Statistical methods I Simo Tuomisto
013456353

Home Exam

Contents

1 Problem 1

1.1 Exercise (ii)

The joint probability that northern lights occur and that the sky is clear is $P(\text{lights})P(\text{clear}) = \frac{4}{5} \times \frac{1}{2} = \frac{4}{10}$. This means that probability of not seeing lights (no lights and sky clear; lights, but no vision; no lights and no vision) is $1 - P(\text{lights})P(\text{clear}) = \frac{6}{10}$. As every night is an independent occurence, the probability of not seeing lights is $P(\text{bad holiday in Lapland}) = \left(\frac{6}{10}\right)^5 \approx 7.77\%$

2 Problem 2

I created a simulator for the plane boarding. Results obtained from it are presented below.

2.1 Exercise (i)

Obtained times were:

$$\mu_1 = 635.74\,\mathrm{s}$$
$$\sigma_1 = 31.82\,\mathrm{s}$$

2.2 Exercise (ii)

Obtained times were:

$$\mu_1 = 718.52\,\mathrm{s}$$
$$\sigma_1 = 42.38\,\mathrm{s}$$

The Outside-In boarding style was significantly faster. The whole story can be seen from the boarding time distributions plotted in a figure below. Random loading creates much more waiting for the passengers and increases the overall boarding time.

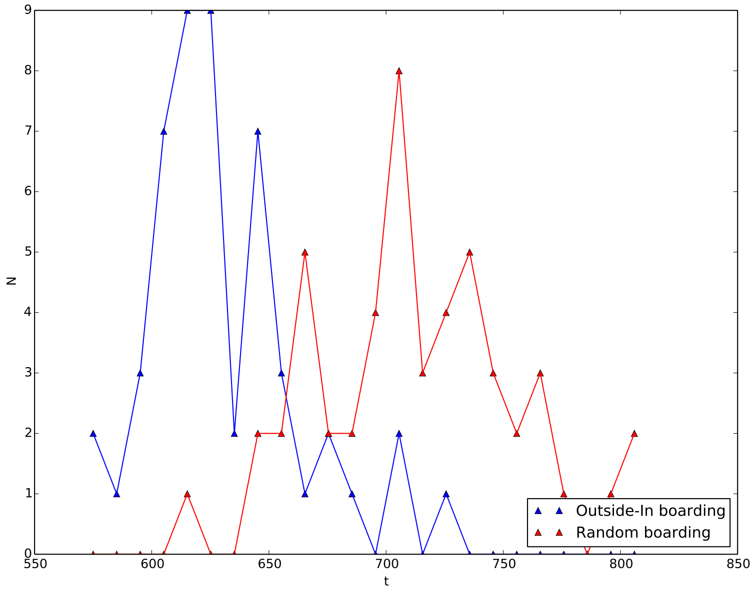

Figure 1: Distributions of boarding times

3 Problem 3

3.1 Exercise (i)

Fisher's discriminant method was used in order to separate the populations. Weights c that were found were:

$$c_1 = 0.7073757$$
$$c_2 = 0.7068377$$

Plot of different x_3 values look like this:

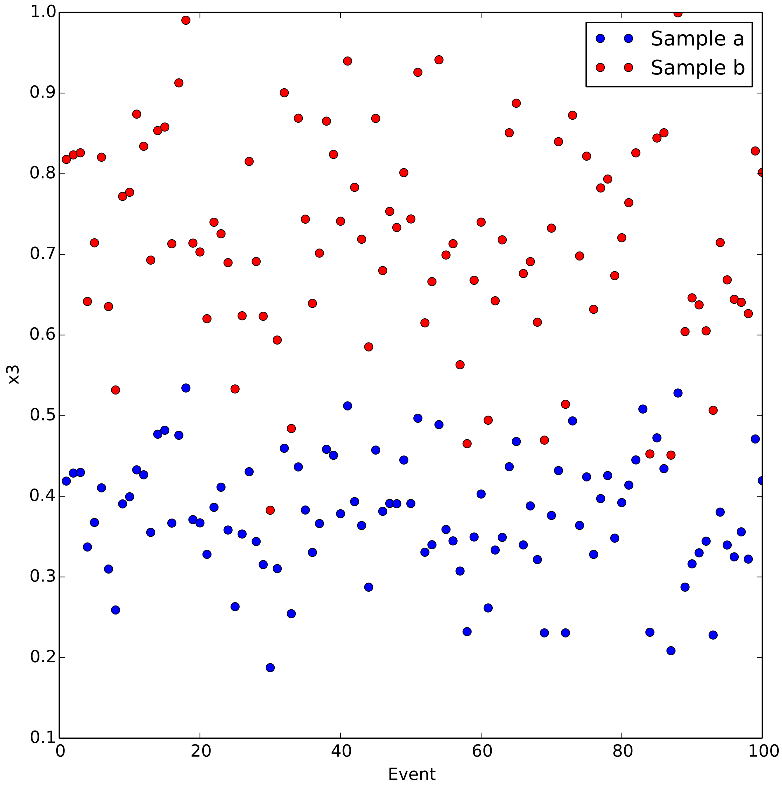

Figure 2: x_3-values

3.2 Exercise (ii)

The 95% rejection limit gives a rejection condition $x_3 < 0.582$. Acceptance efficiency for this limit is 92%. Below is a plot of the samples with the rejection line plotted.

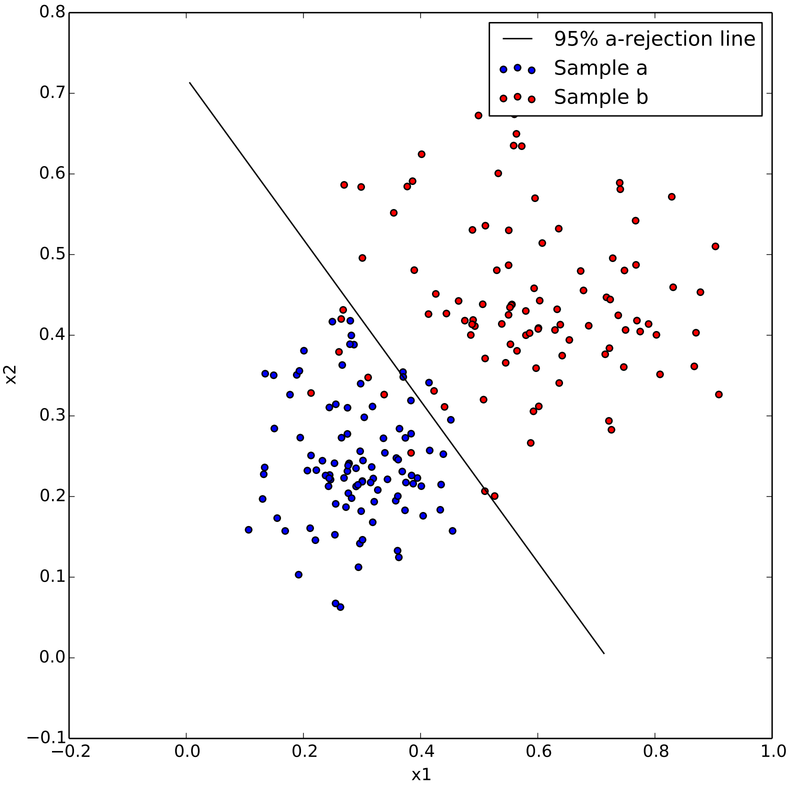

Figure 3: 95%-rejection of Sample A

3.3 Exercise (iii)

With gaussian assumption the probabilty of obtaining the given values of x_1 and x_2 from sample A is:

$$P(x|A) = P_1(x_1|A)P_2(x_2|A)$$

where $P_i(x_i|A)$ are gaussian distributions of each direction.

Same goes for assumption that it was obtained from sample B. Now the probability of sample A or B being correct when $\mathbf{x}$ has been obtained is:

$$P(A|x) = \frac{P(x|A)}{P(x|A) + P(x|B)} \approx 36.6\%$$
$$P(B|x) = \frac{P(x|B)}{P(x|A) + P(x|B)} \approx 36.6\%$$

4 Problem 4

4.1 Exercise (i)

First step is to find t_{expected} as a function of $\Delta T_{\text{observed}}$ with parameter C.

$$\begin{aligned} C &= \frac{E}{m\Delta T} \\ &= \frac{UIt}{m\Delta T} \Leftrightarrow \\ t &= \frac{mC\Delta T}{UI} \end{aligned}$$

The least squares fit for t_{observed} is

$$\begin{aligned} \chi^2 &= \sum_i \frac{(t_i - f(\Delta T_i, C))^2}{\sigma_{t,i}^2} \\ &= \sum_i \frac{\left(t_i - \frac{mC\Delta T_i}{UI}\right)^2}{\sigma_{t,i}^2} \end{aligned}$$

Using this a value for parameter C was obtained with Scipy-library's ODR-libary:

$$\begin{aligned} C &= 4.18797 \pm 0.01086 \frac{\text{J}}{\text{g K}} \\ \chi^2 &= 7.8932 \\ p &= 0.99503 \end{aligned}$$

4.2 Exercise (ii)

RFC-bound was written from the analytic form of χ^2

$$
\begin{aligned}
\ln L = -\frac{\chi^2}{2} &= -\frac{1}{2}\sum_i \frac{\left(t_i - \dfrac{mC\Delta T_i}{UI}\right)^2}{\sigma_{t,i}^2} \\
\frac{\partial \ln L}{\partial C} &= \frac{1}{2}\sum_i \frac{m\Delta T_i}{UI}\frac{\left(t_i - \dfrac{mC\Delta T_i}{UI}\right)}{\sigma_{t,i}^2} \\
-\frac{\partial^2 \ln L}{\partial C^2} &= \frac{1}{2}\sum_i \left(\frac{m\Delta T_i}{\sigma_{t,i}UI}\right)^2 \\
E\left[-\frac{\partial^2 \ln L}{\partial C^2}\right] &= \frac{1}{2}\sum_i \left(\frac{m}{\sigma_{t,i}UI}\right)^2 E[\Delta T_i^2] \\
&= \frac{1}{2}\sum_i \left(\frac{m}{\sigma_{t,i}UI}\right)^2 \left(E[\Delta T_i]^2 + V[\Delta T_i]\right) \\
&= \frac{1}{2}\sum_i \left(\frac{m}{\sigma_{t,i}UI}\right)^2 \left(\mu_{\Delta T,i}^2 + \sigma_{\Delta T,i}^2\right) \\
V[C] = \left(-\frac{\partial^2 \ln L}{\partial C^2}\right)^{-1} &= 2\left(\sum_i \left(\frac{m}{\sigma_{t,i}UI}\right)^2 \left(\mu_{\Delta T,i}^2 + \sigma_{\Delta T,i}^2\right)\right)^{-1}
\end{aligned}
$$

Obtained value for this was

$$\sigma_{\text{rfc}} = 0.01446$$

This is *below* the obtained variance, so I think I have made a mistake somewhere. Unfortunately, I don't have time to fix it.

4.3 Exercise (iii)

It took me some time to understand how this problem was to be solved. Finally I found a page that explained that the situation was a *total least squares problem*, that wasn't teached on the course. I found a solver for this from Scipy's ODR-package.

Obtained values were:

$$C = 4.18589 \pm 0.01115 \frac{\text{J}}{\text{g K}}$$
$$\chi^2 = 4.6199$$
$$p = 0.9684$$

4.4 Exercise (iv)

I'd say that the errors are so small, that the chemist must have an exceptionally good measuring apparatus. P-values are enormous. The result is almost identical to water's heat capacity of 4.181 J/gK.

5 Problem 5

5.1 Exercise (i)

I tested the measured data with respect to the Monte Carlo data with Pearson's χ^2 test. The test works best when the number of samples per bin is more than 5, but I couldn't come up with any better one.

Obtained χ^2- and p-values were (index describes the background):

$$\chi_1^2 = 1673.26$$
$$p_1 = 0$$
$$\chi_2^2 = 29.97$$
$$p_2 = 0.05217$$

Plot of the backgrounds is below.

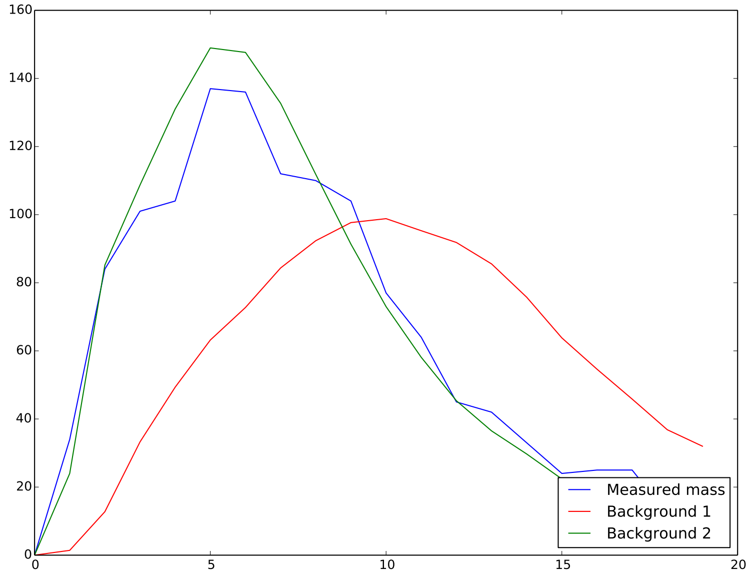

Figure 4: Backgrounds

5.2 Exercise (ii)

I created pseudoexperiments and measured χ^2- and p-value from them:

$$\chi^2_{\text{pseudo},1} = 2101.39$$
$$p_{\text{pseudo},1} = 0$$
$$\chi^2_{\text{pseudo},2} = 83.28$$
$$p_{\text{pseudo},2} = 2.27 \times 10^{-10}$$

As can be seen, the p-value of background 2 gets very low.

Plot of the pseudoexperiments is below.

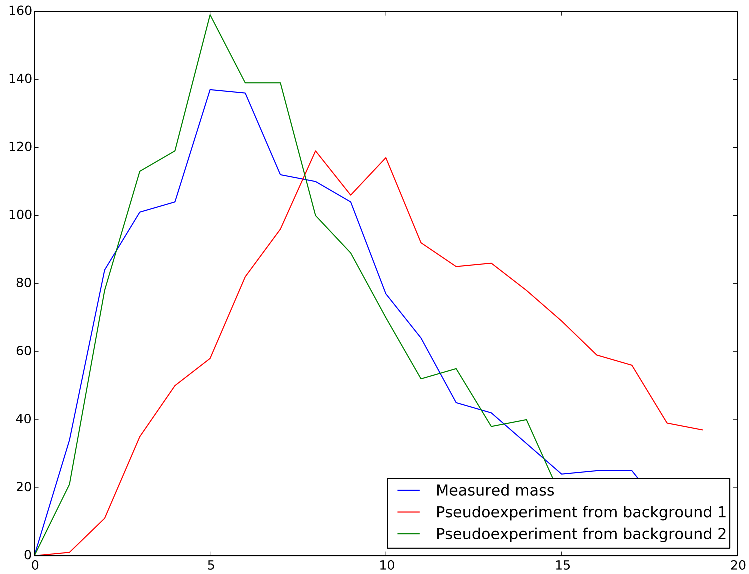

Figure 5: Pseudoexperiment

5.3 Exercise (iii)

I ran a minimization algorithm that found the optimal value for $a = 0.1044$. The χ^2- and p-values for it and it's pseudoexperiments were

$$\begin{aligned}\chi^2_{\text{optimal}} &= 20.496\\ p_{\text{optimal}} &= 0.365\\ \chi^2_{\text{pseudo,optimal}} &= 23.852\\ p_{\text{pseudo,optimal}} &= 0.1599\end{aligned}$$

Below is a plot of the optimal background

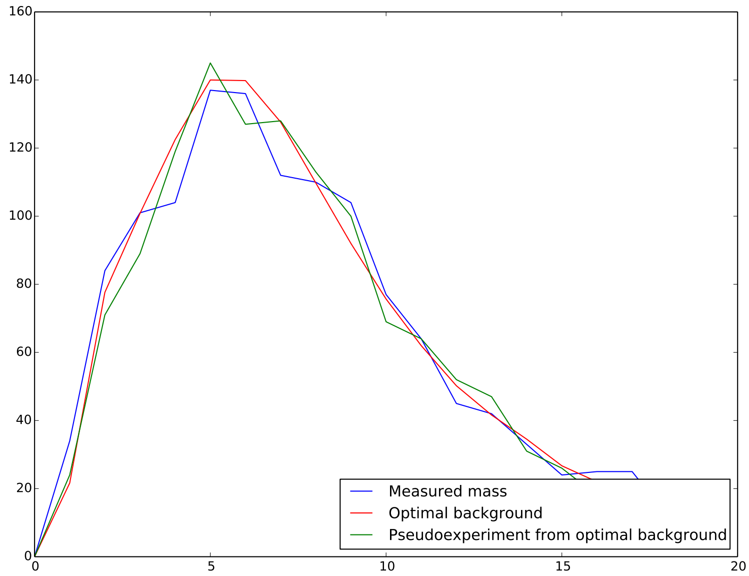

Figure 6: Optimal background

6 Problem 6

6.1 Exercise (i)

The excepted number of decays can be found by integrating R over a window of width Δt:

$$
\begin{aligned}
R(t) &= \frac{dN}{dt} = Ae^{-\lambda t} + B \\
N\big|_{t-\Delta t/2}^{t+\Delta t/2} &= \int_{t-\Delta t/2}^{t+\Delta t/2} dt'\, R(t') \\
&= \int_{t-\Delta t/2}^{t+\Delta t/2} dt'\, Ae^{-\lambda t'} + B \\
&= \left(-\frac{A}{\lambda}e^{-\lambda t'} + Bt'\right)_{t-\Delta t/2}^{t+\Delta t/2} \\
&= \left(-\frac{A}{\lambda}e^{-\lambda(t+\Delta t/2)} + B(t+\Delta t/2)\right) - \left(-\frac{A}{\lambda}e^{-\lambda(t-\Delta t/2)} + B(t-\Delta t/2)\right) \\
&= \frac{A}{\lambda}e^{-\lambda t}\left(e^{\lambda\Delta t/2} - e^{-\lambda\Delta t/2}\right) + B\Delta t \\
&= \frac{2A}{\lambda}e^{-\lambda t}\sinh\frac{\lambda\Delta t}{2} + B\Delta t
\end{aligned}
$$

Fitting to this gives parameters:

$$
\begin{aligned}
A &= 609661 \pm 5.9 \\
\lambda &= 7.0226 \pm 0.000746 \times 10^{-4}\,\mathrm{s}^{-1} \\
B &= 6080 \pm 0.80976 \times 10^{-4}\,\mathrm{s}^{-1}
\end{aligned}
$$

Uncertainities were found by finding $\chi^2(\theta+\sigma) = \chi^2(\theta) + 1$.

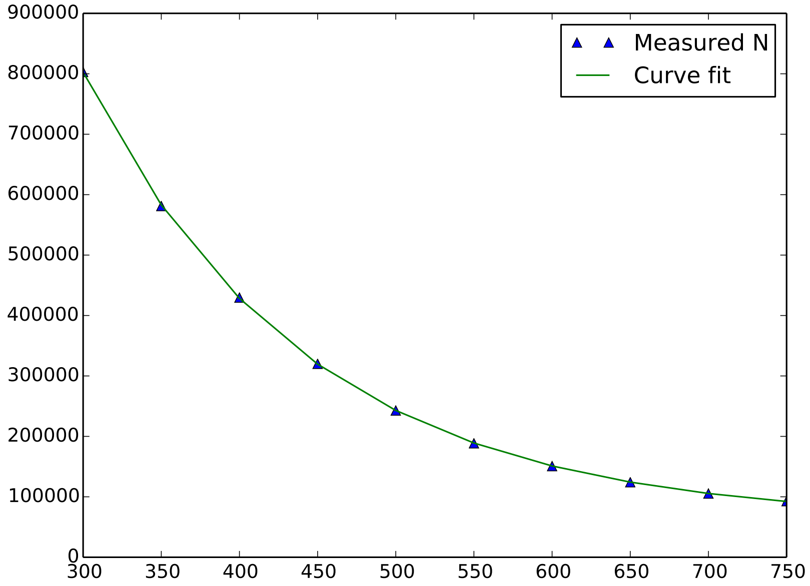

Figure 7: Fit plotted

6.2 Exercise (ii)

Lifetime can be calculated from λ:

$$\lambda = \frac{\ln 2}{\tau} \Leftrightarrow$$
$$\tau = \frac{\ln 2}{\lambda} \approx 98.702\,\mathrm{s}$$
$$\sigma_\tau = \sqrt{\left(\frac{\partial \tau}{\partial \lambda}\right)^2 \sigma_\lambda^2}$$
$$\sigma_\tau = \left|\frac{\partial \tau}{\partial \lambda}\right| \sigma_\lambda$$
$$= \left|-\frac{\ln 2}{\lambda^2}\right| \sigma_\lambda \approx 1.048 \times 10^{-3}\,\mathrm{s}$$