Introduction à la Modélisation Statistique

Contrôle Continu – 3 Novembre 2015

Durée : 1h30

Documents interdits – calculatrices UPPA autorisées

Chaque réponse devra être justifiée et rédigée de manière rigoureuse. La qualité de la rédaction, la clarté et la précision des raisonnements interviendront pour une part importante dans l'appréciation des copies.

Exercice 1 (5pts). Soit X une variable aléatoire discrète définie sur un espacé probabilisé $(\Omega, \mathcal{A}, P)$.

a) Définir l'espérance $\mathrm{E}[X]$ et la variance $\mathrm{Var}[X]$ de la variable aléatoire X.

b) Soient $a, b \in \mathbb{R}$. Supposons que l'univers image $X(\Omega)$ est fini, c-à-d. $X(\Omega) = \{x_1, \ldots, x_n\}$.

   i) Démontrer que $\mathrm{E}[aX + b] = a\,\mathrm{E}[X] + b$.

   ii) En déduire que $\mathrm{Var}[aX + b] = a^2\,\mathrm{Var}[X]$.

Exercice 2 (10pts). Pour embaucher ses cadres une entreprise fait appel à un cabinet de recrutement. La procédure retenue est la suivante. Le cabinet effectue une première sélection de candidats sur dossier. 40% des dossiers reçus sont validés et transmis à l'entreprise. Les candidats ainsi sélectionnés passent un premier entretien à l'issue duquel 70% d'entre eux sont retenus. Ces derniers sont convoqués à un ultime entretien avec le directeur des ressources humaines qui recrutera 25% des candidats rencontrés.

a) On choisit au hasard le dossier d'un candidat et on considère les événements :
   - $D = \{$“ Le candidat est retenu sur dossier ”$\}$,
   - $E_1 = \{$“ Le candidat est retenu à l'issue du premier entretien ”$\}$,
   - $E_2 = \{$“ Le candidat est recruté à l'issue du ultime entretien ”$\}$.

   i) Construire un arbre pondéré à l'aide de D, E_1 et E_2 qui exprime la procédure de recrutement.

   ii) Calculer la probabilité de l'événement E_1.

   iii) On note $F = \{$“ Le candidat n'est pas récruté ”$\}$. Calculer $P(F)$.

b) Cinq amis postulent à un emploi de cadre dans cette entreprise. Les études de leur dossier sont faites indépendamment les unes des autres. On admet que la probabilité que chacun d'eux soit recruté est égale à $0,07$. On désigne par X la variable aléatoire donnant le nombre de personnes recrutées parmi ces cinq candidats.

   i) Quel est la probabilité que aucun soit recruté ? et que les cinq soient recrutés ?

   ii) Quel est la probabilité que au moins deux soient recrutés ?

   iii) Les coûts de préparer les dossiers (impressions de cv/lettre de motivation, enveloppe, timbres,...) sont de 3€ par chaque candidature. Par contre, l'entreprise rembourse une quantité de 12€ aux candidats qui sont finalement recrutés. Quels sont les dépenses attendues par le groupe de cinq amis ?

Exercice 3 (6pts). En 1955, Welcher a proposé de mesurer le QI (Quotient Intellectuel) des adultes grâce à deux échelles permettant de mesurer les compétences verbales et les compétences non verbales. Suite à la réalisation d'un étude statistique, on détermine que le QI des jeunes de la Région Aquitaine est exprimée par une variable aléatoire X en suivant une loi normale de moyenne 100 et écart-type 15.

a) Quel est le pourcentage de jeunes dont le QI est inférieur à 80 ? et un QI entre 90 et 120 ?

b) On considère qu'une personne est *surdoué* par rapport à l'échelle de Welcher si son QI est supérieur à 130. Le Ministère de l'Éducation nationale veut ouvrir un lycée d'étudiants surdoués à Pau. S'il y a 150 000 lycéens en Aquitaine, quel doit être au minimum l'offre de places dans un tel lycée ?

c) Déterminer l'intervalle de la forme $I = [100 - a, 100 + a]$, pour un certain $a > 0$, tel que le 90% des jeunes aient un QI contenu dans I.

Introduction à la Modélisation Statistique

Solutions détaillées

Solution 1. Soit X une v.a.d et $X(\Omega)$ l'univers image de X.

a) On appelle :

– *espérance* de X au nombre réel donné par la somme pondéré

$$\mathrm{E}[X] = \sum_{x \in X(\Omega)} x \cdot P(X = x)$$

Plus généralement, si ϕ est une fonction réelle définie sur $X(\Omega)$:

$$\mathrm{E}[\phi(X)] = \sum_{x \in X(\Omega)} \phi(x) \cdot P(X = x)$$

– *variance* de X au nombre réel définie par

$$\mathrm{Var}[X] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \sum_{x \in X(\Omega)} (x - \mathrm{E}[X])^2 \cdot P(X = x)$$

On assume que l'univers image est fini de la forme $X(\Omega) = \{x_1, \ldots, x_n\}$. Soient $a, b \in \mathbb{R}$:

a) Par définition d'espérance :

$$\begin{aligned}
\mathrm{E}[aX + b] &= \sum_{x \in X(\Omega)} (ax + b) \cdot P(X = x) = \sum_{i=1}^{n} (ax_i + b) \cdot P(X = x_i) \\
&= a \underbrace{\sum_{i=1}^{n} x \cdot P(X = x_i)}_{= \mathrm{E}[X] \text{ par défn.}} + b \underbrace{\sum_{i=1}^{n} P(X = x_i)}_{=1} \\
&= a\,\mathrm{E}[X] + b
\end{aligned}$$

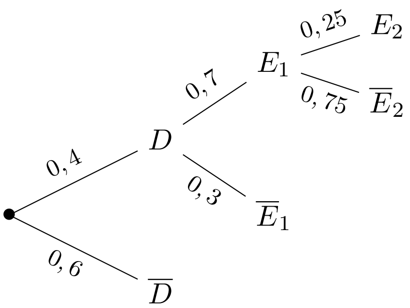
b) Par définition de la variance, on a :

$$\mathrm{Var}[aX + b] = \mathrm{E}\left[((aX + b) - \mathrm{E}[aX + b])^2\right]$$

En utilisant la linéarité sur l'espérance qu'on vient de prouver dans (a) :

$$\begin{aligned}
\mathrm{Var}[aX + b] &= \mathrm{E}\left[((aX + b) - \mathrm{E}[aX + b])^2\right] = \mathrm{E}\left[(aX + b - a\,\mathrm{E}[X] + b)^2\right] \\
&= \mathrm{E}\left[(aX - a\,\mathrm{E}[X])^2\right] = \mathrm{E}\left[a^2\,(X - \mathrm{E}[X])^2\right] \\
&= a^2\,\mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = a^2\,\mathrm{Var}[X]
\end{aligned}$$

Solution 2. a) i) À partir des donnés dans l'énoncé :



ii) On peut décomposer E_1 comme réunion disjointe de $(E_1 \cap D)$ et $(E_1 \cap \overline{D})$, alors :

$$P(E) = P(E_1 \cap D) + P(E_1 \cap \overline{D})$$

Il est clair que E_1 et $\overline{D}$ sont des événements incompatibles. En utilisant des propriétés du cours sur le calcul de l'intersection de deux événements à partir de la probabilité conditionnel, on a :

$$P(E_1) = P(E_1 \cap D) = P(D) \cdot P(E_1 \mid D)$$

À partir de l'arbre pondéré, on peut calculer :

$$\boxed{P(E_1) = P(D) \cdot P(E_1 \mid D) = 0,4 \cdot 0,7 = 0,28}$$

iii) On peut calculer $P(F)$ de deux façons, ou bien en décrivant F comme la réunion disjointe de $\overline{D}$, $D \cap \overline{E}_1$ et $D \cap \overline{E}_2$ et en calculant par des propriétés de probabilité conditionnelle via l'arbre pondéré :

$$\begin{aligned} P(F) &= P(\overline{D} \cup (D \cap \overline{E}_1) \cup (D \cap \overline{E}_2)) = P(\overline{D}) + P(D \cap \overline{E}_1) + P(D \cap \overline{E}_2) \\ &= P(\overline{D}) + P(D) \cdot P(\overline{E}_1 \mid D) + P(D) \cdot P(E_1 \mid D) \cdot P(\overline{E}_2 \mid E_1 \cap D) \\ &= 0,6 + 0,4 \cdot 0,3 + 0,4 \cdot 0,7 \cdot 0,75 = 0,93 \end{aligned}$$

Ou bien, on peut remarquer que $F = \{$“ Le candidat n'est pas récruté ”$\} = \overline{E}_2$, dans ce cas :

$$P(F) = 1 - P(E_2)$$

Si on calcule la probabilité de $E_2 = D \cap E_1 \cap E_2$:

$$\begin{aligned} P(E_2) &= P(D \cap E_1 \cap E_2) = P(D)P(E_1 \cap E_2 \mid D) = P(D)P(E_1 \mid D)P(E_2 \mid D \cap E_1) \\ &= 0,4 \cdot 0,7 \cdot 0,25 = 0,07 \end{aligned}$$

D'où :

$$\boxed{P(F) = 1 - P(E_2) = 0,93}$$

b) Chaque candidat que postule parmi le cinq correspond à une épreuve de Bernoulli de succès $S = \{$“ être recruté$\}$ et probabilité $0,07$. On a cinq épreuves de Bernoulli indépendantes, et la variable aléatoire X compte le nombre de candidats recrutés parmi le cinq. Par définition, on voit bien que X suit une loi binomiale de paramètres $n = 5$ et $p = 0,07$, c-à-d. $X \sim \mathcal{B}(5; 0,07)$. En conséquence, X prends valeurs dans $\{0, 1, 2, 3, 4, 5\}$ avec probabilité :

$$P(X = k) = C_5^k (0,07)^k (0,93)^{5-k}, \quad k \in \{0, 1, 2, 3, 4, 5\}$$

On peut facilement calculer :

i) $P(X = 0) = C_5^0 (0,07)^0 (0,93)^5 = (0,93)^5 = 0,6957$.
$P(X = 5) = C_5^5 (0,07)^5 (0,93)^0 = (0,07)^5 = 0,000001$.

ii) $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 1 - 0,6957 - C_5^1 (0,07)^1 (0,93)^4 = 0,3043 - 5(0,07)^1 (0,93)^4 = 0,0424$.

iii) On peut exprimer les dépenses du groupe par la variable aléatoire $Y = 15 - 12X$ où $15 = 3 \times 5$ sont les dépenses fixes du groupe de cinq amis et $12X$ sont les remboursements pour ces qui sont finalement recrutés. On va calculer $\mathrm{E}[Y]$ comme quantité attendue de dépenses. Par linéarité de l'espérance (Ex. 1-(a)), on sait que :

$$\mathrm{E}[Y] = \mathrm{E}[15 - 12X] = 15 - 12\,\mathrm{E}[X]$$

Il suffit de calculer $\mathrm{E}[X]$. On sait que X suit une loi binomiale de paramètres $n = 5$ et $p = 0,07$, et on a prouvé dans le cours que

$$\mathrm{E}[X] = np = 5 \cdot 0,07 = 0,35$$

D'où :

$$\boxed{\mathrm{E}[Y] = 15 - 12 \cdot 0,35 = 10,8}$$

Les dépenses attendues du groupe sont de 10,8€.

Solution 3. On a la v.a. X ="QI des jeunes de la Région Aquitaine" suit une loi normale de moyenne 100 et écart type 15, c-à-d. $X \sim \mathcal{N}(100; 15^2)$. On a vu dans le cours que en renormalisant la v.a. normale, $Z = \frac{X-100}{15}$ suit une loi binomiale centrée réduite $\mathcal{N}(0,1)$. On dénote par $\Phi(z) = P(Z < z)$ la fonction de distribution de la loi normale centrée réduite. On utilisera la table de valeurs approximatives de $\Phi(z)$.

a) – $P(X < 80) = P\left(\frac{X-100}{15} < \frac{80-100}{15}\right) = P(Z < -1,33) = \Phi(-1,33) = 1 - \Phi(1,33) \simeq 1 - 0,9082 = 0,0918$. Alors, vers le 9,1% des jeunes ont une QI inférieur à 80 en Aquitaine.

– $P(90 < X < 120) = P\left(\frac{90-100}{15} < \frac{X-100}{15} < \frac{120-100}{15}\right) = P(-0,66 < Z < 1,33) = \Phi(1,33) - \Phi(-0,66) = \Phi(1,33) - 1 + \Phi(0,66) \simeq 0,9082 - 1 + 0,7454 = 0,6536$. Alors, vers le 65,3% des jeunes ont une QI inférieur à 120 et supérieur à 90 en Aquitaine.

b) On calcule d'abord le pourcentage de jeunes surdoués en Aquitaine :

$$P(X > 130) = P\left(\frac{X-100}{15} > \frac{130-100}{15}\right) = P(Z > 2) = 1 - \Phi(2) \simeq 1 - 0,9772 = 0,0228.$$

Alors, il faut prévoir au moins $0,0228 \cdot 150000 = 3420$ places au nouveau lycée.

c) On doit trouver $a > 0$ tel que

$$P(X \in [100 - a, 100 + a]) = 0,9$$

En utilisant la renormalisation, on peut trouver une condition équivalente :

$$\begin{aligned} 0,9 = P(100 - a < X < 100 + a) &= P\left(-\frac{a}{15} < \frac{X-100}{15} < \frac{a}{15}\right) \\ &= P\left(-\frac{a}{15} < Z < \frac{a}{15}\right) = \Phi\left(\frac{a}{15}\right) - \Phi\left(-\frac{a}{15}\right) \\ &= \Phi\left(\frac{a}{15}\right) - 1 + \Phi\left(\frac{a}{15}\right) \\ &= 2\Phi\left(\frac{a}{15}\right) - 1 \end{aligned}$$

D',où $a > 0$ doit vérifier :

$$\Phi\left(\frac{a}{15}\right) = 1,9/2 = 0,95$$

Si on cherche dans la table de $\Phi(z)$, on trouve que :

$$\Phi(1,645) \simeq 0,95$$

Alors, approximativement :

$$\frac{a}{15} = 1,645 \Longleftrightarrow a = 24,675.$$

On conclut que environ 90% des des jeunes en Aquitaine ont un QI dans $[75,325; 124,675]$.