# École Polytechnique Fédérale de Lausanne

# MATH-234(a)– Midterm
# Probabilités et statistiques (GC)

Answer Sheet

Novembre 2021

1. **(a) (4 points)**

   **(i)** A boxplot is a graphical summary of a sample of observations. It has the following elements :
   - A (central) box which shows the sample lower and upper quartiles ($\widehat{q}(0.25)$ and $\widehat{q}(0.75)$).
   - A line inside the box which shows the sample median ($\widehat{q}(0.5)$).
   - The whiskers which show the most extreme observations lying inside the numbers $\widehat{q}(0.25) - C$ and $\widehat{q}(0.75) + C$, where $C = 1.5 \times \{\widehat{q}(0.75) - \widehat{q}(0.25)\}$. Observations that are more extreme than the whiskers are shown individually.

   **(ii)** The left panel shows that at least 75% of observations in each sample are below 15ppb (the upper part of boxes are below the horizontal line). Also, the spreads of samples 2 and 3 might be slightly larger than that for sample 1.

   In addition, (log) lead levels' tend to decrease when one increases the duration where tap water is let run, with the exception of the 3 largest observations in sample 2. This trend is at the distribution level ; one cannot draw conclusions at the house level. Indeed, (log) lead levels could increase for houses with low lead levels and decrease (drastically) for houses with high lead levels.

   **(b) (3 points)**

   **(i)** The calculation corresponds to $\Pr(X \geq 45)$ for a random variable $X \sim \mathit{Bernoulli}(n = 271, p = 0.1)$. In the context of Flint's tap water, this is the probability of observing 45 samples or more among the 271 for which the level of lead exceeds 15ppb, assuming that lead levels are independent between houses and have a probability of 0.1 to exceed 15ppb.

   **(ii)** The proportion of elements of sample 1 that exceed 15ppb is $45/271 = 16.61\%$. However, the fact that the observed proportion exceeds the limit of 10% could be due to chance, even if the true proportion is less than 10%. However this event has probability only 0.0003, which strongly suggests that the true level is over 15ppb and thus that action should be taken.

   **(c) (1 point)** This is a normal QQplot, and a perfect normal sample would fall on a straight line. The data seem to follow a slightly convex shape, at least for the lower portion, so the data seem to be slightly skewed to the right relative to a normal sample. However there is no evidence of outliers or heavy tails.

2. 1. [*2 marks*] Let $c_1$ and $c_2$ denote the outcomes of the first and second coin respectively. The sample space is $\Omega = \{(c_1, c_2) : c_1, c_2 \in \{H, T\}\} = \{(H,H),(T,T),(H,T),(T,H)\}$. The probabilities assigned to the elements of $\Omega$ are $\Pr\{(H,H)\} = p^2$, $\Pr\{(T,T)\} = (1-p)^2$ and $\Pr\{(H,T)\} = \Pr\{(T,H)\} = p(1-p)$.
   2. [*3 marks*] We have $A_1 = \{(H,H),(H,T)\}$, $A_2 = \{(H,H),(T,H)\}$ and $B = \{(H,T),(T,H)\}$. The corresponding probabilities are

   $$\Pr(A_1) = p^2 + p(1-p) = p, \quad \Pr(A_2) = p^2 + p(1-p) = p, \quad \Pr(B) = 2p(1-p).$$

   3. [*3 marks*] In this case $\Pr(A_1) = \Pr(A_2) = \Pr(B) = 1/2$, and

   $$\Pr(A_1 \cap A_2) = \Pr(A_1 \cap B) = \Pr(A_2 \cap B) = 1/4,$$

   so they are pairwise independent. However

   $$0 = \Pr(A_1 \cap A_2 \cap B) \neq \Pr(A_1)\Pr(A_2)\Pr(B)$$

   so they are not mutually independent.
   4. [*2 marks*] We have

   $$\Pr(A_1 \mid B) = \frac{\Pr(A_1 \cap B)}{\Pr(B)} = \frac{\Pr\{(H,T)\}}{2p(1-p)} = \frac{1}{2}.$$

   Even if the coin is biased, since $\Pr(A_1 \mid B) = \Pr(A_2 \mid B) = 1/2$, one can get the result of a fair coin by performing the experiment until $B$ occurs, and then checking whether the first coin shows a head.

3. [*3 marks*] Define the event $B$ 'the bus is late' and write $T$ for the time taken by the bus to reach the station; $\Pr(B) = 0.2$. Conditional on $B^c$, $T = 10$ minutes with probability one, and conditional on $B$, $T \sim N(10+3, 1)$. Then

$$\begin{aligned}
\Pr(T \leq 14) &= \Pr(T \leq 14 \mid B)\Pr(B) + \Pr(T \leq 14 \mid B^c)\Pr(B^c) \\
&= \Phi\{(14-13)/1\} \times 0.2 + 1 \times 0.8 \\
&= 0.8413 \times 0.2 + 0.8 \\
&\approx 0.9683.
\end{aligned}$$

**Alternatively** : let the random variable $L = 1$ if the bus is late, and $L = 0$ otherwise. So, $L \sim \text{Bern}(p = 0.2)$, and $X$ and $L$ are independent. Let the random variable $Y$ denote the delay of tomorrows' bus journey. Then $Y = LX$.
The probability that I will catch the train tomorrow is

$$\begin{aligned}
\Pr(Y \leq 4) &= \Pr(LX \leq 4) \\
&= \Pr(LX \leq 4 \mid L = 1)\Pr(L = 1) + \Pr(LX \leq 4 \mid L = 0)\Pr(L = 0) \\
&= \Pr(X \leq 4) \times p + 1 \times (1-p)
\end{aligned}$$

by the independence of $X$ and $L$. Now,

$$\Pr(X \leq 4) = \Pr\left(\tfrac{X-3}{1} \leq 1\right) = \Pr(Z \leq 1) = \Phi(1) \approx 0.8413.$$

since $Z$ is a standard normal random variable. Thus,

$$\Pr(Y \leq 4) = 0.2 \times \Phi(1) + 1 \times 0.8 \approx 0.9683.$$

[*2 marks*] We seek

$$\begin{aligned}
\Pr(T > 14 \mid T > 13) &= \tfrac{\Pr(T>14, T>13)}{\Pr(T>13)} \\
&= \tfrac{\Pr(T>14|B)\Pr(B)+\Pr(T>14|B^c)\Pr(B^c)}{\Pr(T>13|B)\Pr(B)+\Pr(T>13|B^c)\Pr(B^c)} \\
&= \tfrac{\Pr(T>14|B)\Pr(B)}{\Pr(T>13|B)\Pr(B)} \\
&= \tfrac{\Pr(T>14|B)}{\Pr(T>13|B)} \\
&= \tfrac{1-\Phi(14-13)}{1-\Phi(13-13)} \\
&= 2\{1-\Phi(1)\} \approx 0.3174.
\end{aligned}$$

**Alternatively** : the requested probability is

$$\begin{aligned}
\Pr(Y > 4 \mid Y \geq 3) &= \Pr(LX > 4 \mid LX > 3) = \Pr(LX > 4 \mid L = 1, X > 3) \\
&= \Pr(X > 4 \mid L = 1, X > 3) = \tfrac{\Pr(L>1, X>4)}{\Pr(L>1, X>3)} \\
&= \tfrac{\Pr(L>1)\Pr(X>4)}{\Pr(L>1)\Pr(X>3)} = \tfrac{1-\Pr(X\leq 4)}{1-\Pr(X\leq 3)} \\
&= \tfrac{1-\Phi(1)}{1-\Phi(0)} \approx 2(1-0.8413) = 0.3174.
\end{aligned}$$

[*2 marks*] We use Bayes' theorem and the result from (a) to obtain

$$\begin{aligned}
\Pr(B \mid T \leq 14) &= \tfrac{\Pr(T\leq 14|B)\Pr(B)}{\Pr(T\leq 14)} \\
&= \tfrac{\Phi(14-13)\times 0.2}{0.9683} \\
&\approx 0.1738.
\end{aligned}$$

**Alternatively** : using Bayes' theorem and the law of total probability, the requested probability is

$$\begin{aligned}
\Pr(L = 1 \mid Y \leq 4) &= \tfrac{\Pr(Y\leq 4|L=1)\Pr(L=1)}{\Pr(Y\leq 4)} \\
&= \tfrac{\Pr(LX\leq 4|L=1)\times p}{1-p\{1-\Phi(1)\}} \\
&= \tfrac{\Pr(X\leq 4)\times p}{1-p\{1-\Phi(1)\}} \\
&= \tfrac{0.2\times\Phi(1)}{1-0.2\{1-\Phi(1)\}} \\
&\approx 0.1738.
\end{aligned}$$

**Alternatively** : the requested probability is

$$\begin{aligned}
\Pr(L=1 \mid Y \leq 4) &= \Pr(L=1 \mid LX \leq 4) = \Pr(L=1 \mid \{L=0\} \cup \{X \leq 4\}) \\
&= \tfrac{\Pr(\{L=1\} \cap [\{L=0\} \cup \{X \leq 4\}])}{\Pr(\{L=0\} \cup \{X \leq 4\})} \\
&= \tfrac{\Pr(\{L=1\} \cap \{X \leq 4\})}{\Pr(L=0) + \Pr(X \leq 4) - \Pr(\{L=0\} \cap \{X \leq 4\})} \\
&= \tfrac{\Pr(L=1)\Pr(X \leq 4)}{\Pr(L=0) + \Pr(X \leq 4) - \Pr(L=0)\Pr(X \leq 4)} \\
&= \tfrac{0.2 \times \Phi(1)}{0.8 + \Phi(1) - 0.8 \times \Phi(1)} \approx 0.1738.
\end{aligned}$$

[*3 marks*] The expected bus journey duration is

$$\begin{aligned}
\mathrm{E}(T) &= \mathrm{E}(T \mid B)\Pr(B) + \mathrm{E}(T \mid B^c)\Pr(B^c) \\
&= 13 \times 0.2 + 10 \times 0.8 \\
&= 10 + 3 \times 0.2 = 10.6 \text{ minute,}
\end{aligned}$$

and the expected arrival time at the station is 7.55 and 36 seconds.

**Alternatively** : from the wording, the joint density function of $L \sim \mathrm{Bern}(p=0.2)$ and $X \sim N(\mu = 3, \sigma^2 = 1)$ is

$$f_{L,X}(l,x) = \{pI(l=1) + (1-p)I(l=0)\} \times \phi_{\mu,\sigma^2}(x), \quad l \in \{0,1\}, x \in \mathbb{R},$$

where $I(\cdot)$ is the indicator function, and $\phi_{\mu,\sigma^2}(x)$ denotes the density function of the normal distribution with mean $\mu$ and variance $\sigma^2$. So, the expected delay is

$$\begin{aligned}
\mathrm{E}(Y) &= \mathrm{E}(LX) \\
&= \int_{-\infty}^{\infty} \sum_{l=0}^{1} lx \times [\{pI(l=1) + (1-p)I(l=0)\} \times \phi_{\mu,\sigma^2}(x)] \,\mathrm{d}x \\
&= p \int_{-\infty}^{\infty} x\phi_{\mu,\sigma^2}(x) \,\mathrm{d}x = p\mathrm{E}(X) = 0.6 \text{ minute,}
\end{aligned}$$

and the expected arrival time at the station is 7.55 and 36 seconds.

**Alternatively** : by the law of iterated expectations, the expected delay is

$$\mathrm{E}(Y) = \mathrm{E}(LX) = \mathrm{E}_L\{\mathrm{E}(LX \mid L)\}.$$

Now, by the independence of $X$ and $L$,

$$\mathrm{E}(LX \mid L=k) = \mathrm{E}(kX \mid L=k) = k\mathrm{E}(X \mid L=k) = k\mathrm{E}(X) = 3k,$$

and therefore $\mathrm{E}(IX \mid L) = 3I$. Thus,

$$\mathrm{E}(Y) = \mathrm{E}_L(3L) = 3p = 0.6 \text{ minute,}$$

and the expected arrival time at the station is 7.55 and 36 seconds.

4. L'évènement "la pièce a survécu jusqu'au temps $t$" pet s'écrire comme $\{T \geq t\}$. Ici, on utilise le théorème de Bayes :

$$\begin{aligned}
\pi(t) = \Pr(\text{piratée} \mid T \geq t) \\
&= \frac{\Pr(T \geq t \mid \text{piratéePr(piratée)}}{\Pr(T \geq t \mid \text{piratée})\Pr(\text{piratée}) + \Pr(T \geq t \mid \text{non-piratée})\Pr(\text{nonpiratée})} \\
&= \frac{\exp(-5t) \cdot \frac{1}{4}}{\exp(-5t) \cdot \frac{1}{4} + \exp(-2t) \cdot \frac{3}{4}} \\
&= \frac{1}{1 + 3\exp(3t)}.
\end{aligned}$$

On constate donc que $\pi(t)$ tend zers zéro lorsque $t$ tend vers l'infini, c'est-à-dire plus longtemps la pièce survit, moins il y a de chances qu'elle ait été piratée.