Eléments du calcul des probabilités

Répétition 5

19 avril 2016

Correction de l'exercice suggéré 4.11

Une variable aléatoire $\mathcal{X}$ est uniforme sur $[-\frac{1}{2}, \frac{1}{2}]$. Une variable aléatoire $\mathcal{Y}$ est exponentielle de moyenne $\frac{1}{a}$ *sur* $[0, \infty[$ $(a > \frac{1}{2})$. Ces deux variables étant indépendantes, quelle est la probabilité pour que $\mathcal{X} + \mathcal{Y} < a$?

Exercices

5.1. EXAMEN JANVIER 2012
Les variables aléatoires $\mathcal{X}$ et $\mathcal{Y}$ sont indépendantes et distribuées uniformément respectivement dans $[0, a]$ et $[0, b]$, avec $a > 0$ et $b > 0$.

a) Déterminez la densité de probabilité de $\mathcal{X}\mathcal{Y}$.

b) Calculer $P(\mathcal{X}\mathcal{Y} \geq \frac{ab}{e} | \mathcal{X}\mathcal{Y} \leq \frac{ab}{2})$.

5.2. Un ingénieur souhaite analyser le coût $\mathcal{Y}$ de son trajet domicile-travail en fonction de deux paramètres indépendants : le choix $\mathcal{X}_1$ du carburant (C_1 ou C_2) et le choix $\mathcal{X}_2$ du trajet (T_1 ou T_2).

Le carburant C_1 est un diesel classique coûtant 1,4 €/l, tandis que le carburant C_2 est un diesel premium coûtant 1,5 €/l. Le trajet T_1 est le plus court (45km) mais il passe en ville et prend souvent plus de temps que le trajet T_2 (58km) qui est principalement sur autoroute (le trafic y est généralement bien moins dense). On suppose que l'ingénieur a autant de chance de choisir le carburant C_1 que C_2 et autant de chance de choisir le trajet T_1 que T_2. La table 1 donne la consommation de la voiture en fonction des paramètres.

Table 1 – Consommation de la voiture

	Carburant C_1	Carburant C_2
Trajet T_1	Espérance : 8 l/100km Ecart-type : 0,5 l/100km	Espérance : 6,8 l/100km Ecart-type 0,6 l/100km
Trajet T_2	Espérance : 7,4 l/100km Ecart-type : 0,4 l/100km	Espérance : 6,3 l/100km Ecart-type : 0,4 l/100km

a) Calculez $E\{\mathcal{Y}\}$ et $V\{\mathcal{Y}\}$.

b) Calculez $V\{E\{\mathcal{Y}|\mathcal{X}_1\}\}$ et $E\{V\{\mathcal{Y}|\mathcal{X}_1\}\}$. $\mathcal{X}_1$ est-elle une bonne variable explicative de $\mathcal{Y}$? Justifiez.

c) Calculez $V\{E\{\mathcal{Y}|\mathcal{X}_2\}\}$ et $E\{V\{\mathcal{Y}|\mathcal{X}_2\}\}$. $\mathcal{X}_2$ est-elle une meilleure variable explicative de $\mathcal{Y}$ que $\mathcal{X}_1$? Justifiez.

d) Représentez graphiquement les différentes valeurs calculées aux points précédents, en illustrant le théorème de la variance totale.

e) En supposant que toutes les distributions de la table 1 sont gaussiennes, déterminez l'allure de la distribution de $\mathcal{Y}$. Pour ce faire, dessinez l'arbre de décision auquel le conducteur est confronté, lorsque son choix initial est le carburant et son second choix le trajet.

f) Quel est le meilleur choix de paramètres pour optimiser le coût ? Optimise-t-il aussi la variance ? Si l'ingénieur ne connaît pas les valeurs de la table 1, comment peut-il les estimer ?

Exercices suggérés

5.3. Soit $\mathcal{X}$ une variable aléatoire exponentielle de paramètre $\lambda > 0$. On considère la variable aléatoire $\mathcal{Y} = e^{\mathcal{X}}$.

a) Déterminer la densité de probabilité de $\mathcal{Y}$.

b) Calculer l'espérance mathématique de $\mathcal{Y}$ en précisant les valeurs de λ pour lesquelles elle est définie.

c) Calculer la variance de $\mathcal{Y}$ en précisant les valeurs de λ pour lesquelles elle est définie.

5.4. EXAMEN MAI 2014

Un ingénieur travaillant pour une nouvelle centrale électrique liégoise a la responsabilité d'estimer la demande d'électricité $\mathcal{Y}$ à laquelle la centrale doit s'attendre dans le courant de la journée. Il se base sur deux variables explicatives : la température extérieure $\mathcal{X}_1$ et le vent $\mathcal{X}_2$. On sait en effet que lorsque la température extérieure diminue, la demande en électricité augmente. On sait également que quand le vent augmente, il a plus de concurrence car les producteurs travaillant en énergies renouvelables produisent plus et augmentent ainsi l'offre sur la marché, réduisant la demande de la centrale électrique. $\mathcal{X}_1$ et $\mathcal{X}_2$ ne sont pas indépendantes : la probabilité d'avoir un vent faible est en effet plus élevée lorsqu'il fait froid. La table 2 donne la probabilité d'avoir un vent faible/fort en fonction du type de temps. La table 4 donne la demande en électricité en fonction du temps et du vent. On suppose qu'en Belgique la probabilité d'avoir un temps froid est de 35%, celle d'avoir un temps agréable est de 40% et celle d'avoir un temps chaud est de 25%.

Table 2 – Probabilité conditionnelle de vent faible/fort en fonction du temps.

	Temps froid	Temps agréable	Temps chaud
Vent faible	70%	25%	10%
Vent fort	30%	75%	90%

Table 3 – Demande conditionnelle d'électricité en fonction du temps et du vent.

	Temps froid	Temps agréable	Temps chaud
Vent faible	Espérance : 300 GWh Ecart-type : 20 GWh	Espérance : 260 GWh Ecart-type : 30 GWh	Espérance : 200 GWh Ecart-type : 50 GWh
Vent fort	Espérance : 250 GWh Ecart-type : 30 GWh	Espérance : 200 GWh Ecart-type : 40 GWh	Espérance : 140 GWh Ecart-type : 60 GWh

a) En supposant que toutes les distributions de la table 4 sont gaussiennes, déterminez la probabilité, par temps froid, d'avoir une demande supérieure à 290 GWh. (6 points)

b) Calculez $E\{\mathcal{Y}\}$. (4 points)

c) Calculez $V\{E\{\mathcal{Y}|\mathcal{X}_1, \mathcal{X}_2\}\}$ et $E\{V\{\mathcal{Y}|\mathcal{X}_1, \mathcal{X}_2\}\}$. Quel pourcentage de $\mathcal{Y}$, les variables $\mathcal{X}_1$ et $\mathcal{X}_2$ expliquent-elles à elles-deux ? (4 points)

d) Calculez $V\{E\{\mathcal{Y}|\mathcal{X}_1\}\}$ et $E\{V\{\mathcal{Y}|\mathcal{X}_1\}\}$. Quel pourcentage de $\mathcal{Y}$ la variable $\mathcal{X}_1$ explique-t-elle ? (4 points)

e) Les pourcentages trouvés aux sous-questions (c) et (d) permettent-ils de déduire quel pourcentage de $\mathcal{Y}$ la variable $\mathcal{X}_2$ explique ? Justifiez ! (2 points)

5.5. EXAMEN JUIN 2015

Le système élo est un système d'évaluation du niveau relatif d'un joueur d'échecs, basé sur la loi de Gauss. Ce système attribue aux joueurs, en fonction de leurs performances passées, un nombre de points (la cote élo) tel que deux joueurs supposés de la même force se voient attribuer le même nombre de points. Plus le joueur a de points élos, plus il est estimé être performant. François et Romain sont deux joueurs d'échecs belges, dont les cotes élo sont respectivement de 2103 et 2051. En Belgique, 3051 joueurs d'échecs actifs ont actuellement une cote élo. Soit $\mathcal{X}$ la variable aléatoire rendant la cote élo d'un joueur pris au hasard. Cette variable suit une loi normale, dont la moyenne est de 1641 points et l'écart-type de 310 points.

a) Calculez le pourcentage de joueurs qui sont plus forts que François. (2 points)

b) Déterminez le nombre de joueurs qui sont moins forts que François, mais plus forts que Romain. (2 points)

La cote élo représente le niveau d'un joueur d'échecs en moyenne sur une longue période. Bien entendu, lors d'une partie en particulier, le joueur peut avoir un niveau élo effectif légèrement différent de sa cote élo officielle. Par exemple, s'il est fort fatigué, il jouera moins bien que d'habitude et vaudra donc moins d'élo lors de la partie. Inversement, s'il a les blancs, il aura plus de chance de gagner, car avoir les blancs est un avantage réel aux échecs. On essaie d'estimer le niveau élo effectif $\mathcal{Y}$ de Romain lors d'une partie, en fonction de deux paramètres : la forme du jour de Romain $\mathcal{X}_1$ et la couleur des pièces attribuées à Romain $\mathcal{X}_2$. On considère que Romain a 30% de chance de ne pas être en forme, 50% d'être dans un état de forme moyen et 20% de chance d'être en super forme. Aux échecs, un joueur a en toute généralité 50% de chance d'avoir les blancs et 50% de chance d'avoir les noirs. La table 4 donne le niveau effectif de Romain en fonction des deux paramètres.

TABLE 4 – Niveau effectif de Romain en fonction des deux paramètres étudiés

	Pas en forme	Etat de forme moyen	En super forme
Romain a les blancs	Espérance : 2026 élo Ecart-type : 100 élo	Espérance : 2076 élo Ecart-type : 50 élo	Espérance : 2151 élo Ecart-type : 25 élo
Romain a les noirs	Espérance : 1976 élo Ecart-type : 120 élo	Espérance : 2026 élo Ecart-type : 60 élo	Espérance : 2101 élo Ecart-type : 30 élo

c) Calculez $E\{\mathcal{Y}\}$. (2 points)

d) Calculez $V\{E\{\mathcal{Y}|\mathcal{X}_1, \mathcal{X}_2\}\}$ et $E\{V\{\mathcal{Y}|\mathcal{X}_1, \mathcal{X}_2\}\}$. Quel pourcentage de $\mathcal{Y}$, les variables $\mathcal{X}_1$ et $\mathcal{X}_2$ expliquent-elles à elles-deux ? (3 points)

e) Calculez $V\{E\{\mathcal{Y}|\mathcal{X}_1\}\}$ et $E\{V\{\mathcal{Y}|\mathcal{X}_1\}\}$. Quel pourcentage de $\mathcal{Y}$ la variable $\mathcal{X}_1$ explique-t-elle ? (3 points)

f) Les pourcentages trouvés aux sous-questions (d) et (e) permettent-ils de déduire quel pourcentage de $\mathcal{Y}$ la variable $\mathcal{X}_2$ explique ? Justifiez ! (2 points)

g) Y a-t-il indépendance entre $\mathcal{Y}|\mathcal{X}_1$ et $\mathcal{Y}|\mathcal{X}_2$? (2 points)

Romain propose à François de faire une partie d'échecs appellée un blitz. Lors de ce type de parties, les joueurs doivent jouer très rapidement, chacun à leur tour. Le temps, exprimé en secondes, qu'un joueur prend pour effectuer un coup, peut se modéliser à l'aide d'une variable aléatoire continue. Dans le cas de Romain, ce temps suit une variable aléatoire exponentielle de paramètre $\lambda = \frac{1}{5}$. Dans le cas de François, il suit une variable aléatoire exponentielle de paramètre $\lambda = \frac{1}{6}$. Les temps de réflexion de Romain et François sont indépendants.

h) François et Romain viennent de jouer 20 coups chacuns. Quelle est la probabilité que la partie ait commencé il y a moins de 2 minutes ? (4 points)

Solutions des exercices suggérés

5.3. a) $f_{\mathcal{Y}}(y) = \begin{cases} \frac{\lambda}{y^{\lambda+1}} & \text{si} \quad y \geq 1 \\ 0 & \text{sinon} \end{cases}$

b) $E\{\mathcal{Y}\} = \frac{\lambda}{\lambda-1}$ défini pour $\lambda > 1$

c) $V\{\mathcal{Y}\} = \frac{\lambda}{(\lambda-1)^2(\lambda-2)}$ défini pour $\lambda > 2$

5.4. a) 51,15%

b) $E\{\mathcal{Y}\} = 222{,}25$ GWh

c) $V\{E\{\mathcal{Y}|\mathcal{X}_1, \mathcal{X}_2\}\} = 3388$ (GWh)^2 $\quad E\{V\{\mathcal{Y}|\mathcal{X}_1, \mathcal{X}_2\}\} = 1635$(GWh)^2

A elles deux les variables $\mathcal{X}_1$ et $\mathcal{X}_2$ expliquent $\pm$ deux tiers de la variable $\mathcal{Y}$.

d) $V\{E\{\mathcal{Y}|\mathcal{X}_1\}\} = 2852,7$ (GWh)^2 $\quad E\{V\{\mathcal{Y}|\mathcal{X}_1\}\} = 2170,3$ (GWh)^2

La variable $\mathcal{X}_1$ explique $\pm$ 57% de la variable $\mathcal{Y}$.

e) Le pourcentage de la sous-question (c) n'est pas égal à la somme des pourcentages d'explications des variables $\mathcal{X}_1$ et $\mathcal{X}_2$. Ce ne serait même pas le cas si elles étaient indépendantes. Il n'est pas possible de déduire celle de $\mathcal{X}_2$ sans faire explicitement les calculs.

5.5. a) 6,81% b) 2,5% c) 2051 élo

d) 2500 (élo)^2 et 5337,5 (élo)^2 expliquent $\pm$68%

e) 1875 (élo)^2 et 5962,5 (élo)^2 explique $\pm 23,9\%$

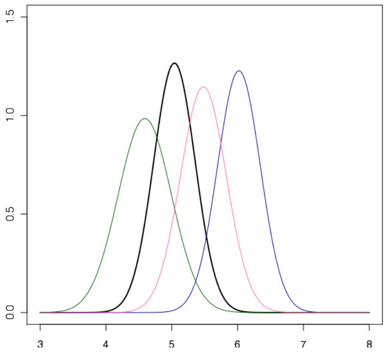
f) Non, le pourcentage calculé au point (d) n'est pas égal à la somme des pourcentages d'explications des variables $\mathcal{X}_1$ et $\mathcal{X}_2$. Il n'est pas possible d'obtenir le pourcentage demandé sans faire explicitement les calculs.

g) $\mathcal{Y}$ n'est pas expliqué à 100% par $\mathcal{X}_1$ et $\mathcal{X}_2$ et dépend donc d'une ou plusieurs autre(s) variable(s) non observée(s). $\mathcal{Y}|\mathcal{X}_1$ et $\mathcal{Y}|\mathcal{X}_2$ dépendent donc toutes deux de cette/ces autre(s) variable(s) et ne sont pas indépendants.

h) 29,869%

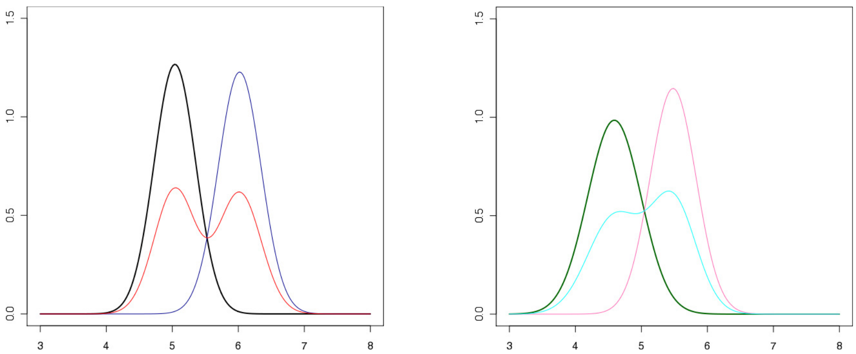
Annexe

Figure 1 - gaussiennes de l'exercice 5.2 (sous-question e)



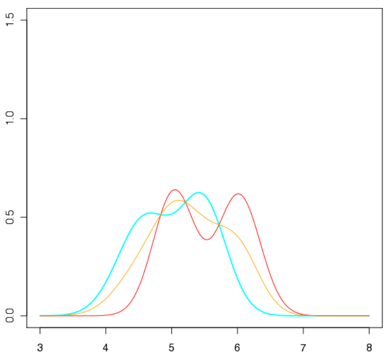
La noire correspond à C_1T_1, la bleue foncé à C_1T_2, la verte à C_2T_1, la rose à C_2T_2.

Figure 2 - distributions pour les carburants C_1 et C_2



Ces deux figures montrent d'une part ce qui se passe lorsqu'on moyennise la noire et la bleue foncée pour obtenir la rouge (distribution du cas C_1) et d'autre part lorsqu'on moyennise la rose et la verte pour obtenir la bleue clair (cas C_2).

Figure 3 - distribution globale du coût du trajet



La combinaison des courbes rouges et bleue claire donne la distribution de $\mathcal{Y}$ (orange).

Figure 4 - arbre de décision