Statistiques Inférentielles II
Correction du TD 0 : Auto évaluation

Licence 3 MIASHS - IDS (2021-2022)

Jairo Cuglirari, Guillaume Metzler
Institut de Communication (ICOM)
Université de Lyon, Université Lumière Lyon 2
Laboratoire ERIC UR 3083, Lyon, France

jairo.cugliari@univ-lyon2.fr guillaume.metzler@univ-lyon2.fr

Exercice 1 : Variable aléatoire continue

Soit X une variable aléatoire distribuée selon une loi uniforme sur l'intervalle $[a, b],\ a < b$.

1. Quelle est sa fonction de densité ?

   La loi uniforme se caractérise par le fait que la densité prend la même valeur quelque soit l'endroit où nous nous trouvons sur le support de cette distribution.

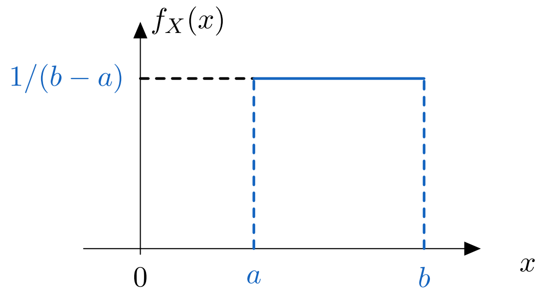
   Cela exclu immédiatement la réponse c) dont la fonction dépend de la valeur de x. En outre, cette fonction f ne définie par une densité.

   Les réponses a) et b) définissent bien une densité et vérifie bien la définition énoncée précédemment. En effet,

   $$\int_a^b f_X(x)\, dx = \int_a^b \frac{1}{b-a}\, dx = \frac{1}{b-a}\int_a^b dx = \frac{1}{b-a}\,[x]_a^b = 1.$$

   Il suffit remplacer a et b par les valeurs 0 et 1 respectivement pour la réponse a).
2. Quelle est la représentation graphique de sa fonction de densité ?

En reprenant l'explication ainsi que la réponse à la question précédente, seul le graphique b) représente la densité (ou fonction de masse) de la variable aléatoire X.



3. Connaissez vous sa fonction génératrice de moments ?

La fonction génératrice des moments M_X d'une variable aléatoire X est définie pour tout $t \in \mathbb{R}$ par

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \int_{\mathbb{R}} e^{tx} f_X(x)\, dx,$$

où t est tel que cette espérance existe.

Comme son nom l'indique, cette fonction permet de déterminer très aisément les différents moments d'une variable aléatoire X.
Pour cela, il suffit d'utiliser un développement en série entière de la fonction exponentielle, *i.e.*

$$e^{tx} = 1 + \frac{(tx)^1}{1!} + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \frac{(tx)^4}{4!} + \ \ldots = \sum_{k=0}^{\infty} \frac{(tx)^k}{k!}.$$

Mais quel est le lien avec les moments d'une variable aléatoire ? Nous y sommes presque ! On rappelle aussi que le moment d'ordre k, noté m_k d'une variable aléatoire X de densité f_X est donné par

$$m_k = \mathbb{E}[X^k] = \int_I x^k f_X(x)\, dx,$$

où I désigne le support de la variable aléatoire X, *i.e.* son intervalle de définition où la fonction prend une valeur non nulle. En observant la définition des moments et le développement en série entière de la fonction exponentielle, on commence à voir le lien qui se dessine, ainsi :

$$M_X(t) = \ \mathbb{E}\left[e^{tX}\right],$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} e^{tx} f_X(x)\, dx, \\
&\downarrow \text{ développement en série entière de l'exponentielle} \\
&= \int_I \left(1 + \frac{(tx)^1}{1!} + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \frac{(tx)^4}{4!} + \ \dots\right) f_X(x)\, dx, \\
&\downarrow \text{ on peut échange la somme et l'intégration} \\
&= \int_I f_X(x)\, dx + \int_I \frac{(tx)^1}{1!} f_X(x)\, dx + \int_I \frac{(tx)^2}{2!} f_X(x)\, dx + \int_I \frac{(tx)^3}{3!} f_X(x)\, dx + \ \dots\ , \\
&\downarrow \text{ définition du moment d'ordre } k \\
&= \ 1 + \frac{t^1 m_1}{1!} + \frac{t^2 m_2}{2!} + \frac{t^3 m_3}{3!} + \ \dots\ .
\end{aligned}
$$

A partir de la fonction $M_X(t)$ on peut donc aisément calculer le moment d'ordre k, plus précisément :

$$m_k = M_X^{(k)}(0),$$

ou $M_X^{(k)}$ désigne la dérivée d'ordre k de la fonction génératrice des moments. Avec cette définition, on peut donc calculer l'espérance de la variable aléatoire X comme suit :

$$\mathbb{E}[X] = m_1 = M_X'(0).$$

De, même pour la variance, nous avons :

$$\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = M_X"(0) - M_X'(0)^2.$$

Revenons maintenant à notre question, et appliquons ce qui précède

$$
\begin{aligned}
M_X(t) = &\ \mathbb{E}\left[e^{tX}\right], \\
= &\int_a^b e^{tx} \frac{1}{b-a}\, dx, \\
= &\ \frac{1}{b-a}\left[\frac{1}{t} e^{tx}\right]_a^b, \\
= &\ \frac{1}{t(b-a)}\left(e^{tb} - e^{ta}\right).
\end{aligned}
$$

4. Quelle est sa fonction de répartition ?

La fonction de répartition F_X d'une variable aléatoire X est définie pour tout $t \in \mathbb{R}$ par $\mathbb{P}[X \leq t]$, ainsi

$$F_X(t) = \mathbb{P}[X \leq t] = \int_{-\infty}^{t} f_X(x)\, dx.$$

Cette fonction prends ses valeurs $[0, 1]$. Elle jouera un rôle primordial dans la suite du cours portant sur l'estimation non paramétrique.

Pour une loi uniforme, cette fonction de répartition est

$$\begin{aligned} F_X(t) &= \int_{-\infty}^{t} \frac{1}{b-a}\, dx, \\ &= \frac{1}{b-a}\, [x]_a^t, \\ &= \frac{1}{b-a}\, (x-a). \end{aligned}$$

Exercice 2 : Variable aléatoire discrète

Soit X une variable aléatoire distribuée selon une loi de géométrique, c'est-à-dire la loi du nombre d'épreuves de Bernoulli indépendantes de probabilité de succès $p \in]0, 1[$ nécessaire pour obtenir le premier succès.

1. Quel est le support de X ?

   Théoriquement, il peut nécessiter un nombre infiniment grand d'expériences pour que le premier succès intervienne. En outre, s'agissant d'une variable aléatoire discrète, son support est donc l'ensemble $\mathbb{N}^\star$. En effet, 0 n'est pas inclus car il faut bien effectuer au moins une expérience pour qu'un succès intervienne.

2. Quelle est sa fonction de masse ?

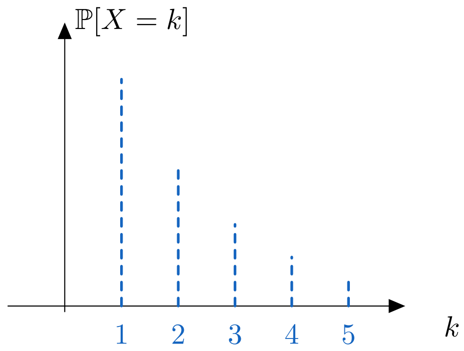
   D'après la définition de la loi géométrique, la probabilité pour que le premier succès intervienne au bout de $k \geq 1$ expérience(s) est

   $$\mathbb{P}[X = k] = q^{k-1}p = (1-p)^{k-1}p.$$

   En effet, pour que le premier succès intervienne au bout de $k \geq 1$ expérience(s), il faut que les $k-1$ premières soient un échec. On notera donc que l'ordre à une importance ici.

3. Quelle est la représentation graphique de sa fonction de masse ?

   Pour illustrer la fonction de masse de cette variable aléatoire, nous prendrons $p = 0.4$, donc $q = 1 - p = 0.6$ pour des valeurs de k allant de 1 à 5.



4. Connaissez vous sa fonction génératrice de moments ?
   La fonction génératrices de moments $M_X(t)$ d'une variable aléatoire X discrète est donnée, pour tout $t \in [0,1]$ par

$$M_X(t) = \sum_{k=0}^{\infty} \mathbb{P}[X=k] t^k$$

   On va appliquer la définition présentée à l'exercice précédent. Ainsi pour tout $t \in [0, \ln(1/q)]$ nous avons

$$\begin{aligned}
M_X(t) &= \mathbb{E}\left[e^{tX}\right], \\
&= \int_{\mathbb{R}_+^\star} e^{tk} q^{k-1} p \, dk, \\
&\downarrow \text{ on utilisera la mesure de comptage pour calculer cette intégrale} \\
&= \sum_{k=1}^{\infty} e^{tk} q^{k-1} p, \\
&\downarrow \text{ on factorise par } e^t \\
&= pe^t \sum_{k=1}^{\infty} e^{t(k-1)} q^{k-1}, \\
&\downarrow \text{ on change l'indice de sommation et on factorise} \\
&= pe^t \sum_{k=0}^{\infty} (qe^t)^k, \\
&\downarrow \text{ notre série géométrique est bien convergente car } qe^t \in [0,1] \\
&= pe^t \frac{1}{1-qe^t}.
\end{aligned}$$

5. Quelle est sa fonction de répartition ?

La fonction de répartition garde la même définition que dans le cas continue, on remplacera simplement l'intégrale par une somme partielle. Donc pour tout $s \in \mathbb{N}^\star$

$$\begin{aligned}
F_X(s) &= \mathbb{P}[X \leq s], \\
&= \int_0^s q^{k-1} p \, dk, \\
&= \sum_{k=1}^{s} q^{k-1} p, \\
&= p \sum_{k=0}^{s} q^k, \\
&\downarrow \text{ somme des termes d'une suite géométrique} \\
&= p \frac{1-q^s}{1-q}, \\
&\downarrow \text{ car } p = 1-q \\
&= 1 - q^s.
\end{aligned}$$

Exercice 3 : Principes basiques de l'inférence statistique

Dans la construction dune pièce par un procédé industriel, une machine produit de boulons de diamètre 9mm et moyenne avec un écart-type de 0.3mm. Le service de contrôle de qualité est contacté car dans un échantillon de 64 pièces produites, on a observé un diamètre moyen de 9.2mm. Le service de contrôle de qualité doit déterminer si la machine n'est plus calibrée.

1. Définir la variable aléatoire d'intérêt que nous noterons X.

   La variable aléatoire X étudiée ici est le diamètre des boulons produits par la machine.
2. Identifier $\mu = \mathbb{E}[X]$ et $\sigma^2 = Var[X]$, ainsi que les équivalents empiriques $\bar{x}$ et s^2.

   La moyenne du processus de production est de $\mu = 9$mm, l'écart-type associé au processus de production est de $\sigma = 0.3$mm.
   La moyenne empirique évaluée sur un échantillon de 64 pièces est de 9.2mm, nous n'avons en revanche aucune information relative à la variance.
3. Pour faire le test, les services de contrôle de qualité considèrent les hypothèses

$$H_0 : \mu = 9 \quad \text{vs} \quad \mu > 9.$$

et la statistique du test de la forme $\dfrac{\bar{X}-9}{\sqrt{Var[\bar{X}]}}$.
Identifier la distribution de cette quantité. Quelles sont les hypothèses nécessaires ?

Comme nous connaissons la variance associée au processus de fabrication, que l'échantillon est de taille raisonnable, *i.e. supérieure ou égale à* 30, la statistique de test est distribuée selon une **loi normale centrée-réduite**.

Nous ne savons pas si l'échantillon étudié est gaussien. En revanche, l'échantillon étudié est grand, on peut donc supposer que la variable aléatoire $\bar{X}$ est distribuée selon une loi normale de paramètres μ et $\dfrac{\sigma^2}{n}$.

4. Définir la région critique de ce test.

La formulation du test nous invite à faire un test unilatéral supérieur, la région critique est donc définie par l'intervalle

$$[z_{1-\alpha}, +\infty[\,,$$

où α désigne le risque d'erreur du test (ou le risque de première espèce), appelé aussi seuil de signification du test et $z_{1-\alpha}$ désigne le quantile d'ordre $1-\alpha$ de la loi normale centrée réduite.

5. Conclure sur le besoin de réaliser une maintenance à la machine.

L'énoncé ne le précise pas, mais on effectuera le test au seuil $\alpha = 5\%$. On va simplement réaliser notre application numérique. La valeur de notre statistique de test est

$$\frac{9.2-9}{\frac{0.3}{\sqrt{64}}} = \frac{0.2}{\frac{0.3}{8}} = 8\frac{2}{3} = \frac{16}{3}.$$

Or $z_{0.95} = 1.65$ et $z_{0.95} < \dfrac{16}{3}$, on peut donc rejeter l'hypothèse nulle au risque α et conclure que la machine n'est plus correctement calibrée et qu'elle nécessite une maintenance.