product of the N functions of λ, $P(x_n|\lambda)$.
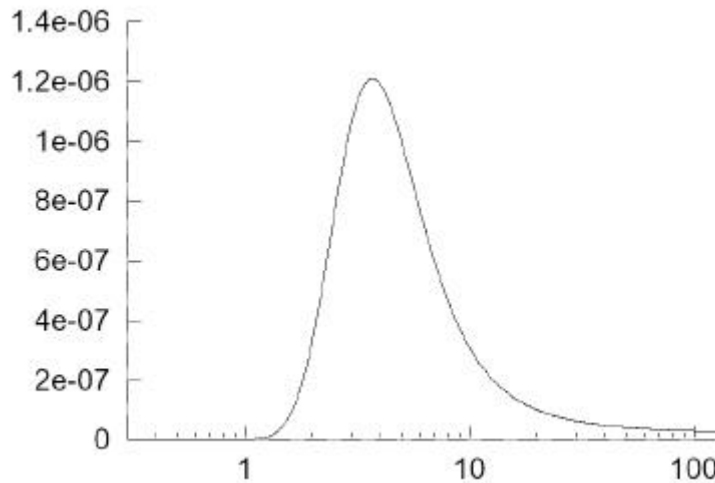
Assumptions on inference

- Inference is conditional on assumptions, that are explicit, which has several benefits:

  - Once assumptions are made, the inferences are objective and unique, reproducible with complete agreement by anyone who has the same information and makes the same assumptions.

  - When the assumptions are explicit, they are easier to criticise, and easier to modify.

  - When we are not sure which of various alternative assumptions is the most appropriate for a problem, we can treat this question as another inference task.

  - We can take into account our uncertainty regarding such assumptions when we make subsequent predictions.

Bent coin problem

- A bent coin is tossed F times; a sequence s of heads and tails is observed (denoted by the symbols a and b). What is the bias of the coin (p_a), and what is the probability that the next toss will result in a head?
- The solution:
  - The probability that F tosses result in a sequence s that contains $\{F_a, F_b\}$ counts of the two outcomes is (assumptions are called H_1):

$$P(s \mid p_a, F, \mathrm{H}_1) = p_a^{F_a}(1 - p_a)^{F_b}$$

  - A uniform prior distribution is assumed:

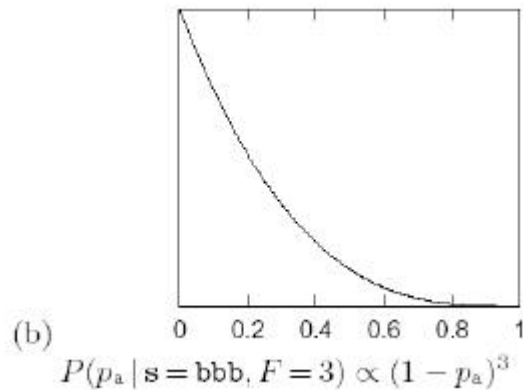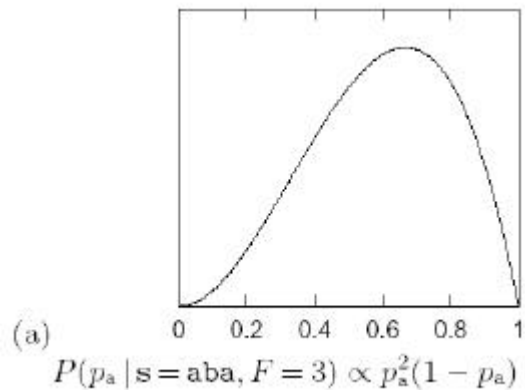$$P(p_a \mid \mathrm{H}_1) = 1, \quad \mathrm{p}_a \in [0,1]$$

  - The posterior distribution is obtained by multiplying the prior by the likelihood (and divided by the evidence).

Inferring the bias

- Assuming H_1 to be true, the posterior probability of p_a, given a string s of length F that has counts $\{F_a, F_b\}$, is:

$$P(p_a \mid s, F, H_1) = \frac{P(s \mid p_a, F, H_1)P(p_a \mid H_1)}{P(s \mid F, H_1)} = \frac{p_a^{F_a}(1 - p_a)^{F_b}}{P(s \mid F, H_1)}$$

$$P(s \mid F, H_1) = \int_0^1 p_a^{F_a}(1 - p_a)^{F_b}\, dp_a = \frac{F_a! F_b!}{(F_a + F_b + 1)!}$$

(a)   0   0.2   0.4   0.6   0.8   1

$P(p_\mathrm{a} \mid s = \mathrm{aba}, F = 3) \propto p_\mathrm{a}^2(1 - p_\mathrm{a})$

(b)   0   0.2   0.4   0.6   0.8   1

$P(p_\mathrm{a} \mid s = \mathrm{bbb}, F = 3) \propto (1 - p_\mathrm{a})^3$

Predicting the next toss

- The prediction about the next toss, i.e. the probability that next toss is a head, is obtained by integrating over p_a. By the sum rule,

$$P(a \mid s, F) = \int P(a \mid p_a) P(p_a \mid s, F) dp_a = \int p_a \frac{p_a^{F_a}(1-p_a)^{F_b}}{P(s \mid F, \mathrm{H}_1)} dp_a =$$

$$\frac{(F_a + F_b + 1)!}{F_a! F_b!} \int p_a^{F_a+1}(1-p_a)^{F_b} dp_a = \frac{F_a + 1}{F_a + F_b + 2}$$

Model comparison

- Introducing a new hypothesis H_0:

  - The source is not really a bent coin but is really a perfectly formed die with one face painted heads and the other five painted tails ->
    $P_a = P_0 = 1/6$

- How probable H_1 is relative to H_0:

$$P(H_n \mid s,F) = \frac{P(s \mid F, H_n)P(H_n)}{P(s \mid F)}, \quad n = 0,1$$

$$P(s \mid F) = \sum_0^1 P(s \mid F, H_n)P(H_n)$$

$$P(H_0) = P(H_1) = 0.5$$

$$P(s \mid F, H_0) = p_0^{F_a}(1 - p_0)^{F_b}$$

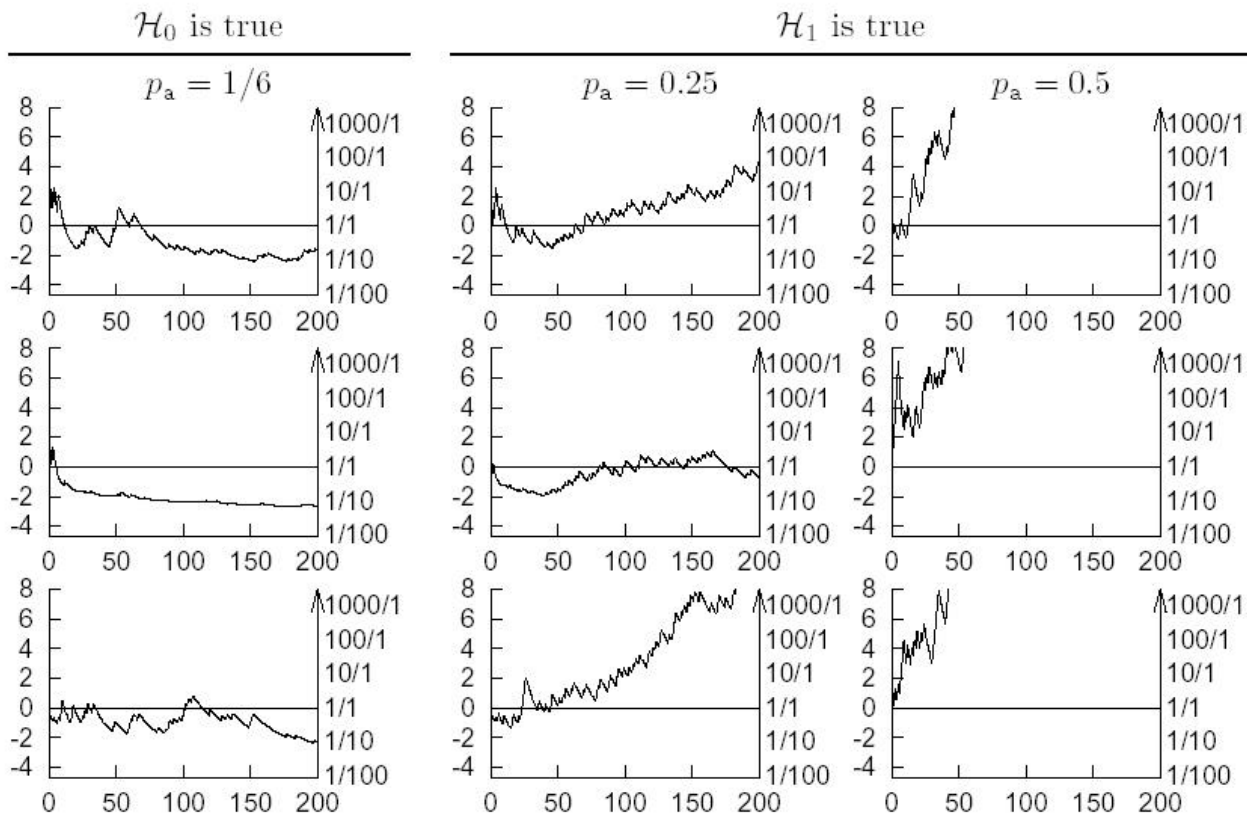$$P(s \mid F, H_1) = \frac{F_a! \, F_b!}{(F_a + F_b + 1)!}$$

Posterior probability ratio

- The posterior probability ratio of model H_1 to H_0 is:

$$\frac{P(H_1 \mid s,F)}{P(H_0 \mid s,F)} = \frac{P(s \mid F,H_1)P(H_1)}{P(s \mid F,H_0)P(H_0)} = \frac{\dfrac{F_a! F_b!}{(F_a + F_b + 1)!}}{p_0^{F_a}(1-p_0)^{F_b}}$$

F	Data (F_a, F_b)	$\dfrac{P(\mathcal{H}_1 \mid \mathrm{s}, F)}{P(\mathcal{H}_0 \mid \mathrm{s}, F)}$	
6	$(5, 1)$	222.2	
6	$(3, 3)$	2.67	
6	$(2, 4)$	0.71	$= 1/1.4$
6	$(1, 5)$	0.356	$= 1/2.8$
6	$(0, 6)$	0.427	$= 1/2.3$
20	$(10, 10)$	96.5	
20	$(3, 17)$	0.2	$= 1/5$
20	$(0, 20)$	1.83	

Typical behaviour of the evidence

Legal evidence problem

- Two people have left traces of their own blood at the scene of a crime. A suspect, Oliver, is tested and found to have type O blood. The blood groups of the two traces are found to be of type O (a common type in the local population, having frequency 60%) and of type AB (a rare type, with frequency 1%). Do these data give evidence in favour of the proposition that Oliver was one of the two people present at the crime?

Solution

- Denote with

  - S the proposition "Oliver and one unknown person were present".

  - $\overline{S}$ the proposition "two unknown people from the population were present".

- The prior in this problem is the prior probability ratio between the propositions S and $\overline{S}$.

- The task is to evaluate the contribution made by the data D, that is, the likelihood ratio, $P(D \mid S, \mathrm{H}) / P(D \mid \overline{S}, \mathrm{H})$.

$$P(D|S,\mathrm{H}) = p_{AB}$$

$$P(D|\overline{S},\mathrm{H}) = 2p_O p_{AB}$$

$$P(D \mid S, \mathrm{H}) / P(D \mid \overline{S}, \mathrm{H}) = \frac{1}{2p_O} = 0.83$$

Case: Alberto

- Consider the case of another suspect, Alberto, who has type AB.
- Denote *S'* the proposition "Alberto and one unknown person were present".
- The likelihood ratio in this case is:

$$P(D|S',\mathrm{H})/P(D|\overline{S},\mathrm{H}) = \frac{1}{2p_{AB}} = 50$$

Another consideration

- Let's imagine that 99% of people are of blood type O, and the rest are of type AB. Only these two blood types exist in the population.

    - Intuitively, we still believe that the presence of the rare AB blood provides positive evidence that Alberto was there.

    - Does the fact that type O blood was detected at the scene favour the hypothesis that Oliver was present? ->
      Everyone in the population would be under greater suspicion.

- The data may be compatible with any suspect of either blood type being present, but if they provide evidence for some theories, they must also provide evidence against other theories.

And yet another

- Let's imagine that instead of two people's blood stains there are ten, and that in the entire local population of one hundred, there are ninety type O suspects and ten type AB suspects.

  - Without any other information, and before the blood test results come in, there is a one in 10 chance that Oliver was at the scene, since we know that 10 out of the 100 suspects were present.

  - The results of blood tests tell that nine of the ten stains are of type AB, and one of the stains is of type O. ->
    There is now only a one in ninety chance that Oliver was there, since we know that only one person present was of type O.

The general case

- n_O blood stains of individuals of type O are found, and n_{AB} of type AB, a total of N individuals in all, and unknown people come from a large population with fractions p_O and p_{AB} (there may be other blood types too).
- The task is to evaluate the likelihood ratio for the two hypotheses:
  - S, "the type O suspect and N-1 unknown others left N stains".
  $$P(n_O, n_{AB} \mid S) = \frac{(N-1)!}{(n_O - 1)! \, n_{AB}!} \, p_O^{n_O - 1} p_{AB}^{n_{AB}}$$
  - $\overline{S}$, "N unknowns left N stains".
  $$P(n_O, n_{AB} \mid \overline{S}) = \frac{N!}{n_O! \, n_{AB}} \, p_O^{n_O} p_{AB}^{n_{AB}}$$
- The likelihood ratio is:
  $$\frac{P(n_O, n_{AB} \mid S)}{P(n_O, n_{AB} \mid \overline{S})} = \frac{n_O / N}{p_O}$$

Lessons learned

- The essence of the Bayes' theorem is:

  What you know about the world after the data arrive is what you knew before (prior distribution), and what the data told you (posterior distribution).

- Probability theory reaches parts that ad hoc (orthodox statistics') methods cannot reach.

- Inference cannot be done without making assumptions.

Home exercises

- Exercise 3.10. Another example in which the emphasis is not on priors. You visit a family whose three children are all at the local school...

- Exercise 3.12. A bag contains one counter, known to be either white or black...