EPFL
Spring Semester 2021, **21.6.2021**

Probability and statistics for materials science
Prof. Philip Moll

EXAM

Time: The exam starts at 16:15 and finishes at 18:15.

Points: The exam consists of 4 questions. Each question and sub-steps have individual points associated to them, indicated in brackets.

General remarks:

1) Please provide your answers **only** on the supplied quad paper. The question sheets and your draft sheets will not be graded.
2) Explain your reasoning well to justify your answers. For all answers, the computations leading to the results have to be clear to achieve full points.
3) Drafting sheets and additional paper is supplied as well. **Drafting paper will not be graded**. Please write your name and page number on each sheet.
4) Round numerical values to the second digit ($\pi = 3.14$).
5) **Allowed supporting material for you are one double-sided A4 sheet of notes and a non-programmable calculator.**

Question	Points Total	Points achieved
1	17	
2	23	
3	19	
4	27	
Total	**86**	

1 Landing the Mars rover

If you ever flew a drone in your backyard, you can begin to imagine how difficult it must be to shoot a robot to Mars and have it operate - autonomously, in a very harsh environment and 300 million km away. You enjoy tackling hard problems, and you join the EPFL Mars rover challenge - land a rover on Mars in 2024. Your team is charged with designing the titanium landing gear of the robot. Your best design for the spring suspension survives the harsh impact on Mars only with a probability of $p = 90\%$.

a) [1 pt] Your rover design features 6 individually suspended wheels. What is the probability that all 6 wheels will survive the impact?

b) [1 pt] The design team tells you that the rover should remain operational if at least 5 of the 6 wheels survive. What is the probability that it will be able to drive on Mars?

c) [3 pt] You design an alternative set of suspensions that survive with a lower probability, $p_2 = 80\%$. However, the rover will remain operational even if at least 4 out of 6 wheels survive. Which design gives your rover a better chance to succeed?

d) [2 pt] Chances are not looking too good for something as expensive as a Mars expedition. When you are ready to give up, Elon Musk saves the day and offers you his new HYPER-GAMMA suspensions he designed for a rocket-propelled submarine. Elon claims his suspensions offer a superb reliability, $p_3 = 99\%$. Having worked on the engineering problem for a long time, you doubt that claim. Insulted by your doubts, Elon lands 100 rovers on Mars and records the numbers of wheels surviving the impact:

Number of surviving wheels	0	1	2	3	4	5	6
Number of rovers	0	0	0	0	0	29	71

Table 1: Data from the 100 rovers Elon landed on Mars. Every single one turned out operational. Impressive!

State the statistical test you would use to verify Elons claim of $p_3 = 99\%$, formulate the null hypothesis H_0 and the alternative hypothesis H_1 for this test.

e) [10 pt] Test your hypothesis identified in d) at a level of significance $\alpha = 0.05$. Does that data support the claim of $p_3 = 99\%$? *You may use the approximation* $(0.01)^n \approx 0$ *for* $n \geq 2$.

2 Urban architecture - in wood

Living in urban areas evolves around tall buildings that support a high population density. For such tall buildings, concrete and steel are ideal and optimized building materials - if it were not for the immense CO_2 footprint of concrete. Your startup challenges this problem by redesigning an old solution: wood nanocomposites as a natural building material. As part of your all-wood skyscraper project, your production plant plans to fabricate I-beams that can replace the standard steel beams on the market. The standard I-beam yield strength required by the building architect is 350MPa. Can your all-wood product replace this?

The production line makes 20 all-wood I-beams for yield strength testing. Here are your 20 test results:

Yield strength X_i (MPa)	341	344	345	346	346	350	352	354	354	358
	360	360	363	368	369	369	370	372	382	425

Table 2: Yield strength of your new all-wood I-beams. *Hint: this question can also be solved without typing all these numbers into your calculator.*

a) [3 pt] Compute the mean of the data in table 2. Unfortunately, your statistics software is broken. Instead of the mean, it computes for you $\sum_{i=1}^{20}(x_i - 10)^2 = 2476262$ and $\sum_{i=1}^{20} x_i^2 = 2618822$ from this dataset.

b) [3 pt] State the empirical quantiles $q_{18\%}, q_{25\%}, q_{50\%}, q_{75\%}$.

c) [5 pt] Analyze your dataset by drawing a boxplot. Describe briefly the relevant parameters of the boxplot, and give their numerical values. What do you notice about the data?

d) [5 pt] A further analysis shows that the datapoint at 425MPa was due to a measurement error, and you remove that datapoint. Next you want to build a model for the remaining n=19 datapoints a model based on a normal distribution. Estimate the standard deviation σ and mean μ from your data.

e) [5 pt] Assuming your yield strength follows a normal distribution $N(\mu, \sigma^2)$, state the probability that a random wooden beam exceeds the 350 MPa yield strength of steel (using the parameters of the previous task d). The engineers of your building require an acceptance rate of at least 70% for safe construction. Does your product make it?

f) [2 pt] State the empirical probability, $P(X \geq 350MPa)$, of a beam to fulfil this criterion based on your data. Does this number make sense compared to the result from your normal distribution model developed in e)?

3 Transport in quantum materials

The challenge in quantum materials research is to synthesize new materials in which quantum mechanical behavior is enhanced, and achieve new functionality thereby. In your study, you synthesize a conductive material, SmB_6, and plan to measure its electrical resistance. To this end, you apply a current, I, to the sample and measure the voltage, U.

The following table states your experimental results:

Applied current I(A)	0	1	2	3
Measured voltage U(V)	-0.1	0.9	2.3	9.6

Table 3: 4 datapoints of your resistance experiment

a) [6 pt] In a metal, you expect the voltage to be linear in current, given by the famous formula $U = R \cdot I$, where R is the unknown resistance of the sample. Apply the principle of least squares to this simple model $\hat{U}^{(1)} = RI$ and compute the coefficient R that best describes your data.

b) [5 pt] At a research conference, you learn of a new theory that predicts a novel conduction mechanism in your material. The theory predicts a response of your voltage to current in quadrature. Hence, you try a second model and describe your data by $\hat{U}^{(2)} = \beta I^2$. Apply the principle of least squares to model 2 and compute the coefficient β that best describes your data.

c) [1 pt] Are $\hat{U}^{(1)}$ and $\hat{U}^{(2)}$ nested models?

d) [7 pt] Compute the Error Sum of Squares (SSE) for both models. Which model explains more of the total error?

4 VegiSTEAK

In our battle with global warming, cattle farming plays an important role. A less CO_2 and methane intense way to obtain a juicy food item that could replace a steak would get us far. You experiment in your kitchen with mushrooms, yeast and algae; and finally make the breakthrough. You cannot believe how its impossible to distinguish your dinner from actual meat. You name your invention the VegiSTEAK!

To get seed funding, you invite 6 friends to a dinner party where you serve them both T-bone steaks and VegiSTEAKS. You ask each of your friends to rank the "meatiness" of both on a scale between 1 (least) to 10 (best). Here are your friends responses:

VegiSTEAK (1)	9	7	6	10	8	7
T-Bone (2)	4	9	6	8	3	4

Table 4: Meat-taste ranking of your friends. Grades between 1 (worst) and 10 (best).

a) [2 pt] For your VegiSTEAK (group 1), compute the group mean $\bar{X}_1$ and the standard deviation s_1. For the T-bone (group 2), the same calculation gives $\bar{X}_2 = 5.67$ and $s_2 = 2.42$.

b) [2 pt] You want to know if your VegiSTEAK tastes better than the T-bone. Formulate the null hypothesis H_0 and alternative hypothesis H_1 for a T-test. Justify your choice of one-sided vs. two-sided T-test.

c) [8 pt] Perform the T-test according to b). Is there a statistically relevant differences at a level of significance $\alpha = 0.05$?

Your friends encourage you to approach investors, and they too are impressed by your product! However, they caution that other meat-free products are on the market and its unclear how your product competes against them. You again invite your friends to a large BBQ, and serve your VegiSTEAK together with the leading brands of meatless grill products on the market. As you now have 3 groups of meatless products to compare, you chose to do an ANOVA analysis. Here is your data:

VegiSTEAK (1)	9	7	6	10	8	7
Quorn Steak (2)	6	6	5	4	6	5
Beyond Burger (3)	5	6	7	4	4	6

Table 5: Meat-taste ranking of your friends. Grades between 1 (worst) and 10 (best). *The results of group 1 are the same as in the previous part - no need to recompute mean and standard deviation.*

You use R to compute the statistical parameters of these new groups 2 and 3, and find $\bar{X}_2 = \bar{X}_3 = 5.34$, $s_2 = 0.82$ and $s_3 = 1.21$.

d) [7 pt] Compute the one-factor ANOVA table for this model.

e) [4 pt] Based on your ANOVA analysis, perform an F-test. Can you state a statistically significant taste difference between these products at a level of significance of $\alpha = 0.05$?

f) [4 pt] Compare and briefly discuss the results of the T-test and the ANOVA analysis. Where does your product stand?

Tableau de la fonction de répartition normale. La fonction $\Phi(z)$ est égale à la proportion des réalisations inférieures ou égales à z d'une varable aléatoire normale centrée et réduite.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0,00	0,500	0,72	0,764	1,44	0,9251	2,16	0,9846	2,88	0,99801	3,80	0,9999277
0,02	0,508	0,74	0,770	1,46	0,9279	2,18	0,9854	2,90	0,99813	3,84	0,9999385
0,04	0,516	0,76	0,776	1,48	0,9306	2,20	0,9861	2,92	0,99825	3,88	0,9999478
0,06	0,524	0,78	0,782	1,50	0,9332	2,22	0,9868	2,94	0,99836	3,92	0,9999557
0,08	0,532	0,80	0,788	1,52	0,9357	2,24	0,9875	2,96	0,99846	3,96	0,9999625
0,10	0,540	0,82	0,794	1,54	0,9382	2,26	0,9881	2,98	0,99856	4,00	0,9999683
0,12	0,548	0,84	0,800	1,56	0,9406	2,28	0,9887	3,00	0,99865	4,04	0,9999733
0,14	0,556	0,86	0,805	1,58	0,9429	2,30	0,9893	3,02	0,99874	4,08	0,9999775
0,16	0,564	0,88	0,811	1,60	0,9452	2,32	0,9898	3,04	0,99882	4,12	0,9999811
0,18	0,571	0,90	0,816	1,62	0,9474	2,34	0,9904	3,06	0,99889	4,16	0,9999841
0,20	0,579	0,92	0,821	1,64	0,9495	2,36	0,9909	3,08	0,99996	4,20	0,9999867
0,22	0,587	0,94	0,826	1,66	0,9515	2,38	0,9913	3,10	0,99903	4,24	0,9999888
0,24	0,595	0,96	0,831	1,68	0,9535	2,40	0,9918	3,12	0,99910	4,28	0,9999907
0,26	0,603	0,98	0,836	1,70	0,9554	2,42	0,9922	3,14	0,99916	4,32	0,9999922
0,28	0,610	1,00	0,841	1,72	0,9573	2,44	0,9927	3,16	0,99921	4,36	0,9999935
0,30	0,618	1,02	0,846	1,74	0,9591	2,46	0,9931	3,18	0,99926	4,40	0,9999946
0,32	0,626	1,04	0,851	1,76	0,9608	2,48	0,9934	3,20	0,99931	4,44	0,9999955
0,34	0,633	1,06	0,855	1,78	0,9625	2,50	0,9938	3,22	0,99936	4,48	0,9999963
0,36	0,641	1,08	0,860	1,80	0,9641	2,52	0,9941	3,24	0,99940	4,52	0,9999969
0,38	0,648	1,10	0,864	1,82	0,9656	2,54	0,9945	3,26	0,99944	4,56	0,9999974
0,40	0,655	1,12	0,869	1,84	0,9671	2,56	0,9948	3,28	0,99948	4,60	0,9999979
0,42	0,663	1,14	0,873	1,86	0,9686	2,58	0,9951	3,30	0,99952	4,64	0,9999983
0,44	0,670	1,16	0,877	1,88	0,9799	2,60	0,9953	3,32	0,99955	4,68	0,9999986
0,46	0,677	1,18	0,881	1,90	0,9713	2,62	0,9956	3,34	0,99958	4,72	0,9999988
0,48	0,684	1,20	0,885	1,92	0,9726	2,64	0,9959	3,36	0,99961	4,76	0,9999990
0,50	0,691	1,22	0,889	1,94	0,9738	2,66	0,9961	3,38	0,99964	4,80	0,9999992
0,52	0,698	1,24	0,893	1,96	0,9750	2,68	0,9963	3,40	0,99966	4,84	0,9999994
0,54	0,705	1,26	0,896	1,98	0,9761	2,70	0,9965	3,42	0,99969	4,88	0,9999995
0,56	0,712	1,28	0,900	2,00	0,9772	2,72	0,9967	3,44	0,99971	4,92	0,9999996
0,58	0,719	1,30	0,903	2,02	0,9783	2,74	0,9969	3,46	0,99973	4,96	0,9999996
0,60	0,726	1,32	0,907	2,04	0,9793	2,76	0,9971	3,48	0,99975	5,00	0,9999997
0,62	0,732	1,34	0,910	2,06	0,9803	2,78	0,9973	3,50	0,99977	5,04	0,9999998
0,64	0,739	1,36	0,913	2,08	0,9812	2,80	0,9974	3,52	0,99978	5,08	0,9999998
0,66	0,745	1,38	0,916	2,10	0,9821	2,82	0,9976	3,54	0,99980	5,12	0,9999998
0,68	0,752	1,40	0,919	2,12	0,9830	2,84	0,9977	3,56	0,99981	5,16	0,9999999
0,70	0,758	1,42	0,922	2,14	0,9838	2,86	0,9979	3,58	0,99983	5,20	0,9999999

Quantiles de la loi t_ν de Student

ν	$qt_\nu(95\%)$	$qt_\nu(97{,}5\%)$	$qt_\nu(99\%)$	ν	$qt_\nu(95\%)$	$qt_\nu(97{,}5\%)$	$qt_\nu(99\%)$
1	6,314	12,71	31,82	21	1,721	2,080	2,518
2	2,920	4,303	6,965	22	1,717	2,074	2,508
3	2,353	3,182	4,541	23	1,714	2,069	2,500
4	2,132	2,776	3,747	24	1,711	2,064	2,492
5	2,015	2,571	3,365	25	1,708	2,060	2,485
6	1,943	2,447	3,143	26	1,706	2,056	2,479
7	1,895	2,365	2,998	27	1,703	2,052	2,473
8	1,860	2,306	2,896	28	1,701	2,048	2,467
9	1,833	2,262	2,821	29	1,699	2,045	2,462
10	1,812	2,228	2,764	30	1,697	2,042	2,457
11	1,796	2,201	2,718	32	1,694	2,037	2,449
12	1,782	2,179	2,681	34	1,691	2,032	2,441
13	1,771	2,160	2,650	36	1,688	2,028	2,434
14	1,761	2,145	2,624	38	1,686	2,024	2,429
15	1,753	2,131	2,602	40	1,684	2,021	2,423
16	1,746	2,120	2,583	50	1,676	2,009	2,403
17	1,740	2,110	2,567	60	1,671	2,000	2,390
18	1,734	2,101	2,552	120	1,658	1,980	2,358
19	1,729	2,093	2,539	∞	1,645	1,960	2,326
20	1,725	2,086	2,528				

Table 1 – Quantiles des distributions t_ν. Pour ν suffisamment grand, on peut substituer aux quantiles de la loi t_ν les quantiles de la loi normale.

Quantiles de la loi khi–deux

ν	$q\chi^2_\nu(1\%)$	$q\chi^2_\nu(2{,}5\%)$	$q\chi^2_\nu(5\%)$	$q\chi^2_\nu(95\%)$	$q\chi^2_\nu(97{,}5\%)$	$q\chi^2_\nu(99\%)$
1	$0,0^3 1571$	$0,0^3 9821$	0,003932	3,841	5,024	6,635
2	0,02010	0,05064	0,1026	5,991	7,378	9,210
3	0,1148	0,2158	0,3518	7,815	9,348	11,34
4	0,2971	0,4844	0,7107	9,488	11,14	13,28
5	0,5543	0,8312	1,145	11,07	12,83	15,09
6	0,8721	1,237	1,635	12,59	14,45	16,81
7	1,239	1,690	2,167	14,07	16,01	18,48
8	1,646	2,180	2,733	15,51	17,53	20,09
9	2,088	2,700	3,325	16,92	19,02	21,67
10	2,558	3,247	3,940	18,31	20,48	23,21
11	3,053	3,816	4,575	19,68	21,92	24,72
12	3,571	4,404	5,226	21,03	23,34	26,22
13	4,107	5,009	5,892	22,36	24,74	27,69
14	4,660	5,629	6,571	23,68	26,12	29,14
15	5,229	6,262	7,261	25,00	27,49	30,58
16	5,812	6,908	7,962	26,30	28,85	32,00
17	6,408	7,564	8,672	27,59	30,19	33,41
18	7,015	8,231	9,390	28,87	31,53	34,81
19	7,633	8,907	10,12	30,14	32,85	36,19
20	8,260	9,591	10,85	31,41	34,17	37,57

TABLE 5 – Quantiles des distributions χ^2_ν. Le chiffres pour $\nu = 21, \ldots, 100$ se trouvent à la page suivante.

ν	$q\chi^2_\nu(1\%)$	$q\chi^2_\nu(2{,}5\%)$	$q\chi^2_\nu(5\%)$	$q\chi^2_\nu(95\%)$	$q\chi^2_\nu(97{,}5\%)$	$q\chi^2_\nu(99\%)$
21	8,897	10,28	11,59	32,67	35,48	38,93
22	9,542	10,98	12,34	33,92	36,78	40,29
23	10,20	11,69	13,09	35,17	38,08	41,64
24	10,86	12,40	13,85	36,42	39,36	42,98
25	11,52	13,12	14,61	37,65	40,65	44,31
26	12,20	13,84	15,38	38,89	41,92	45,64
27	12,88	14,57	16,15	40,11	43,19	46,96
28	13,56	15,31	16,93	41,34	44,46	48,28
29	14,26	16,05	17,71	42,56	45,72	49,59
30	14,95	16,79	18,49	43,77	46,98	50,89
32	16,36	18,29	20,07	46,19	49,48	53,49
34	17,79	19,81	21,66	48,60	51,97	56,06
36	19,23	21,34	23,27	51,00	54,44	58,62
38	20,69	22,88	24,88	53,38	56,90	61,16
40	22,16	24,43	26,51	55,76	59,34	63,69
50	29,71	32,36	34,76	67,50	71,42	76,15
60	37,48	40,48	43,19	79,08	83,30	88,38
70	45,44	48,76	51,74	90,53	95,02	100,4
80	53,54	57,15	60,39	101,9	106,6	112,3
90	61,75	65,65	69,13	113,1	118,1	124,1
100	70,06	74,22	77,93	124,3	129,6	135,8

TABLE 6 – Quantiles des distributions χ^2_ν, dont la figure A.8 montre quelques densités.
L'approximation normale est $q\chi^2_\nu(p) \approx \nu + \sqrt{2\nu}\, q_{\mathrm{normal}}(p)$, valable pour ν suffisamment grand. Pour $\nu = 20$, par exemple, cette formule donne $20 + 6,32 \times 1,64 = 30,37$ comme valeur approximative du 95%-quantile d'une loi χ^2_{20}.
Une meilleure approximation est celle de Wilson-Hilferty, définie par
$q\chi^2_\nu(p) \approx \nu\left(1 - 2/(9\nu) + \sqrt{2/(9\nu)}\, q_{\mathrm{normal}}(p)\right)^3$. Dans l'exemple précédent cette formule donne la valeur approximative de $20(0,99 + 0,11 \times 1,64)^3 = 31,36$ pour $q\chi^2_{20}(95\%)$.

Quantiles de la loi F_{ν_1,ν_2} de Fisher

	$\nu_1 = 1$	2	3	4	5	6	7	8	10	12	24	∞
$\nu_2 = 1$	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	241,9	243,9	249,1	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,40	19,41	19,45	19,50
3	10,13	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,786	8,745	8,639	8,526
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,964	5,912	5,774	5,628
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,735	4,678	4,527	4,365
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,060	4,000	3,841	3,669
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,637	3,575	3,410	3,230
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,347	3,284	3,115	3,928
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,137	3,073	2,900	2,707
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	2,978	2,913	2,737	2,538
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,854	2,788	2,609	2,404
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,753	2,687	2,505	2,296
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,671	2,604	2,420	2,206
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,602	2,534	2,349	2,131
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,544	2,475	2,288	2,066
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,494	2,425	2,235	2,010
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,450	2,381	2,190	1,960
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,412	2,342	2,150	1,917
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,378	2,308	2,114	1,878
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,348	2,278	2,082	1,843
21	4,325	3,467	3,072	2,840	2,685	2,573	2,488	2,420	2,321	2,250	2,054	1,812
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,297	2,226	2,028	1,783
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,275	2,204	2,005	1,757
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,255	2,183	1,984	1,733
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,236	2,165	1,964	1,711
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,220	2,148	1,946	1,691
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,204	2,132	1,930	1,672
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,190	2,118	1,915	1,654
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,177	2,104	1,901	1,638
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,165	2,092	1,887	1,622
32	4,149	3,295	2,901	2,668	2,512	2,399	2,313	2,244	2,142	2,070	1,864	1,594
34	4,130	3,276	2,883	2,650	2,494	2,380	2,294	2,225	2,123	2,050	1,843	1,569
36	4,113	3,259	2,866	2,634	2,477	2,364	2,277	2,209	2,106	2,033	1,824	1,547
38	4,098	3,245	2,852	2,619	2,463	2,349	2,262	2,194	2,091	2,017	1,808	1,527
40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,077	2,003	1,793	1,509
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	1,993	1,917	1,700	1,389
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,910	1,834	1,608	1,254
∞	3,841	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,831	1,752	1,517	1,000

TABLE 5 – Les 95%-quantiles, $qF_{\nu_1,\nu_2}(95\%)$, des distributions F_{ν_1,ν_2}.