Sheet 12

Please put your solution in the box of your corresponding instructor before **Friday, May 22nd** at **12h00** and name your sheet "<First name>_<Last name>.pdf".

Exercise 1. A train line connects two cities, A and B. Each day at 7h00, 1600 people arrive at Station A and want to board a train. Since this is too many people for one train, there are two trains waiting for them. Each person chooses one train at random (with equal probabilities, independent from other people), and they don't have time to change trains before they leave.

How many seats should each train have so that, with probability less than 0.003, none of the two trains is full? You can use the table at the end of the sheet, which present values of the function

$$\text{erfc} : x \mapsto \mathbb{P}[|X| \geq x],$$

where X is a Gaussian random variable with mean 0 and variance $\frac{1}{2}$.

Solution of exercise 1. Since 1600 is a large number, we will apply the Central Limit Theorem.

For $1 \leq k \leq 1600$, we denote by X_k the random variable equal to 0 if the person k boards the first train, and to 1 if they board the second train. Then X_k is a sequence of independent Bernoulli's with parameter $\frac{1}{2}$; the mean is $\frac{1}{2}$ and the variance is $\frac{1}{4}$. Both are finite so we can apply the Central Limit Theorem.

The number of people boarding the second train is $\sum_{k=1}^{1600} X_k$. By the Central Limit Theorem, the distribution of this random variable is close to $\mathcal{N}(800, 400)$ (indeed $1600\mathbb{E}[X_1] = 800$ and $1600\mathbb{V}[X_1] = 400$).

To use the table, we write $\mathcal{N}(800, 400) = 800 + 20\sqrt{2}\mathcal{N}(0, \frac{1}{2})$.

From the table, we know that the probability that a Gaussian random variable with mean 0 and variance $\frac{1}{2}$ is greater, in absolute value, than 2.1, is less than 0.003.

Since $20\sqrt{2} \times 2.1 \approx 59.4$, we conclude that the probability that 860 people want to board the same train is less than 0.003.

In conclusion, each train should have 860 seats. □

Exercise 2. We wish to prove a transfer theorem relatively to the CLT.

Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of real random variables and $m, \sigma > 0$ be real numbers, such that, as $n \to +\infty$,

$$\sqrt{n}(Y_n - m) \Rightarrow \mathcal{N}(0, \sigma^2).$$

(For instance, this happens if $Y_n = (X_1 + \ldots + X_n)/n$, where the X_i's are independent, identically distributed real random variables with mean m and variance σ^2).

Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function and suppose that g is differentiable at m. We wish to show that

$$\sqrt{n}(g(Y_n) - g(m)) \Rightarrow \mathbb{N}(0, (g'(m))^2\sigma^2).$$

1. Show that Y_n converges to m in probability.

2. Let $\Delta : \mathbb{R} \to \mathbb{R}$ be the following function:

$$\Delta : x \mapsto \begin{cases} \frac{g(x)-g(m)}{x-m} & \text{if } x \neq m \\ g'(m) & \text{if } x = m. \end{cases}$$

   Show that $\Delta(Y_n)$ converges in probability to $g'(m)$.

3. Conclude (you may use the fact that, if $X_n \Rightarrow X$ and $Z_n \to c$ in probability, then $X_n + Z_n$ and $X_n Z_n$ converge in distribution, respectively to $X + c$ and Xc).

Solution of exercise 2.

1. By the Portmanteau theorem, for all $C > 0$, as $n \to +\infty$, one has

$$\limsup \mathbb{P}[\sqrt{n}(X_n - m) \geq C] \leq \frac{1}{2}\operatorname{erfc}(C/2\sigma),$$

   In particular, for all $\epsilon > 0$, since for n large enough $\epsilon\sqrt{n} \geq C$,

$$\limsup \mathbb{P}[(X_n - m) \geq \epsilon] \leq \limsup \mathbb{P}[\sqrt{n}(X_n - m) \geq C] \leq \operatorname{erfc}(C/\sigma).$$

   In particular, since this is true for all $C > 0$, there holds, for all $\epsilon > 0$, as $n \to +\infty$,

$$\mathbb{P}[X_n - m \geq \epsilon] \to 0.$$

   Since the normal distribution is symmetric, one has

$$\sqrt{n}(m - X_n) \Rightarrow \mathbb{N}(0, \sigma^2)$$

   so that, repeating the previous lines, for all $\epsilon > 0$,

$$\mathbb{P}[X_n - m \leq -\epsilon] \to 0.$$

   Thus, X_n converges to m in probability.

2. It suffices to show that Δ is continuous on $\mathbb{R}$; then, we can apply the transfer theorem: if Δ is continuous, since $Y_n \Rightarrow m$, then $\Delta(Y_n) \Rightarrow \Delta(m) = g'(m)$. Since $\Delta(Y_n)$ converges in distribution to a constant, it converges in probabilities to the same constant.

   The function Δ is obviously continuous on $\mathbb{R} \setminus \{m\}$ because it is the quotient of two continuous functions, the second one being non-vanishing.

   By definition, the claim that g is differentiable at m is equivalent to the claim that $\Delta(x) \to \Delta(m)$ as $x \to m$. This concludes the proof.

3. As suggested by the two previous questions, we write

$$\sqrt{n}(g(Y_n) - g(m)) = \sqrt{n}(Y_n - m)\frac{g(Y_n) - g(m)}{Y_n - m} = \sqrt{n}(Y_n - m)\Delta(Y_n).$$

By Sheet 9, exercise 3, this converges in distribution to

$$g'(m)\mathcal{N}(0, \sigma^2) = \mathcal{N}(0, (g'(m))^2\sigma^2).$$

□

Exercise 3. You just read an article in a scientific journal claiming that pigeons in Zürich have a mean weight of 300g. You want to verify this for yourself, so you catch 40 pigeons in the street and weight them.

Your results are as follows: the mean weight of your 40 pigeons is 312g and the empirical variance (see previous sheet) is $(31\text{g})^2$. You know (from the previous sheet) that, at least if the weight of a pigeon follows a Gaussian distribution, the mean and variance of this distribution should be close to 312g and $(31\text{g})^2$, respectively.

1. Use your measured empirical variance, and suppose that the weight of pigeons in Zürich follows a Gaussian distribution with mean 300g and variance $(31\text{g})^2$. How likely is it that, if you catch 40 pigeons, their empirical mean weight differs by more than 12g from the expected value of 300g?

2. Would you say that your results invalidate the claim?

Solution of exercise 3.

1. We apply the Central Limit Theorem (Gaussian distributions have finite expectation and variance). For $1 \le k \le 40$, let X_k be a random variable following $\mathcal{N}(300, 31^2)$, and such that the X_k's are independent. Then, by the CLT, the distribution of $\frac{X_1+\ldots+X_{40}}{40}$ is close to that of $\mathcal{N}(300, 31^2/40) \approx 300 + 7\mathcal{N}(0, \frac{1}{2})$.

   From the table of the function erfc, the probability that this mean differs from 300 by more than 12 is $\text{erfc}(12/7) \approx 0.015$.

2. If the claim is right, then you were suspiciously unlucky when catching your pigeons! Either the claim is wrong, or for some reason the pigeons near your place are bigger than average. More pigeon-study is advised!

□

Exercise 4. Let $\lambda > 0$. Let $(X_n)_{n\ge1}$ be a sequence of independent, identically distributed random variables, such that, for every $k \in \mathbb{N}$,

$$\mathbb{P}(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

We define the sequence $(Z_n)_{n\ge1}$ of random variables as $Z_n = \prod_{m=1}^{n} X_m$.

1. Compute $P(Z_n \neq 0)$, conclude that $(Z_n)_{n\geq 1}$ converges in probability to zero.
2. Does one have almost sure convergence? Does one have convergence in L^1?
3. Let $W = \sum_{n=0}^{+\infty} Z_n$. Show that $W < \infty$ almost surely; what is $\mathbb{E}[W]$?

Solution of exercise 4.

1. For $Z_n \neq 0$, it is necessary and sufficient that all X_k's are $\neq 0$ for $1 \leq k \leq n$. Since they are independent, identically distributed, and $\mathbb{P}[X_1 \neq 0] = 1 - e^{-\lambda}$, one has

   $$\mathbb{P}[Z_n \neq 0] = (1 - e^{-\lambda})^n$$

   In particular, as $n \to +\infty$, one has $\mathbb{P}[Z_n \neq 0] \to 0$, so that $Z_n \to 0$ in probability.

2. We prove almost sure convergence using the Borel-Cantelli lemma on the sequence of independent random variables (X_k).

   Since

   $$\sum_{k=1}^{+\infty} \mathbb{P}[X_k = 0] = \sum_{k=1}^{+\infty} e^{-\lambda} = +\infty,$$

   by the second part of the Borel-Cantelli lemma, almost surely, there exists an infinite number of k's such that $X_k = 0$.

   The property $X_k = 0$ implies $Z_n = 0$ for all $n \geq k$, so, in particular, almost surely, for all n large enough $Z_n = 0$. Hence, almost surely, $Z_n \to 0$.

   To check convergence in L^1 of the sequence of positive random variables $(Z_n)_{n\in\mathbb{N}}$ to 0, we use the independence of the X_k's to compute

   $$\mathbb{E}[Z_n] = \prod_{k=1}^{n} \mathbb{E}[X_k] = \lambda^n.$$

   In particular, if $\lambda < 1$ then $Z_n \to 0$ in L^1, if $\lambda \geq 1$ then L^1 convergence is false.

3. Since $Z_n \to 0$ almost surely, then the sum $W = \sum_{n=0}^{+\infty} Z_n$ contains, almost surely, a finite number of terms. Moreover, each of the Z_n's is finite almost surely, so that we can conclude: W is finite almost surely.

   The expectation of W is

   $$\mathbb{E}[W] = \sum_{n=1}^{+\infty} \lambda^n.$$

   In particular, this expectation is finite if $\lambda < 1$ and infinite if $\lambda \geq 1$.

□

Complementary Error Function Table

x	erfc(x)	x	erfc(x)	x	erfc(x)	x	erfc(x)	x	erfc(x)	x	erfc(x)	x	erfc(x)
0	1.000000	0.5	0.479500	1	0.157299	1.5	0.033895	2	0.004678	2.5	0.000407	3	0.00002209
0.01	0.988717	0.51	0.470756	1.01	0.153190	1.51	0.032723	2.01	0.004475	2.51	0.000386	3.01	0.00002074
0.02	0.977435	0.52	0.462101	1.02	0.149162	1.52	0.031587	2.02	0.004281	2.52	0.000365	3.02	0.00001947
0.03	0.966159	0.53	0.453536	1.03	0.145216	1.53	0.030484	2.03	0.004094	2.53	0.000346	3.03	0.00001827
0.04	0.954889	0.54	0.445061	1.04	0.141350	1.54	0.029414	2.04	0.003914	2.54	0.000328	3.04	0.00001714
0.05	0.943628	0.55	0.436677	1.05	0.137564	1.55	0.028377	2.05	0.003742	2.55	0.000311	3.05	0.00001608
0.06	0.932378	0.56	0.428384	1.06	0.133856	1.56	0.027372	2.06	0.003577	2.56	0.000294	3.06	0.00001508
0.07	0.921142	0.57	0.420184	1.07	0.130227	1.57	0.026397	2.07	0.003418	2.57	0.000278	3.07	0.00001414
0.08	0.909922	0.58	0.412077	1.08	0.126674	1.58	0.025453	2.08	0.003266	2.58	0.000264	3.08	0.00001326
0.09	0.898719	0.59	0.404064	1.09	0.123197	1.59	0.024538	2.09	0.003120	2.59	0.000249	3.09	0.00001243
0.1	0.887537	0.6	0.396144	1.1	0.119795	1.6	0.023652	2.1	0.002979	2.6	0.000236	3.1	0.00001165
0.11	0.876377	0.61	0.388319	1.11	0.116467	1.61	0.022793	2.11	0.002845	2.61	0.000223	3.11	0.00001092
0.12	0.865242	0.62	0.380589	1.12	0.113212	1.62	0.021962	2.12	0.002716	2.62	0.000211	3.12	0.00001023
0.13	0.854133	0.63	0.372954	1.13	0.110029	1.63	0.021157	2.13	0.002593	2.63	0.000200	3.13	0.00000958
0.14	0.843053	0.64	0.365414	1.14	0.106918	1.64	0.020378	2.14	0.002475	2.64	0.000189	3.14	0.00000897
0.15	0.832004	0.65	0.357971	1.15	0.103876	1.65	0.019624	2.15	0.002361	2.65	0.000178	3.15	0.00000840
0.16	0.820988	0.66	0.350623	1.16	0.100904	1.66	0.018895	2.16	0.002253	2.66	0.000169	3.16	0.00000786
0.17	0.810008	0.67	0.343372	1.17	0.098000	1.67	0.018190	2.17	0.002149	2.67	0.000159	3.17	0.00000736
0.18	0.799064	0.68	0.336218	1.18	0.095163	1.68	0.017507	2.18	0.002049	2.68	0.000151	3.18	0.00000689
0.19	0.788160	0.69	0.329160	1.19	0.092392	1.69	0.016847	2.19	0.001954	2.69	0.000142	3.19	0.00000644
0.2	0.777297	0.7	0.322199	1.2	0.089686	1.7	0.016210	2.2	0.001863	2.7	0.000134	3.2	0.00000603
0.21	0.766478	0.71	0.315335	1.21	0.087045	1.71	0.015593	2.21	0.001776	2.71	0.000127	3.21	0.00000564
0.22	0.755704	0.72	0.308567	1.22	0.084466	1.72	0.014997	2.22	0.001692	2.72	0.000120	3.22	0.00000527
0.23	0.744977	0.73	0.301896	1.23	0.081950	1.73	0.014422	2.23	0.001612	2.73	0.000113	3.23	0.00000493
0.24	0.734300	0.74	0.295322	1.24	0.079495	1.74	0.013865	2.24	0.001536	2.74	0.000107	3.24	0.00000460
0.25	0.723674	0.75	0.288845	1.25	0.077100	1.75	0.013328	2.25	0.001463	2.75	0.000101	3.25	0.00000430
0.26	0.713100	0.76	0.282463	1.26	0.074764	1.76	0.012810	2.26	0.001393	2.76	0.000095	3.26	0.00000402
0.27	0.702582	0.77	0.276179	1.27	0.072486	1.77	0.012309	2.27	0.001326	2.77	0.000090	3.27	0.00000376
0.28	0.692120	0.78	0.269990	1.28	0.070266	1.78	0.011826	2.28	0.001262	2.78	0.000084	3.28	0.00000351
0.29	0.681717	0.79	0.263897	1.29	0.068101	1.79	0.011359	2.29	0.001201	2.79	0.000080	3.29	0.00000328
0.3	0.671373	0.8	0.257899	1.3	0.065992	1.8	0.010909	2.3	0.001143	2.8	0.000075	3.3	0.00000306
0.31	0.661092	0.81	0.251997	1.31	0.063937	1.81	0.010475	2.31	0.001088	2.81	0.000071	3.31	0.00000285
0.32	0.650874	0.82	0.246189	1.32	0.061935	1.82	0.010057	2.32	0.001034	2.82	0.000067	3.32	0.00000266
0.33	0.640721	0.83	0.240476	1.33	0.059985	1.83	0.009653	2.33	0.000984	2.83	0.000063	3.33	0.00000249
0.34	0.630635	0.84	0.234857	1.34	0.058086	1.84	0.009264	2.34	0.000935	2.84	0.000059	3.34	0.00000232
0.35	0.620618	0.85	0.229332	1.35	0.056238	1.85	0.008889	2.35	0.000889	2.85	0.000056	3.35	0.00000216
0.36	0.610670	0.86	0.223900	1.36	0.054439	1.86	0.008528	2.36	0.000845	2.86	0.000052	3.36	0.00000202
0.37	0.600794	0.87	0.218560	1.37	0.052688	1.87	0.008179	2.37	0.000803	2.87	0.000049	3.37	0.00000188
0.38	0.590991	0.88	0.213313	1.38	0.050984	1.88	0.007844	2.38	0.000763	2.88	0.000046	3.38	0.00000175
0.39	0.581261	0.89	0.208157	1.39	0.049327	1.89	0.007521	2.39	0.000725	2.89	0.000044	3.39	0.00000163
0.4	0.571608	0.9	0.203092	1.4	0.047715	1.9	0.007210	2.4	0.000689	2.9	0.000041	3.4	0.00000152
0.41	0.562031	0.91	0.198117	1.41	0.046148	1.91	0.006910	2.41	0.000654	2.91	0.000039	3.41	0.00000142
0.42	0.552532	0.92	0.193232	1.42	0.044624	1.92	0.006622	2.42	0.000621	2.92	0.000036	3.42	0.00000132
0.43	0.543113	0.93	0.188437	1.43	0.043143	1.93	0.006344	2.43	0.000589	2.93	0.000034	3.43	0.00000123
0.44	0.533775	0.94	0.183729	1.44	0.041703	1.94	0.006077	2.44	0.000559	2.94	0.000032	3.44	0.00000115
0.45	0.524518	0.95	0.179109	1.45	0.040305	1.95	0.005821	2.45	0.000531	2.95	0.000030	3.45	0.00000107
0.46	0.515345	0.96	0.174576	1.46	0.038946	1.96	0.005574	2.46	0.000503	2.96	0.000028	3.46	0.00000099
0.47	0.506255	0.97	0.170130	1.47	0.037627	1.97	0.005336	2.47	0.000477	2.97	0.000027	3.47	0.00000092
0.48	0.497250	0.98	0.165769	1.48	0.036346	1.98	0.005108	2.48	0.000453	2.98	0.000025	3.48	0.00000086
0.49	0.488332	0.99	0.161492	1.49	0.035102	1.99	0.004889	2.49	0.000429	2.99	0.000024	3.49	0.00000080