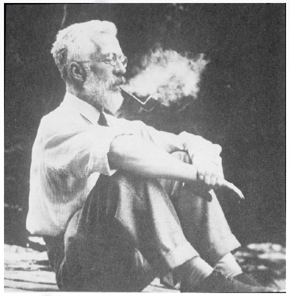
Applied Statistics

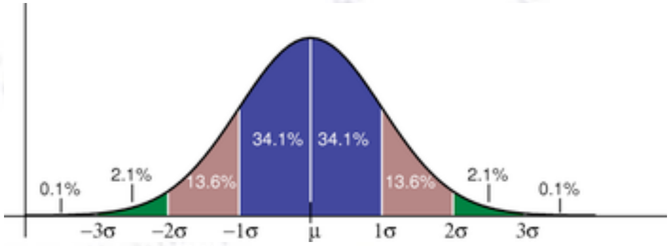
Problem Set Solution and Discussion








Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense"

Overall comments

The problem set is hard!

The problem set is hard, and this one was no exception. If anything, on the contrary.

So if you had a hard time, then there should be no surprise. But the point of the problem set is of course also to give problems, so that every student will be challenged. This problem set (also) managed that…

It closely resembles what to expect for the exam, so you should be well prepared by now.

The solutions

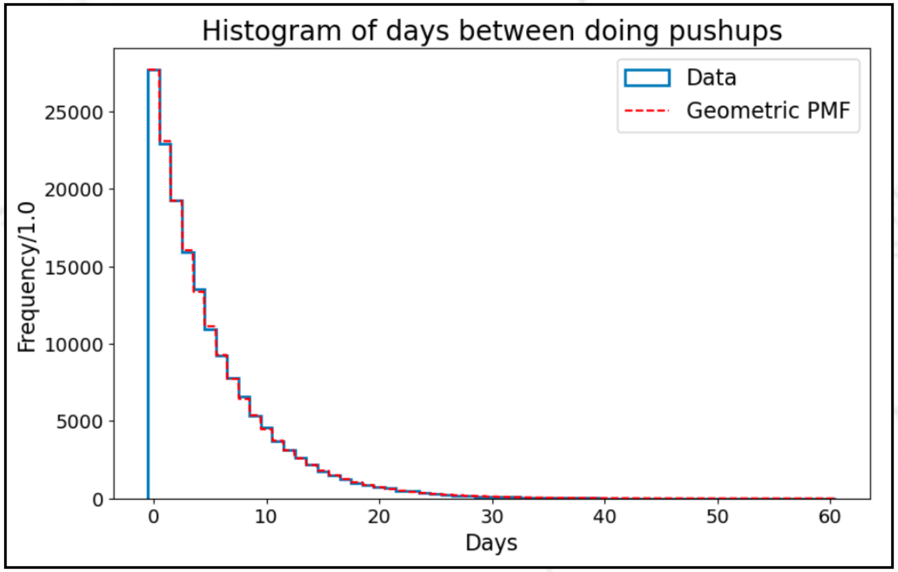
Problem 1.1

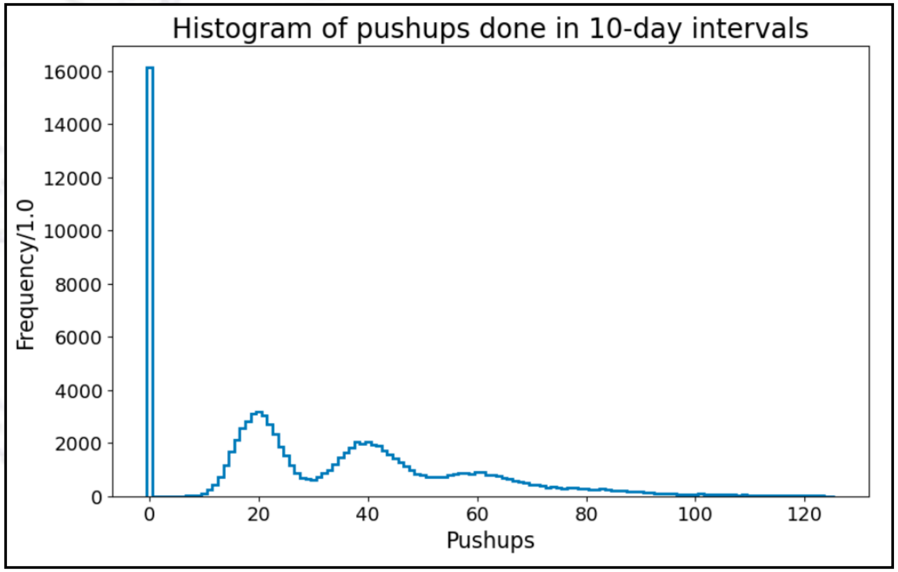
1.1 (6 points) Every day, you roll a normal die, and if you get a six, you do roll the die 120 times and do as many push-ups as you get sixes. Otherwise, you don't do any push-ups.

- What is the distribution of days between doing push-ups?
- What is the mean, median, and standard deviation of number of push-ups in 10 days?

1.1.1: The distribution is **Geometric**, which is an "integer exponential distribution", in this case with an exponent of 5/6 and a normalisation of P(N days = 0) = 1/6.

1.1.2: To compute the distribution, simulation is the easiest to use.



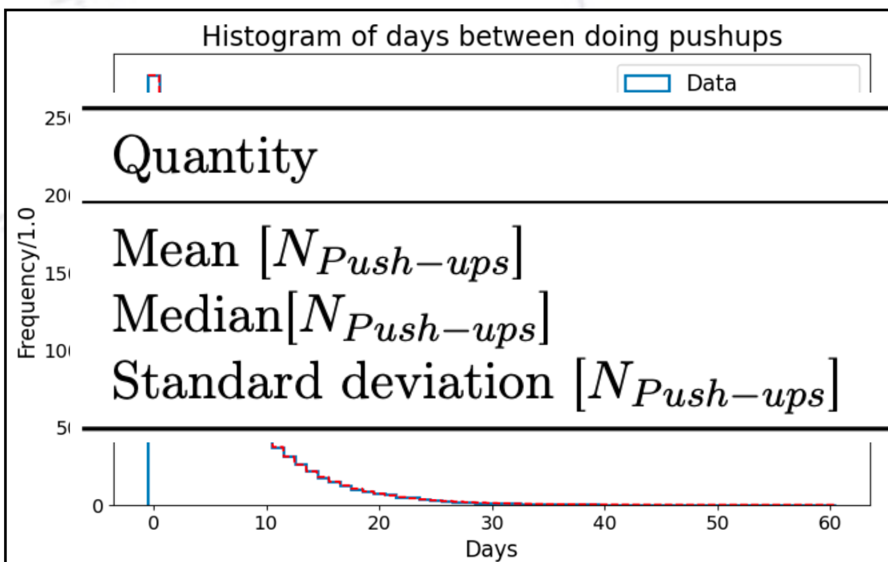


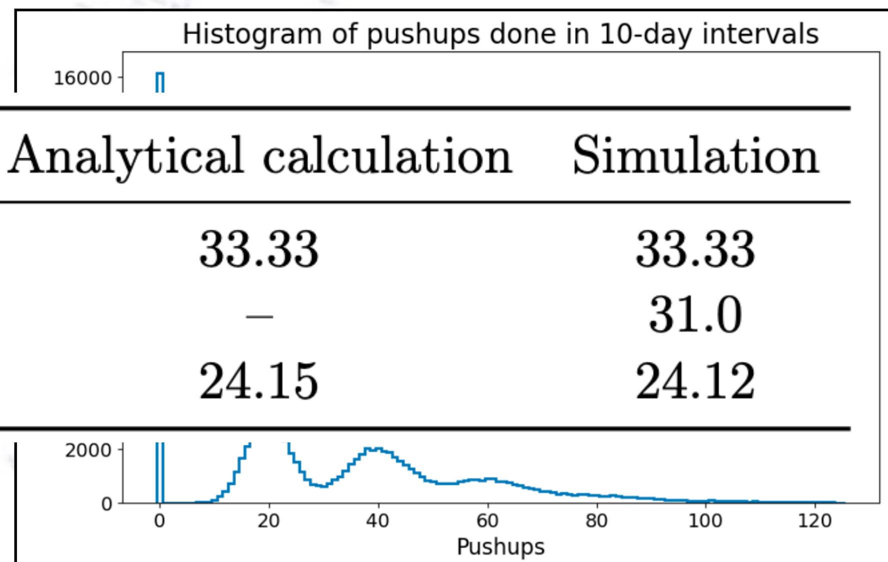
Problem 1.1

> **1.1** (6 points) Every day, you roll a normal die, and if you get a six, you do roll the die 120 times and do as many push-ups as you get sixes. Otherwise, you don't do any push-ups.
>
> - What is the distribution of days between doing push-ups?
> - What is the mean, median, and standard deviation of number of push-ups in 10 days?

1.1.1: The distribution is **Geometric**, which is an "integer exponential distribution", in this case with an exponent of 5/6 and a normalisation of P(N days = 0) = 1/6.

1.1.2: To compute the distribution, simulation is the easiest to use.





Quantity	Analytical calculation	Simulation
Mean $[N_{Push-ups}]$	33.33	33.33
Median$[N_{Push-ups}]$	–	31.0
Standard deviation $[N_{Push-ups}]$	24.15	24.12

Problem 1.2

> **1.2** (4 points) The Djoser pyramide in Egypt is North-South aligned to 3 degrees.
>
> - Estimate the probability that the pyramid is North-South aligned by coincidence.

1.2.1:
The naive probability of **this or something more extreme** is "within 3 degrees to either side".

However, a pyramide has a 4-fold rotational symmetry, which means that out of the full 360 degrees, there are four places where one observes is "this or more extreme".

Thus the result is 6 / 90 = 1/15th. As it happens, the Djoser Pyramide is slightly rectangular, thus in this (rare) case, it only has a 2-fold symmetry. Either answer are accepted.

Note: The problem was designed to catch ChatGPT and other LLMs off guard!

Problem 2.1

2.1 (5 points) Water on Earth ($\oplus$) has a Deuterium to Hydrogen ratio of $r_\oplus = (149 \pm 3) \times 10^{-6}$. The hydrogen of the proto-solar system ($\odot$) has a ratio of $r_\odot = (25 \pm 5) \times 10^{-6}$, while that of comets ($C$) have been measured to be $r_C = (309 \pm 20) \times 10^{-6}$.

- From these numbers, what fraction of water on Earth do you estimate come from the original proto-solar system, and what fraction do you attribute to comets?

2.1.1: The calculation is straight forward error propagation, once finding a formula for the fraction.

$$B = \frac{r_\oplus - r_\odot}{r_C - r_\odot}$$

$$\sigma_B = \sqrt{\frac{\sigma_{r_C}^2 (r_\oplus - r_\odot)^2}{(r_C - r_\odot)^4} + \frac{\sigma_{r_\oplus}^2}{(r_C - r_\odot)^2} + \sigma_{r_\odot}^2 \left(-\frac{1}{r_C - r_\odot} + \frac{r_\oplus - r_\odot}{(r_C - r_\odot)^2}\right)^2}$$

The fraction of water originating from comets is f(A) = 0.437 ± 0.034

The fraction of water originating from the proto-solar system is f(B) = 0.563 ± 0.034

Problem 2.2

2.2 (8 points) You run a detector for a time interval of $\Delta t = 98.4$s, during which the detector yields $N = 1971$ counts. The time interval uncertainty is $\sigma_{\Delta t} = 3.7$s, indepedent of Δt.

- What is the rate $r = N/t$ and its uncertainty?
- How long should you measure to get a relative uncertainty on the rate r below 2.5%?

2.2.1: The uncertainty on N is from its Poissonian nature, and from there error propagation gives the answer.

2.2.2: The uncertainty as a function of time has two components:

N: Scales as 1/sqrt(t) ~ 1/sqrt(N)

t: Scales as 1/t, as the uncertainty is constant while t grows linearly.

The time to get 2.5% relative uncertainty can then be determined to be 193 s.

$$\sigma_r = \sqrt{\left(\frac{\sqrt{N}}{t}\right)^2 + \left(\frac{N\sigma_{\Delta t}}{t^2}\right)^2}.$$

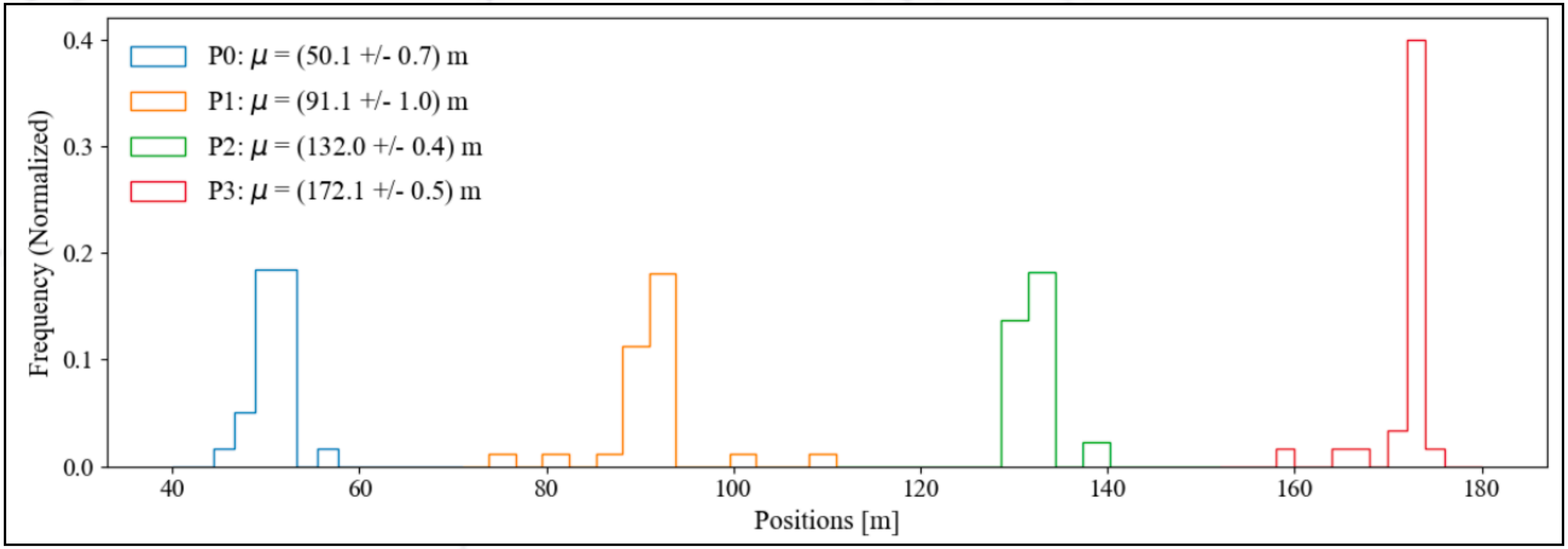
$$t = \frac{1 + \sqrt{1 + 4r^2\sigma_{\Delta t}^2 R^2}}{2R^2 r}$$

Problem 2.3

2.3 (14 points) The file **www.nbi.dk/~petersen/data_PylonPositions.csv** contains position measurements (both with and without uncertainties) of four pylons for a bridge.

- Using measurements without uncertainty, determine the four pylon positions.
- Using measurements with uncertainty, determine the four pylon positions.
- Combine the two measurement groups. Do they match each other?
- Test if the four measured pylon positions are equidistant.

2.3.1: The unweighted pylon positions can be found through "normal" mean, though Chauvenet's Criterion has to be considered.

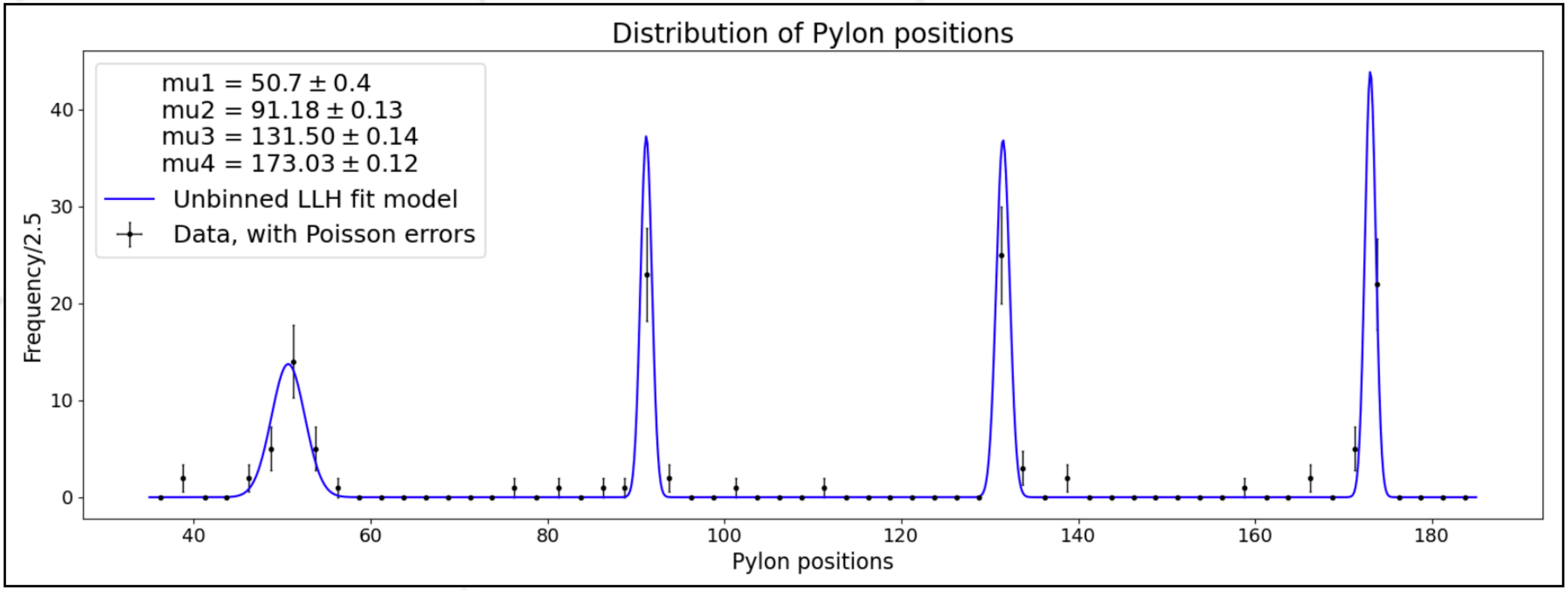


Problem 2.3

2.3 (14 points) The file **www.nbi.dk/~petersen/data_PylonPositions.csv** contains position measurements (both with and without uncertainties) of four pylons for a bridge.

- Using measurements without uncertainty, determine the four pylon positions.
- Using measurements with uncertainty, determine the four pylon positions.
- Combine the two measurement groups. Do they match each other?
- Test if the four measured pylon positions are equidistant.

2.3.2: The weighted means needs cross checks by z-score.



Problem 2.3

2.3 (14 points) The file **www.nbi.dk/~petersen/data_PylonPositions.csv** contains position measurements (both with and without uncertainties) of four pylons for a bridge.

- Using measurements without uncertainty, determine the four pylon positions.
- Using measurements with uncertainty, determine the four pylon positions.
- Combine the two measurement groups. Do they match each other?
- Test if the four measured pylon positions are equidistant.

2.3.3: The combination should be straight forward, as can be seen below.

For the second pylon (here P1) there is a 2 sigma discrepancy. I would note it, but also accept it (remember that there is a trial factor of 4, anyway).

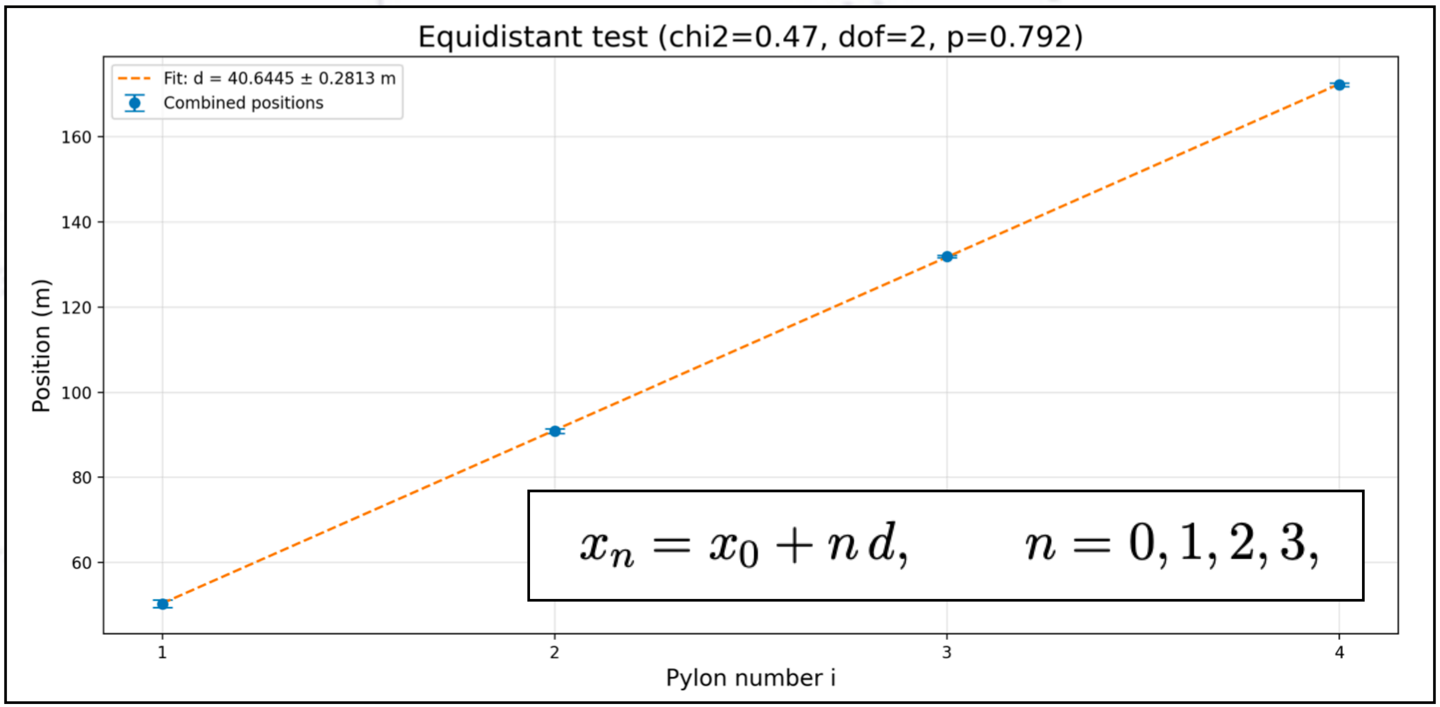
Pylon No.	Unweighted [m]	Weighted [m]	z-value	p-value	Combined [m]
P0	50.1 ± 0.7	50.6 ± 0.2	0.64	0.52	50.54 ± 0.17
P1	91.2 ± 1.0	93.2 ± 0.3	2.01	0.045	93.0 ± 0.3
P2	132.0 ± 0.4	132.4 ± 0.4	0.74	0.46	132.2 ± 0.3
P3	172.1 ± 0.5	171.4 ± 0.4	1.06	0.289	171.7 ± 0.3

Problem 2.3

2.3 (14 points) The file **www.nbi.dk/~petersen/data_PylonPositions.csv** contains position measurements (both with and without uncertainties) of four pylons for a bridge.

- Using measurements without uncertainty, determine the four pylon positions.
- Using measurements with uncertainty, determine the four pylon positions.
- Combine the two measurement groups. Do they match each other?
- Test if the four measured pylon positions are equidistant.

2.3.4: The test is best done with a linear ChiSquare fit (interdistances correlate!)



$$x_n = x_0 + n\,d, \qquad n = 0, 1, 2, 3,$$

Problem 2.2 + 2.3

2.2 (8 points) You run a detector for a time interval of $\Delta t = 98.4$s, during which the detector yields $N = 1971$ counts. The time interval uncertainty is $\sigma_{\Delta t} = 3.7$s, indepedent of Δt.

- What is the rate $r = N/t$ and its uncertainty?
- How long should you measure to get a relative uncertainty on the rate r below 2.5%?

Common mistakes/slips:

2.2.1 - Not using poisson errors for N
2.2.2 - Forgetting to substitute N=rt (some forgot just sigmaN)

2.3.1 - No cleaning of data, giving no error on reported mean
2.3.2 - Similar to 2.3.1. No cleaning of data / discussing outliers
(students that missed this aspect also tended not to plot the data)
2.3.3 - Not combining measurements in weighted mean, or no z-test/t-test
2.3.4 - No straight line fit, or no statistical discussion with chi squared, p-value etc

Problem 3.1

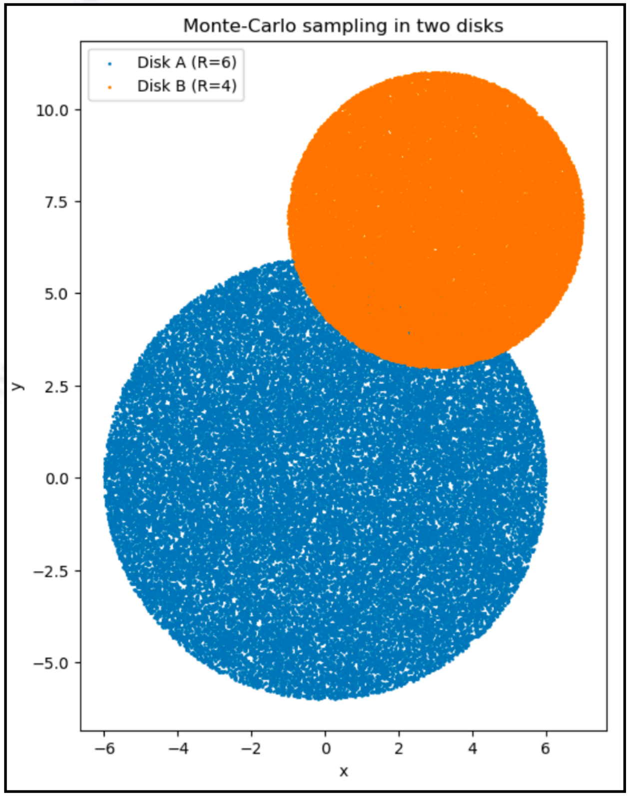
3.1 (8 points) Circles A and B are centered at (0,0) and (3,7) and have radii of 6 and 4, respectively.

- What fraction of A overlaps with B? And conversely, what fraction of B overlaps with A?
- If the circles were 4D "hyperballs" centered at (0,0,0,0) and (3,7,-1,2), respectively, and with the same radii, what would the answers then be?

3.1.1: There is a formula for the overlap area between circles, as follows:

$$A = r_1^2 \arccos\left(\frac{r_1^2 + d^2 - r_2^2}{2r_1 d}\right) + r_2^2 \arccos\left(\frac{r_2^2 + d^2 - r_1^2}{2r_2 d}\right) - \frac{1}{2}\sqrt{(-d + r_1 + r_2)(d + r_1 - r_2)(d - r_1 + r_2)(d + r_1 + r_2)}$$

The best solution is to compare the analytic to a simulated solution, and once these are assured in agreement, proceed the 4D case.



The overlaps:

$$f_{AB} = 0.09088 \pm 0.00001$$

$$f_{BA} = 0.20450 \pm 0.00001$$

$$f_{AB} = 0.01465 \pm 0.00001$$

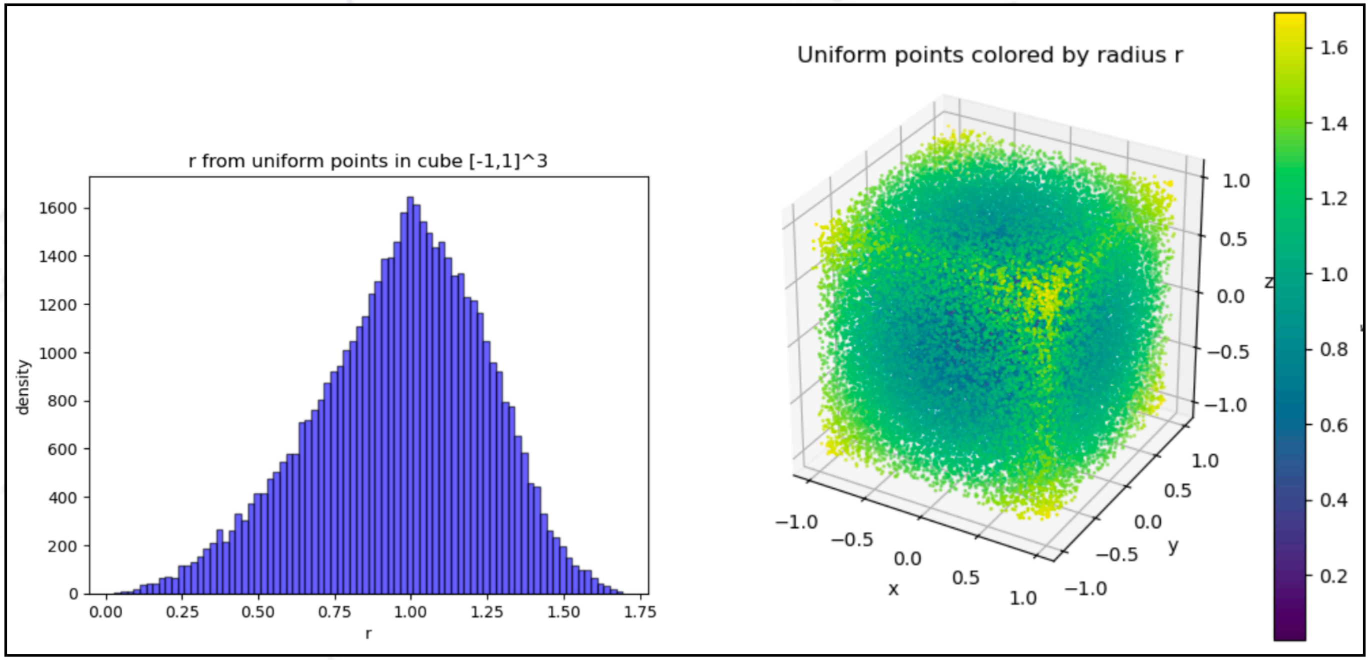
$$f_{BA} = 0.07422 \pm 0.00001$$

Problem 3.2

3.2 (15 points) You want to simulate the radial material distribution $m(r)$ from a uniform explosion.

- Generate 50000 x, y, and z values in the range $[-1, 1]$ and plot the spherical coordinate r.
- Selecting only points with $z > 0$ and $r < 1$, what distributions in θ and ϕ do you obtain?
- How would you produce random velocities v according to $f(v) = (v/v_0)^2 \exp(-v/v_0)$?
- Given $v_0 = 100$ m/s and that the radial distance of material r as a function of velocity is $r(v) = \sin(\theta)v^2/g$ ($g = 9.82\text{m/s}^2$), simulate 10000 values of θ and v. Combine these to obtain values of r, and plot the resulting distribution $m(r)$.

3.2.1: The generation should be straight forward, but is nicely illustrated below.

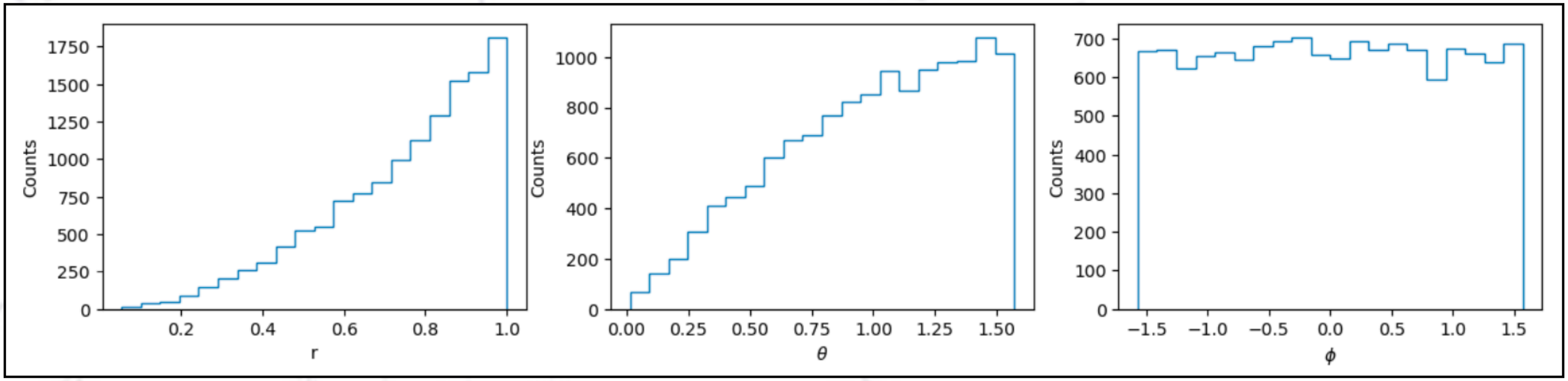


Problem 3.2

3.2 (15 points) You want to simulate the radial material distribution $m(r)$ from a uniform explosion.

- Generate 50000 x, y, and z values in the range $[-1, 1]$ and plot the spherical coordinate r.
- Selecting only points with $z > 0$ and $r < 1$, what distributions in θ and ϕ do you obtain?
- How would you produce random velocities v according to $f(v) = (v/v_0)^2 \exp(-v/v_0)$?
- Given $v_0 = 100$ m/s and that the radial distance of material r as a function of velocity is $r(v) = \sin(\theta)v^2/g$ ($g = 9.82\text{m/s}^2$), simulate 10000 values of θ and v. Combine these to obtain values of r, and plot the resulting distribution $m(r)$.

3.2.2: Simulation yields the below distributions:



Recognising the angular distributions, it is "good sports" to plot the distribution on top.

Problem 3.2

3.2 (15 points) You want to simulate the radial material distribution $m(r)$ from a uniform explosion.

- Generate 50000 x, y, and z values in the range $[-1, 1]$ and plot the spherical coordinate r.
- Selecting only points with $z > 0$ and $r < 1$, what distributions in θ and ϕ do you obtain?
- How would you produce random velocities v according to $f(v) = (v/v_0)^2 \exp(-v/v_0)$?
- Given $v_0 = 100$ m/s and that the radial distance of material r as a function of velocity is $r(v) = \sin(\theta)v^2/g$ ($g = 9.82\text{m/s}^2$), simulate 10000 values of θ and v. Combine these to obtain values of r, and plot the resulting distribution $m(r)$.

<u>Common mistakes/slips:</u>

3.1. No uncertainties when printing out MC simulation results

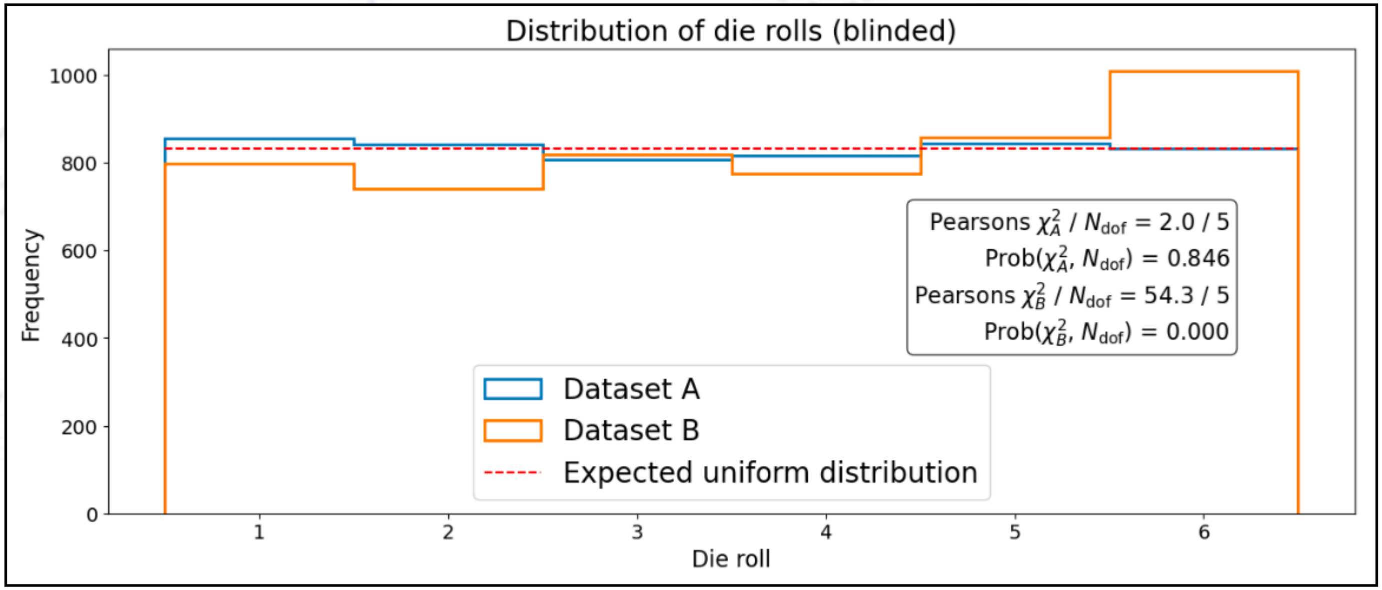
3.2.1 Not plotting the full range
3.2.2 Not plotting distributions (scatter plots are not valid).
3.2.3 Not discussing why it is ok to cut the tail for the accept/reject method.

Problem 4.1

4.1 (12 points) You get a permanently closed box with 25 normal (i.e. six-sided) dices in. One of the dice is potentially fake, with all the sides having the same (unknown) value. You can shake the box and see the resulting 25 dice roll as many times as you like.

- Simulate 200 box rolls and plot the die frequencies, both with and without a fake die in.
- For both of your simulated datasets, test if there is a fake die or not.
- How many rolls would you require before you can tell if there is a fake die or not?

4.1.1 + 4.1.2: The simulation shows that 200 rolls yields a clear difference from the expected uniform distribution. Testing it quantifies this statement. Careful writing "p = 0.0", but rather put "$p < 10^{-20}$" or the likes.



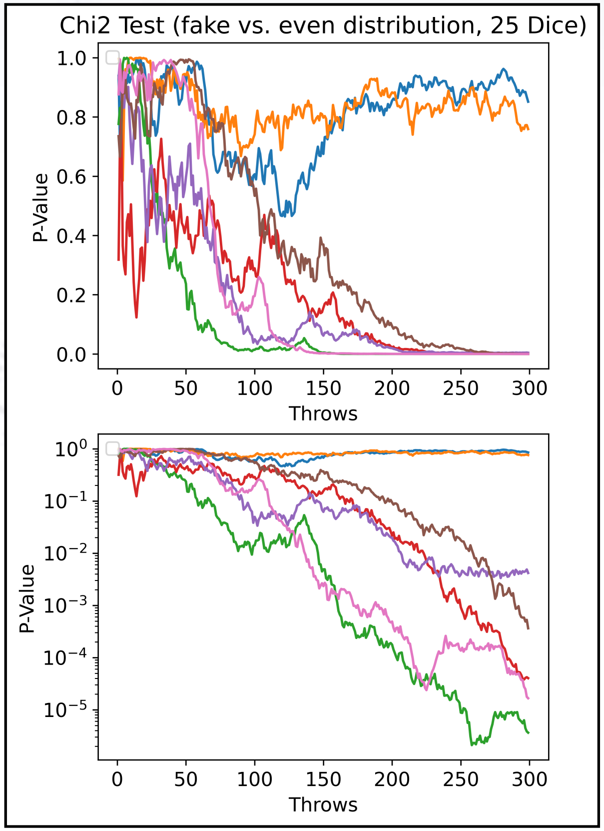
Problem 4.1

4.1 (12 points) You get a permanently closed box with 25 normal (i.e. six-sided) dices in. One of the dice is potentially fake, with all the sides having the same (unknown) value. You can shake the box and see the resulting 25 dice roll as many times as you like.

- Simulate 200 box rolls and plot the die frequencies, both with and without a fake die in.
- For both of your simulated datasets, test if there is a fake die or not.
- How many rolls would you require before you can tell if there is a fake die or not?

4.1.3: The way to investigate if there is a fake die or not is through simulation. One does not obtain the same result each time, and so an average must be made.
Plot the p-value as a function of the number of throws, possibly on a log(p-value) on the y-axis.

Depending on the required p-value the number of throws is around 100-125.

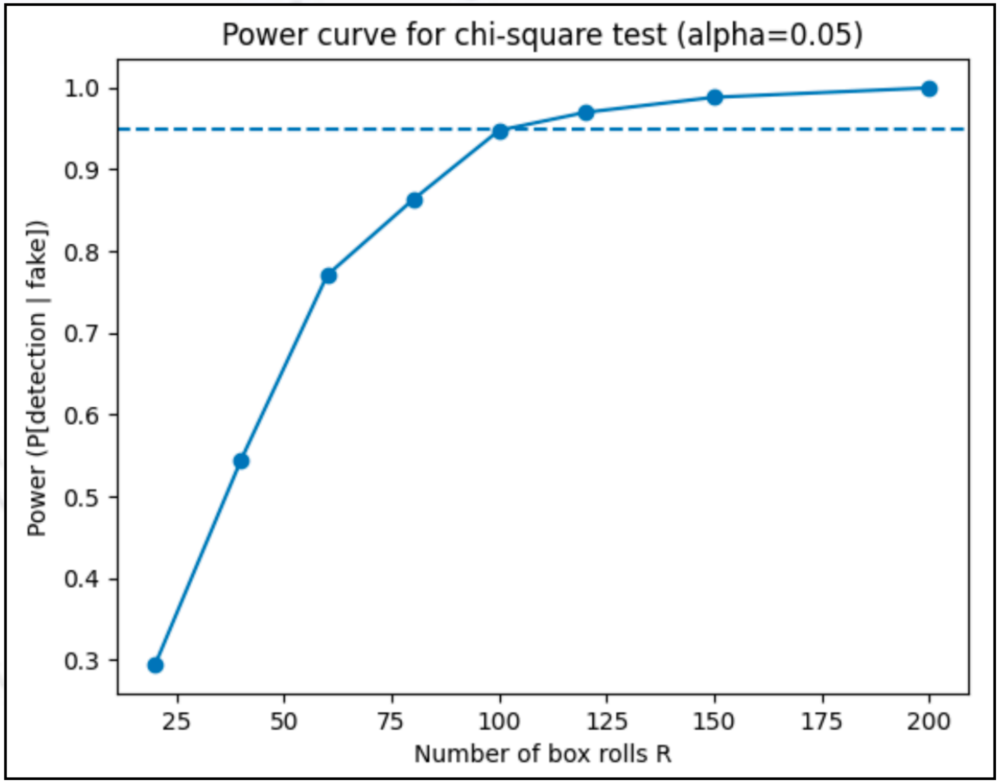


Problem 4.1

4.1 (12 points) You get a permanently closed box with 25 normal (i.e. six-sided) dices in. One of the dice is potentially fake, with all the sides having the same (unknown) value. You can shake the box and see the resulting 25 dice roll as many times as you like.

- Simulate 200 box rolls and plot the die frequencies, both with and without a fake die in.
- For both of your simulated datasets, test if there is a fake die or not.
- How many rolls would you require before you can tell if there is a fake die or not?

4.1.3: The way to investigate if there is a fake die or not is through simulation. One does not obtain the same result each time, and so an average must be made. Plot the p-value as a function of the number of throws, possibly on a log(p-value) on the y-axis. Depending on the required p-value the number of throws is around 100-125.



Problem 4.1

4.1 (12 points) You get a permanently closed box with 25 normal (i.e. six-sided) dices in. One of the dice is potentially fake, with all the sides having the same (unknown) value. You can shake the box and see the resulting 25 dice roll as many times as you like.

- Simulate 200 box rolls and plot the die frequencies, both with and without a fake die in.
- For both of your simulated datasets, test if there is a fake die or not.
- How many rolls would you require before you can tell if there is a fake die or not?

<u>Common mistakes/slips:</u>

4.1.2: Some people just write p=0, or don't mention chi2 (or other test) but then mention the p value.

4.1.3: Just used the mean of the p values

Just do 1 run but say they don’t have statistics

Do several runs but then do a distribution of p values for increasing #runs and take the mean.

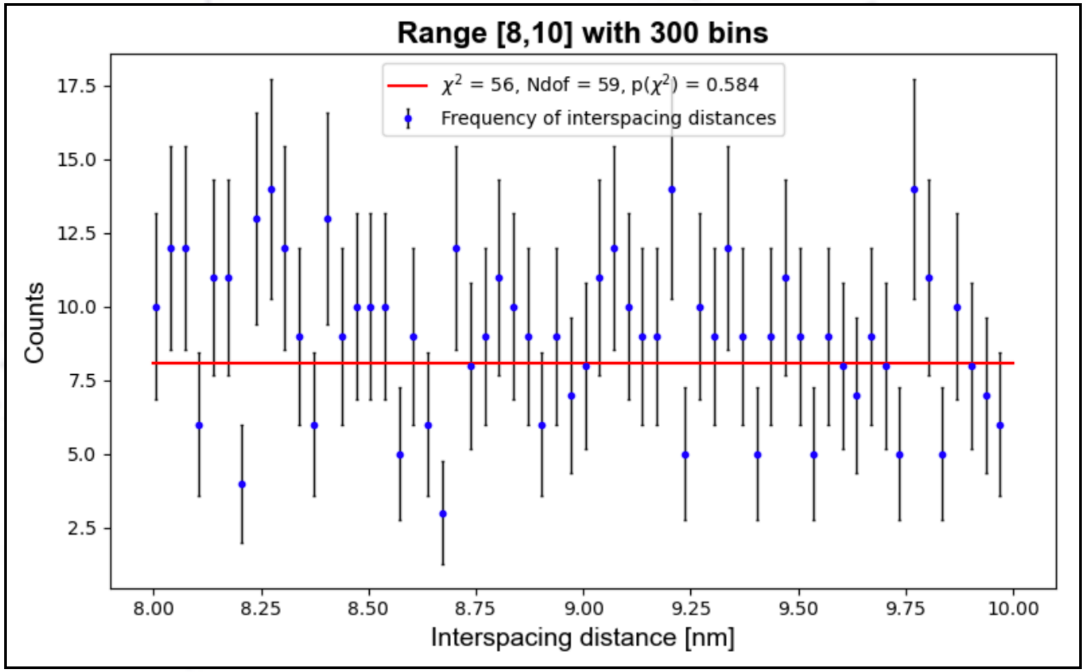
Try to do it theoretically but rely on mean cases or approximations leading to slightly wrong results.

Problem 5.1

5.1 (14 points) The file **www.nbi.dk/∼petersen/data_InconstantBackground.csv** contains molecular interspacing measurements d (in nm) from a scattering experiment.

- Plot the data and test to what extend the background in the range [8,10] is uniform.
- Fit the three Gaussian peaks at around $d = 0.9$, 3.4, and 5.9 nm, including local background.
- Test if the peaks have the same intensity, i.e. number of measurements in them.
- Try to fit the entire spectrum or parts of it best possible and comment on your results.

5.1.1: This is tested by fitting with a uniform distribution. Binning is important! In the below case the binning (and description) is slightly off.

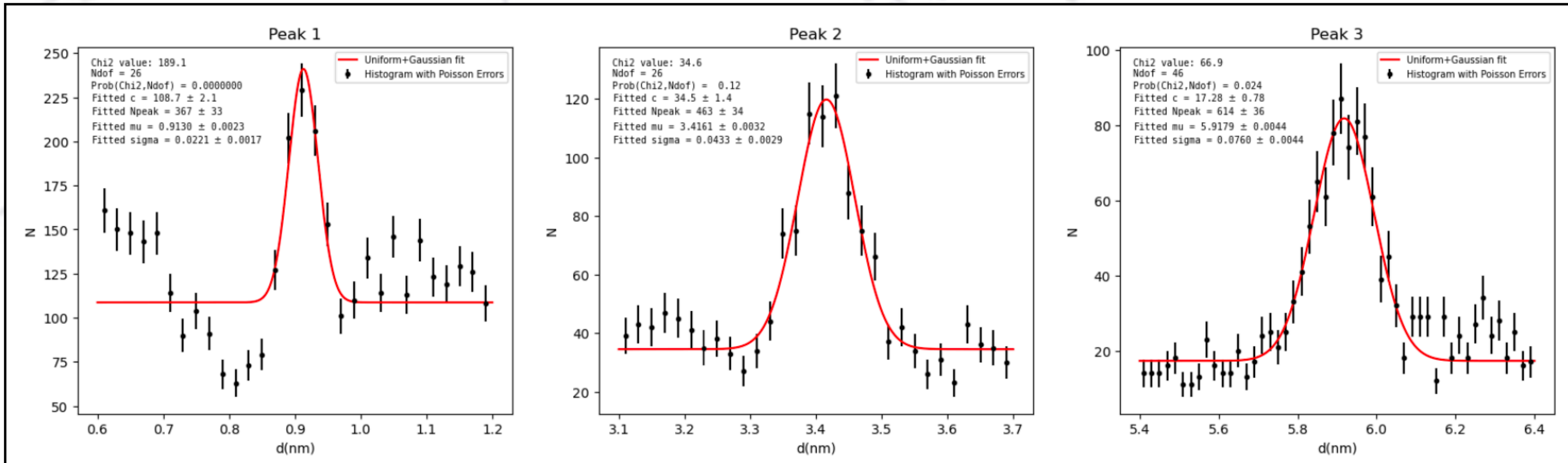


Problem 5.1

5.1 (14 points) The file **www.nbi.dk/~petersen/data_InconstantBackground.csv** contains molecular interspacing measurements d (in nm) from a scattering experiment.

- Plot the data and test to what extend the background in the range [8,10] is uniform.
- Fit the three Gaussian peaks at around $d = 0.9$, 3.4, and 5.9 nm, including local background.
- Test if the peaks have the same intensity, i.e. number of measurements in them.
- Try to fit the entire spectrum or parts of it best possible and comment on your results.

5.1.2+5.1.3: The challenge in these fits is the background, which is different in each case. This also affects the question about intensity, because using a flat background PDF leads to different intensities (they are generated the same).



Problem 5.1

5.1 (14 points) The file **www.nbi.dk/~petersen/data_InconstantBackground.csv** contains molecular interspacing measurements d (in nm) from a scattering experiment.

- Plot the data and test to what extend the background in the range [8,10] is uniform.
- Fit the three Gaussian peaks at around $d = 0.9$, 3.4, and 5.9 nm, including local background.
- Test if the peaks have the same intensity, i.e. number of measurements in them.
- Try to fit the entire spectrum or parts of it best possible and comment on your results.

5.1.4: Once the three peaks are fitted, one can in principle "subtract" these from the data to obtain the background… leading to a simpler fit.

The background is an exponential of a damped harmonic oscillator:

$$f(x) = e^{-x/\tau} \left(A \sin\left(\frac{\omega}{x+\delta} + \phi \right) + c \right)$$

However, people were creative, also including e.g. a Fourrier series:

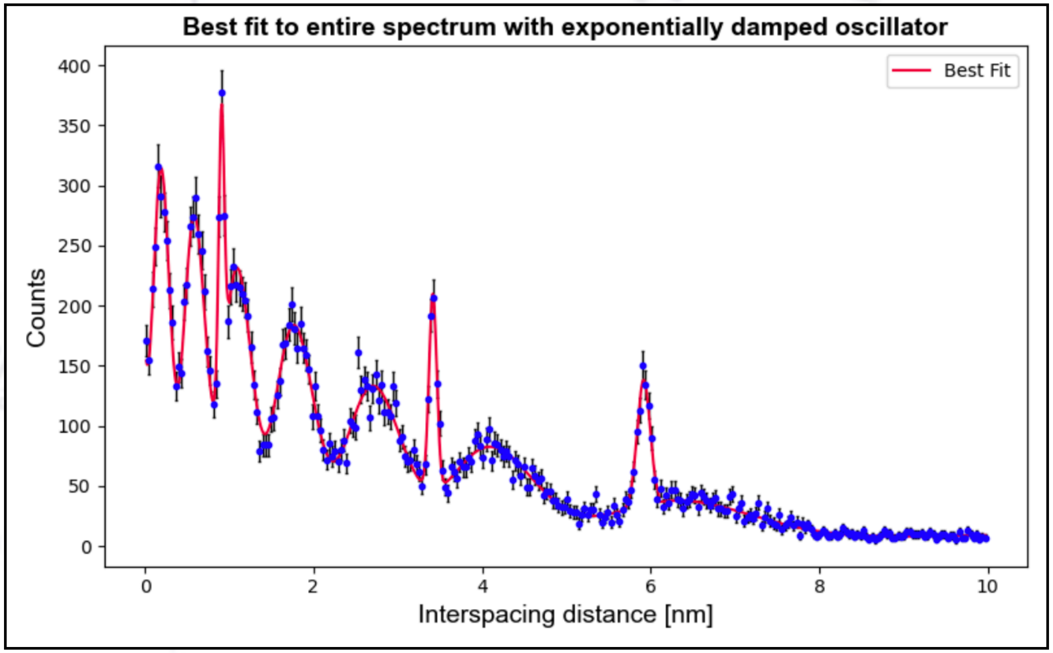
$$f_{\mathrm{bg}}(d) = a\,\mathrm{e}^{-kd} + c + \mathrm{e}^{-\lambda d} \sum_{n=1}^{N} \left[C_n \cos(n\omega d) + S_n \sin(n\omega d) \right],$$

Problem 5.1

5.1 (14 points) The file **www.nbi.dk/~petersen/data_InconstantBackground.csv** contains molecular interspacing measurements d (in nm) from a scattering experiment.

- Plot the data and test to what extend the background in the range [8,10] is uniform.
- Fit the three Gaussian peaks at around $d = 0.9$, 3.4, and 5.9 nm, including local background.
- Test if the peaks have the same intensity, i.e. number of measurements in them.
- Try to fit the entire spectrum or parts of it best possible and comment on your results.

5.1.4: Even when one knows the PDF, it is still not a simple fit to get running. However, at least the parameters of the three Gaussians are known well.



Problem 5.1

5.1 (14 points) The file **www.nbi.dk/~petersen/data_InconstantBackground.csv** contains molecular interspacing measurements d (in nm) from a scattering experiment.

- Plot the data and test to what extend the background in the range [8,10] is uniform.
- Fit the three Gaussian peaks at around $d = 0.9$, 3.4, and 5.9 nm, including local background.
- Test if the peaks have the same intensity, i.e. number of measurements in them.
- Try to fit the entire spectrum or parts of it best possible and comment on your results.

<u>Common mistakes/slips:</u>

5.1.2: Some people do not say what they are using to fit and just show a line or very vague description.

5.1.3: A lot of good reasonings but very bad results from fit.

5.1.4: Some people just assumed as many gaussians as possible, other used just the simple exponential decay.

Problem 5.2

5.2 (14 points) The table below lists the North-South alignment of Egyptian pyramids (in arc minutes).

Pyramid	1.Meidum	2.Bent	3.Red	4.Khufu	5.Khafre	6.Menk.	7.Sahure	8.Nefer.
Align. year	2600 BC	2583 BC	2572 BC	2554 BC	2522 BC	2489 BC	2446 BC	2433 BC
East Align.	-20.6±1.0	-17.3±0.2	-8.7±0.2	-3.4±0.2	6.0±0.2	12.4±1.0	23±10	30±10
West Align.	-18.1±1.0	-11.8±0.2	–	-2.8±0.2	6.0±0.2	14.1±1.8	–	–

- Test to what extend the East (E) and West (W) alignment values agree.
- Combine East and West values. Include systematic uncertainties to ensure agreement.
- If the alignments were done using the stars, the true north would drift with Earth's precession. Test if the alignment of the pyramids shifts linearly as a function of time.
- The astronomically predicted shift as a function of time is 0.274 arc min./year. Does the slope of the linear fit match the astronomically predicted value?
- If the stars used for N-S pyramid alignment pointed towards true north in 2467 BC, then what is your estimate of the alignment year of the Khufu pyramide (historically 2554 ± 100 BC) and its uncertainty?

5.2.1: This problem comes from a 2000 Nature paper by Kate Spence, which I have pursued a bit since… brilliant idea, but poor statistics and hypo-testing.

For example, nowhere in the paper is there any mention of the > 15 sigma discrepancy between East and West alignment of the Bent pyramide!

Problem 5.2

5.2 (14 points) The table below lists the North-South alignment of Egyptian pyramids (in arc minutes).

Pyramid	1.Meidum	2.Bent	3.Red	4.Khufu	5.Khafre	6.Menk.	7.Sahure	8.Nefer.
Align. year	2600 BC	2583 BC	2572 BC	2554 BC	2522 BC	2489 BC	2446 BC	2433 BC
East Align.	-20.6±1.0	-17.3±0.2	-8.7±0.2	-3.4±0.2	6.0±0.2	12.4±1.0	23±10	30±10
West Align.	-18.1±1.0	-11.8±0.2	–	-2.8±0.2	6.0±0.2	14.1±1.8	–	–

- Test to what extend the East (E) and West (W) alignment values agree.
- Combine East and West values. Include systematic uncertainties to ensure agreement.

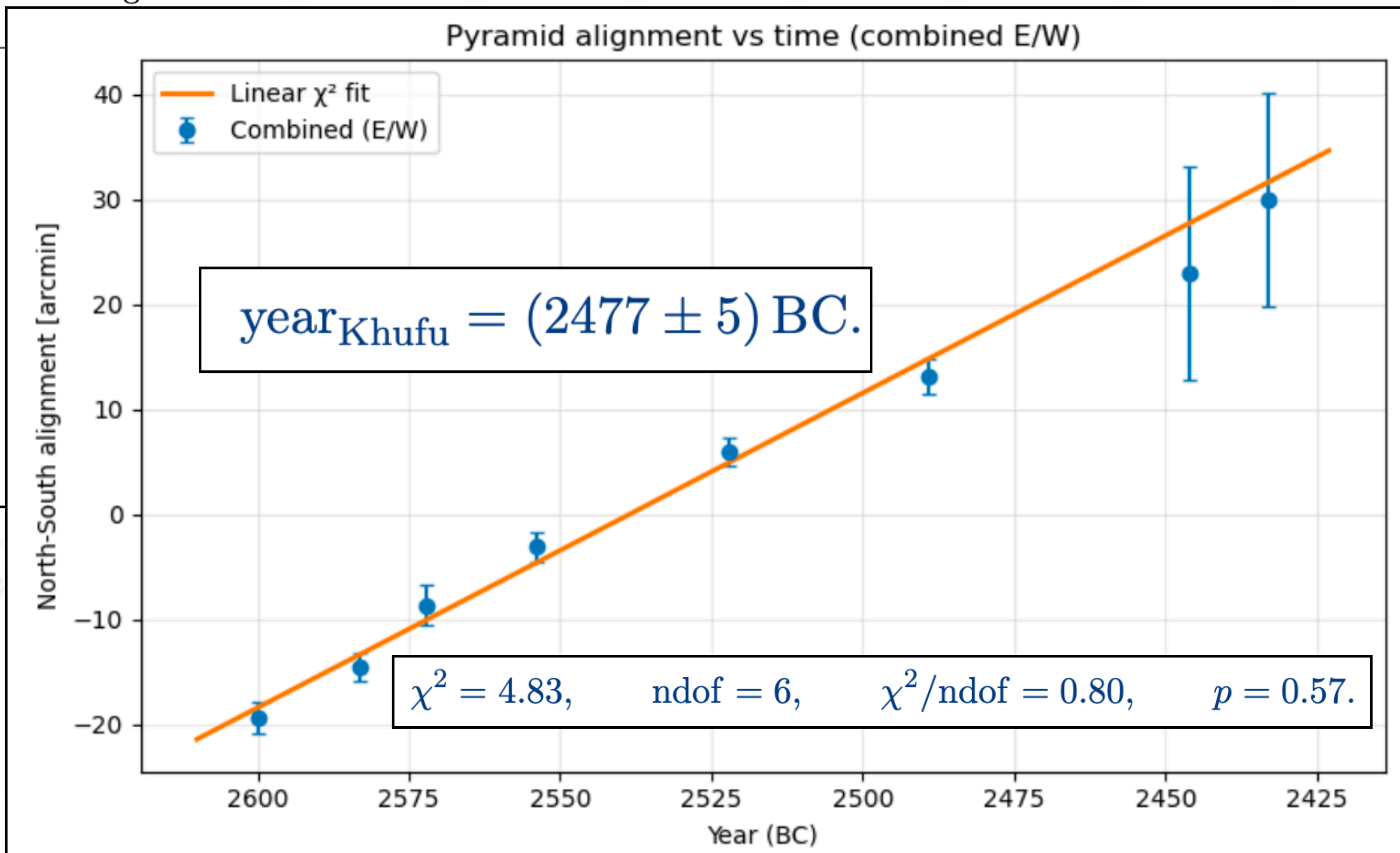
Pyramid	Meidum	Bent	Red	Khufu	Khafre	Menk	Sahure	Nefer.
z	1.77	19.45	–	2.12	0.00	0.83	–	–
p	0.08	0.00	–	0.03	1	0.41	–	–

Pyramid	Meidum	Bent	Khufu	Khafre	Menk
East	-20.6 ± 1.0	$-17.3 \pm 0.2 \pm \mathbf{1.8}$	$-3.4 \pm 0.2 \pm \mathbf{0.1}$	6 ± 0.2	12.4 ± 1.0
West	-18.1 ± 1.0	$-11.8 \pm 0.2 \pm \mathbf{1.8}$	$-2.8 \pm 0.2 \pm \mathbf{0.1}$	6 ± 0.2	14.1 ± 1.8
z	1.77	1.94	1.41	0.00	0.83
p	0.08	0.05	0.16	1	0.41
Comb.	-19.4 ± 0.7	-14.6 ± 1.4	-3.1 ± 0.2	6 ± 0.2	12.8 ± 0.9

Problem 5.2

5.2 (14 points) The table below lists the North-South alignment of Egyptian pyramids (in arc minutes).

Pyramid	1.Meidum	2.Bent	3.Red	4.Khufu	5.Khafre	6.Menk.	7.Sahure	8.Nefer.
Align. year	2600 BC	2583 BC	2572 BC	2554 BC	2522 BC	2489 BC	2446 BC	2433 BC
East Align.	-20.6±1.0	-17.3±0.2	-8.7±0.2	-3.4±0.2	6.0±0.2	12.4±1.0	23±10	30±10



$$\text{year}_{\text{Khufu}} = (2477 \pm 5)\ \text{BC}.$$

$$\chi^2 = 4.83, \quad \text{ndof} = 6, \quad \chi^2/\text{ndof} = 0.80, \quad p = 0.57.$$

5.2.5: If the hypothesis holds, and the right stars are known, and if there are no other major effects, then this allows the dating to become **very accurate** (±5y).

Problem 5.2

5.2 (14 points) The table below lists the North-South alignment of Egyptian pyramids (in arc minutes).

Pyramid	1.Meidum	2.Bent	3.Red	4.Khufu	5.Khafre	6.Menk.	7.Sahure	8.Nefer.
Align. year	2600 BC	2583 BC	2572 BC	2554 BC	2522 BC	2489 BC	2446 BC	2433 BC
East Align.	-20.6 ± 1.0	-17.3 ± 0.2	-8.7 ± 0.2	-3.4 ± 0.2	6.0 ± 0.2	12.4 ± 1.0	23 ± 10	30 ± 10
West Align.	-18.1 ± 1.0	-11.8 ± 0.2	–	-2.8 ± 0.2	6.0 ± 0.2	14.1 ± 1.8	–	–

- Test to what extend the East (E) and West (W) alignment values agree.
- Combine East and West values. Include systematic uncertainties to ensure agreement.
- If the alignments were done using the stars, the true north would drift with Earth's precession. Test if the alignment of the pyramids shifts linearly as a function of time.
- The astronomically predicted shift as a function of time is 0.274 arc min./year. Does the slope of the linear fit match the astronomically predicted value?
- If the stars used for N-S pyramid alignment pointed towards true north in 2467 BC, then what is your estimate of the alignment year of the Khufu pyramide (historically 2554 ± 100 BC) and its uncertainty?

Common mistakes/slips:

5.2.2: We have probably seen ~20 different ways to estimate the systematic uncertainties. A suggested method would be good.

5.2.3: Some people didn't know what to do with the pyramids missing the west alignment measurement and dropped them for the rest of the exercise.

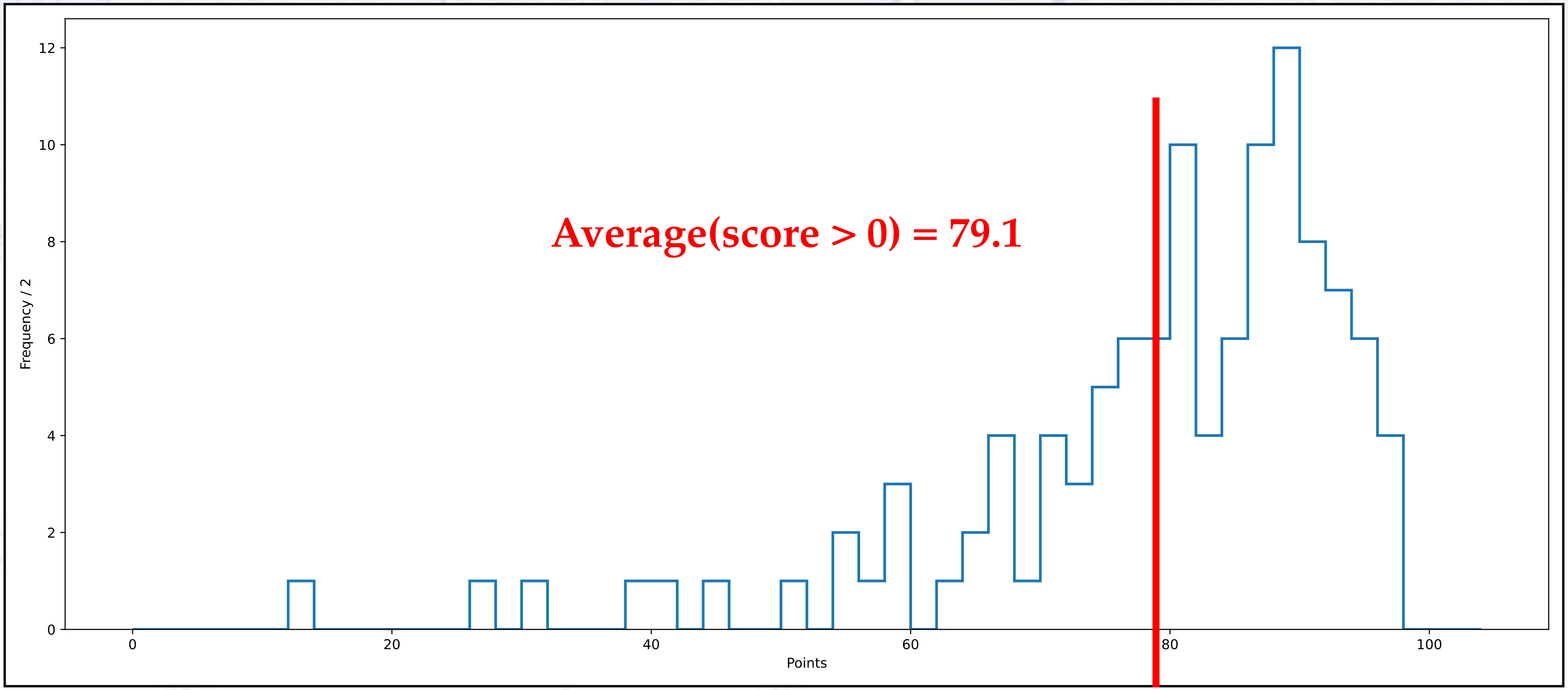
Your scores

General distribution

The distribution of points in the Problem Set was 79.1.

Last years, it was 72.1, 70.8, and 75.2, so “better than normally”. Thanks!

Notice, that the grading scale is not fixed, so nothing is “absolute”.



1.1.1	1.1.2	1.2.1	2.1.1	2.2.1	2.2.2	2.3.1	2.3.2	2.3.3	2.3.4	3.1.1	3.1.2	3.2.1	3.2.2	3.2.3	3.2.4	4.1.1	4.1.2	4.1.3	5.1.1	5.1.2	5.1.3	5.1.4	5.2.1	5.2.2	5.2.3	5.2.4	5.2.5	Total:
3	3	4	5	4	4	4	3	3	4	4	4	3	4	4	4	4	4	4	3	3	4	4	2	3	3	3	3	100
2.61	2.71	3.2	4.21	3.79	3.51	2.71	2.12	2.17	2.73	3.51	3.34	2.88	3.96	3.56	3.79	3.81	3.52	2.27	2.83	2.35	2.17	2.25	1.39	2.21	2.09	2.17	1.94	79.06
0.52	0.6	0.95	1.13	0.48	0.94	0.99	0.74	0.98	1.11	0.66	0.88	0.58	0.89	0.86	0.87	0.73	0.97	1.52	0.57	0.99	1.4	1.47	0.66	1.09	1.08	1.18	1.16	15.37