Exercice 0.1 **(a) (2 points)** A probability space for this experiment is $(\Omega, \mathcal{F}, \Pr)$, where the sample space

$$\Omega = \{\{1,2\},\{1,3\},\ldots,\{n-1,n\}\}$$

consists of all $n(n-1)/2$ unordered distinct pairs of numbers in the range $1,\ldots,n$, the event space $\mathcal{F}$ is the set of all subsets of Ω, so $|\mathcal{F}| = 2^{n(n-1)/2}$, and the probability distribution $P : \mathcal{F} \to [0,1]$ places equal probabilities $1/\{n(n-1)/2\}$ on each element of Ω.

There are $n-1$ pairs $\{1,2\},\{2,3\},\ldots,\{n-1,n\}$, and the probability of choosing one of them at random is $(n-1)/\{n(n-1)/2\} = 2/n$.

(b) (2 points) The total points X obtained at the exam follows a binomial distribution $B(20, 1/5)$. The probability sought is

$$\Pr(X \geq 10) = \sum_{x=10}^{20} \binom{20}{x} (1/5)^x (4/5)^{20-x}.$$

Full marks are obtained by giving this expression (or something equivalent), and a **BONUS (+1)** for working out $\Pr(X \geq 10) \approx 0.0026$.

(c) (2 points) The conditional density is

$$f(x \mid X \geq 2) = \frac{\lambda^x e^{-\lambda}/x!}{1 - e^{-\lambda} - \lambda e^{-\lambda}}, \quad x = 2, 3, \ldots,$$

where the denominator is $\Pr(X \geq 2) = 1 - \Pr(X + 0) - \Pr(X = 1)$. Hence

$$\mathrm{E}(X \mid X \geq 2) = \sum_{x=2}^{\infty} x f(x \mid X \geq 2) = \sum_{x=2}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!(1 - e^{-\lambda} - \lambda e^{-\lambda})} = \frac{\lambda(1 - e^{-\lambda})}{1 - e^{-\lambda} - \lambda e^{-\lambda}}$$

after a little algebra.

Alternatively (better) one can note that

$$\mathrm{E}(X) = \mathrm{E}(X \mid X \geq 2)\Pr(X \geq 2) + \mathrm{E}(X \mid X < 2)\Pr(X < 2),$$

i.e.,

$$\lambda = \mathrm{E}(X \mid X \geq 2)(1 - e^{-\lambda} - \lambda e^{-\lambda}) + \frac{\lambda e^{-\lambda}}{e^{-\lambda} + \lambda e^{-\lambda}} \times (e^{-\lambda} - \lambda e^{-\lambda}),$$

which yields

$$\mathrm{E}(X \mid X \geq 2) = \frac{\lambda(1 - e^{-\lambda})}{1 - e^{-\lambda} - \lambda e^{-\lambda}}.$$

(d) (2 points) This is the Weibull density. Integration gives $F(x) = 1 - \exp(-x^{\alpha})$ for $x > 0$, and the median m satisfies $F(m) = 1/2$, solution to which gives $m = (\log 2)^{1/\alpha}$.

(e) (3 points) We have $X_1, X_1 \overset{\text{iid}}{\sim} \exp(\lambda)$, and Then,

$$M_{X_1}(t) = \mathrm{E}(e^{tX_1}) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x}\, dx = \frac{\lambda}{\lambda - t}, \quad t < \lambda,$$

where the condition on t ensures that the integral is finite. Obviously $M_{X_1}(t) = M_{X_2}(t)$. Now

$$M_{X_1 - X_2}(t) = \mathrm{E}\{e^{t(X_1 - X_2)}\} = \mathrm{E}(e^{tX_1} e^{-tX_2}) = M_{X_1}(t) M_{X_2}(-t) = \frac{\lambda^2}{\lambda^2 - t^2}, \quad |t| < \lambda,$$

where the factorisation is possible because the variables are independent and the more restrictive condition on t holds because $t < \lambda$ and $-t < \lambda$ must both be satisfied for the integrals to be finite.

(f) (2 points) The required variance is

$$\begin{aligned}
\mathrm{var}(4+2X-Y) &= \mathrm{var}(2X-Y) = 2^2\,\mathrm{var}(X) + 2(2)(-1)\,\mathrm{cov}(x,y) + (-1)^2\,\mathrm{var}(Y) \\
&= 4\,\mathrm{var}(X) - 4\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}\mathrm{corr}(X,Y) + \mathrm{var}(Y) \\
&= 4 - 4\sqrt{2}\rho + 2 \\
&= 6 - 4\sqrt{2}\rho.
\end{aligned}$$

Now $\rho \in [-1, 1]$, so the variance is minimised when $\rho = 1$, and is then $6 - 4\sqrt{2} \approx 0.34$.

(g) (1 point) No, because $s^2 \propto \sum(x_j - \overline{x})^2$, so $s^2 = 0$ implies that all the x_j are equal to 0.3. In that case the interquartile range is 0.

(h) (2 points) 'The 95% confidence interval for the unknown is (L, U)' means that under the assumptions made, the probability that the unknown θ (say) is contained in the random interval (L, U) is 0.95, i.e., $\Pr\{\theta \in (L, U)\} = 0.95$. The unknown is regarded a fixed, and the random quantity here is the interval, which is supposed to vary in independent replications of the experiment that lead to the data from which (L, U) are computed.

Exercice 0.2 **(a) (5 points)** Write $S_1, \ldots, S_{50}$ for the amounts of soup put into small bowls and $L_1, \ldots, L_{40}$ for the amounts of soup put into large bowls. We can use the central limit theorem by assuming that 50 and 40 are sufficiently large to get a good normal approximation for the sums. If so, we have that, approximately,

$$\sum_{i=1}^{50} S_i \sim \mathcal{N}(50\times 300, 50\times 30^2) = \mathcal{N}(15, 0.045), \quad \sum_{i=1}^{40} L_i \sim \mathcal{N}(40\times 600, 40\times 60^2) = \mathcal{N}(24, 0.144),$$

where the last numbers are in litres and litres^2.

By properties of independent normal variables, the daily total amount of soup consumed has approximate distribution $T \sim \mathcal{N}(15+24, 0.045+0.144) = \mathcal{N}(39, 0.189)$, in litres and litres^2. The probability of not having enough soup is

$$\begin{aligned}
\Pr(T > 40) &= \Pr\{(T-39)/\sqrt{0.189} \geq (40-39)/\sqrt{0.189})\\
&\approx \Pr(Z \geq 2.3) \\
&= 1 - \Pr(Z \leq 2.3) \\
&= 1 - \Phi(2.3) \approx 1 - 0.989 = 0.011,
\end{aligned}$$

where $Z \sim \mathcal{N}(0, 1)$ and Φ is the cumulative distribution functon of Z.

b) **(3 points)** Sandwiches correspond to independent Bernoulli variables $X_1, \ldots, X_{100}$ with probabilities $1-e^{-t}$ of being sold after time t, so the number left at time t is $Y = \sum_i X_i \sim B(100, 1-e^{-t})$. The probability of having sold all sandwiches after 4 hours is therefore

$$\Pr(Y = 100) = \left(1 - e^{-4}\right)^{100} \approx 0.16.$$

c) **(2 points)** Let A denote the event that there are no more sandwiches, and B the event that there is no more soup. I seek the probability $\Pr(A^c \cup B^c) = 1 - \Pr(A \cap B) = 1 - 0.011 \times 0.16 \approx 0.998$. It is unlikely that I'll find nothing to eat.

Exercice 0.3 (a) **(3 points in total, be flexible)** (i) **(1 point)** This plot is a Q-Q plot, which is one way to compare a sample $X_1, \ldots, X_n$ with a theoretical distribution F (here the standard normal distribution).

The Q-Q plot displays the sample quantiles

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}$$

plotted against

$$F^{-1}\{1/(n+1)\}, F^{-1}\{2/(n+1)\}, \ldots, F^{-1}\{n/(n+1)\}.$$

The closer the graph is to a straight line, the more the data resemble a sample from F.

The slope and intercept of the line provide rough estimates of σ and μ for the hypothesised distribution.

(ii) **(1 point)** The points in the Q-Q plot are quite well aligned, so the sample looks like a sample from a normal distribution, though with so few observations it is hard to be sure.

(1 point) Judging from the plot it seems not impossible that Marco will be fined at the airport. The mean for his measurements appears to be greater than 20 kg.

(b) **(3 points in total, be flexible)** Let Y_i denote the weight of the ith object. The number of objects is large and each is weighed independently, so we can apply the central limit theorem and approximate the distribution of $W = \sum_{i=1}^{49} Y_i$ by a normal distribution with known variance the sum of the individual variances : $W \dot{\sim} N(\mu, 49 \times 0.004^2)$, where μ is the unknown true weight of the suitcase and we have observed $w = 19.9$.

We argue as follows. If $\mu = 20$, then Chris will (just) not be fined, but if $\mu > 20$ then he will be fined. Now if $\tau = 7 \times 0.004$ is the (known) standard deviation of W, then **(2 points)**

$$\Pr(W \leq 19.9) = \Pr\{(W-\mu)/\tau \leq (19.9-\mu)/\tau\} = \Phi\{(19.9-\mu)/\tau\} = \Phi(-3.57) \approx 0.0002$$

when $\mu = 20$. Since this probability would be even lower if $\mu > 20$, Chris is very unlikely to be fined. **(1 point)**

Another possibility is to compute a confidence interval. For example, a two-sided 99% confidence interval for μ is **(2 points)**

$$19.9 \pm z_{0.995} \times 7 \times 0.004 = 19.9 \pm 2.58 \times 0.028 = 19.9 \pm 0.072 = (19.828, 19.972) \text{ kg},$$

so a one-sided upper interval with coverage 99.5% would be $(0, 19.97)$ kg. This does not include 20 kg, so Chris is not likely to be fined. **(1 point)**

All the above discussion assumes that the airport scales are totally precise, so they can measure μ with no error, which seems far-fetched. On the other hand the weight measured at the airport will be regarded as the reference ...

Exercice 0.4 (a) **(3 points)** The likelihood is

$$L(a) = \prod_{i=1}^{n} (x_i/a) e^{-\sum_{i=1}^{n} x_i^2/(2a)}, \quad a > 0,$$

where the x_i are the heights observed in years $i = 1, \ldots, 8$. The log-likelihood is

$$\ell(a) = -n \log a + \sum_{i=1}^{n} \log x_i - \frac{1}{2a} \sum_{i=1}^{n} x_i^2, \quad a > 0,$$

its first two derivatives are

$$\ell'(a) = -\frac{n}{a} + \frac{1}{2a^2} \sum_{i=1}^{n} x_i^2, \quad \ell''(a) = \frac{n}{a^2} - \frac{1}{a^3} \sum_{i=1}^{n} x_i^2, \quad a > 0,$$

and solving $\ell'(a) = 0$ yields

$$\widehat{a} = \frac{1}{2n} \sum_{i=1}^{n} x_i^2, \quad \ell''(\widehat{a}) = -\frac{n}{\widehat{a}^2},$$

so $\widehat{a}$ is indeed the maximum likelihood estimate. Using the $n = 8$ observations yields $\widehat{a} = 2.42$.

(b) **(2 points)** The probability of a catastrophe in any given year equals

$$p(a) = \int_{6}^{\infty} f_X(x)\, dx = e^{-6^2/2a} = e^{-18/a},$$

and $\widehat{p} = p(\widehat{a}) \simeq 5.88 \times 10^{-4}$. On average, there would therefore be one catastrophe every $1/\widehat{p} \simeq 1700$ years, which is less than the insurance company suggests.

BONUS (+1) : The estimated probability of having more than one catastrophe in a thousand years is

$$1 - \left\{ (1 - \widehat{p})^{1000} + 1000(1 - \widehat{p})^{999} \widehat{p} \right\} \simeq 0.11,$$

which is not as small as the previous calculation might suggest.

(c) **(3 points)** The delta method and the fact that $\mathrm{var}(\widehat{a}) \simeq -1/\ell''(\widehat{a})$ give

$$\mathrm{var}(\widehat{p}) = \mathrm{var}\{p(\widehat{a})\} \simeq p'(\widehat{a})^2 \, \mathrm{var}(\widehat{a}) \simeq \{18 e^{-18/\widehat{a}} / \widehat{a}^2\}^2 \times \widehat{a}^2/n = \{18 e^{-18/\widehat{a}} / \widehat{a}\}^2/n,$$

so the standard error is $s = 18 e^{-18/\widehat{a}} / (n^{1/2} \widehat{a}) = 0.0015$. Thus an approximate confidence interval based on the large-sample (ahem, $n = 8$) normal distribution of $\widehat{a}$ would be

$$\widehat{p} \pm 1.96 \times s = (-0.0024, 0.0036),$$

which is not very useful, as it contains negative values. However with such a small sample it is not surprising that 'large-sample' results fail badly.

It would be better to compute the standard error for a and then transform it to one for p using the function $p(a)$, or (not quite so good) to compute the standard error for $\log \widehat{p} = 18/\widehat{a}$, get an interval for this quantity, and then go back to the original scale. There is a **BONUS (+1)** for students who propose something like this.