Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées

DEUG MASS 2 **Année 2000-2001**

Probabilités et Statistique, Corrigé Code de la Route

Ex 1. *Code de la Route I.*

Pour l'examen du Code de la Route, les candidats doivent remplir un questionnaire de 40 questions en choisissant pour chacune d'elles l'une des 4 réponses proposées, dont une seule est exacte. Un candidat totalement ignorant décide de tenter sa chance en cochant complètement au hasard une réponse pour chaque question.

1) Le nombre S de bonnes réponses du candidat est ici le nombre de succès dans une suite de 40 épreuves répétées indépendantes, avec pour chacune probabilité de succès 1/4. La variable aléatoire S suit donc la loi binomiale de paramètres 40 et 1/4.

2) Pour calculer $\mathbf{P}(S \geq 36)$, on utilise la décomposition

$$\mathbf{P}(S \geq 36) = \sum_{k=36}^{40} \mathbf{P}(S = k).$$

$$\begin{aligned}
\mathbf{P}(S \geq 36) &= C_{40}^{36}\left(\frac{1}{4}\right)^{36}\left(\frac{3}{4}\right)^{4} + C_{40}^{37}\left(\frac{1}{4}\right)^{37}\left(\frac{3}{4}\right)^{3} + C_{40}^{38}\left(\frac{1}{4}\right)^{38}\left(\frac{3}{4}\right)^{2} + C_{40}^{39}\left(\frac{1}{4}\right)^{39}\left(\frac{3}{4}\right)^{1} \\
&\quad + C_{40}^{40}\left(\frac{1}{4}\right)^{40}\left(\frac{3}{4}\right)^{0} \\
&= \frac{1}{4^{40}}\Big(C_{40}^{36} \times 3^4 + C_{40}^{37} \times 3^3 + C_{40}^{38} \times 3^2 + C_{40}^{39} \times 3 + C_{40}^{40} \times 1\Big) \\
&= \frac{1}{4^{40}}\Big(91390 \times 81 + 9880 \times 27 + 780 \times 9 + 40 \times 1 + 1 \times 1\Big) \\
&\simeq 6,35 \times 10^{-18}.
\end{aligned}$$

Cette probabilité est infime. Elle est du même ordre de grandeur que celle de trouver 10 milliards de fois *consécutives* (sans tricher !) les 6 bons numéros au Loto en jouant une grille à 6 numéros à chaque tirage. Pour cela, l'heureux candidat et ses descendants devraient jouer pendant 100 millions d'années en gagnant à chaque fois ! On ne prend donc aucun risque en pratique en considérant qu'un candidat ayant obtenu au moins 36 bonnes réponses n'a certainement pas répondu au hasard à *toutes* les questions.

Ex 2. *Code de la Route II.*

Dans un modèle plus réaliste, le candidat répond à coup sûr lorsqu'il connaît la réponse à la question et s'il l'ignore, choisit au hasard entre les 4 réponses proposées. On suppose que toutes les questions sont indépendantes et que pour chacune de ces questions, la probabilité que le candidat connaisse la vraie réponse est p. Ce paramètre p mesure donc le vrai niveau du candidat.

1) Soit A_i l'évènement « le candidat *connaît* la réponse à la $i^{\text{ième}}$ question ». Notons $\mathbf{1}_{A_i}$ la variable aléatoire indicatrice de cet évènement :

$$\mathbf{1}_{A_i} = \begin{cases} 1 & \text{si le candidat connaît la réponse à la } i^{\text{ième}} \text{ question}\,; \\ 0 & \text{s'il l'ignore.} \end{cases}$$

Le nombre S de questions à réponse connue du candidat est donc

$$S = \sum_{i=1}^{40} \mathbf{1}_{A_i}.$$

L'indépendance des questions entraîne celle des évènements A_i. Comme ils ont chacun la même probabilité p, S suit la loi binomiale :

$$S \sim \mathrm{Bin}(40; p).$$

2) On note B_i l'évènement « le candidat donne la bonne réponse à la $i^{\text{ième}}$ question ». On a clairement

$$\mathbf{P}(B_i \mid A_i) = 1 \quad \text{et} \quad \mathbf{P}(B_i \mid A_i^c) = \frac{1}{4}.$$

On calcule alors $\mathbf{P}(B_i)$ en conditionnant par les deux cas possibles :

$$\mathbf{P}(B_i) = \mathbf{P}(B_i \mid A_i)\mathbf{P}(A_i) + \mathbf{P}(B_i \mid A_i^c)\mathbf{P}(A_i^c) = 1 \times p + \frac{1}{4}(1-p) = \frac{1+3p}{4}.$$

3) Introduisons à nouveau les variables aléatoires indicatrices :

$$\mathbf{1}_{B_i} = \begin{cases} 1 & \text{si le candidat donne la bonne réponse à la } i^{\text{ième}} \text{ question}\,; \\ 0 & \text{si le candidat donne une réponse fausse à la } i^{\text{ième}} \text{ question.} \end{cases}$$

Le nombre total U de bonnes réponses données par le candidat peut s'écrire :

$$U = \sum_{i=1}^{40} \mathbf{1}_{B_i}.$$

Les B_i sont indépendants et de même probabilité $(1+3p)/4$, donc U suit la loi binomiale :

$$U \sim \mathrm{Bin}\left(40; \frac{1+3p}{4}\right).$$

De même le nombre total V de réponses fausses données par le candidat est

$$V = \sum_{i=1}^{40} \mathbf{1}_{B_i^c}.$$

Les B_i^c étant indépendants et de même probabilité $1 - (1 + 3p)/4 = 3(1 - p)/4$,

$$V \sim \text{Bin}\Big(40; \frac{3 - 3p}{4}\Big).$$

4) Notons C_i l'évènement « le candidat répond au hasard à la $i^{\text{ème}}$ question *et* trouve la bonne réponse ». On a clairement pour tout $i \in \{1, \dots, 40\}$,

$$B_i = A_i \cup C_i \quad \text{avec} \quad A_i \cap C_i = \emptyset.$$

On en déduit $\mathbf{P}(B_i) = \mathbf{P}(A_i) + \mathbf{P}(C_i)$, d'où

$$\mathbf{P}(C_i) = \mathbf{P}(B_i) - \mathbf{P}(A_i) = \frac{1 + 3p}{4} - p = \frac{1 - p}{4}.$$

Le nombre T de bonnes réponses « chanceuses » est

$$T = \sum_{i=1}^{40} \mathbf{1}_{C_i}.$$

Les C_i étant indépendants et de même probabilité $(1 - p)/4$,

$$T \sim \text{Bin}\Big(40; \frac{1 - p}{4}\Big).$$

Ex 3. *Épreuves répétées à trois issues.*

On considère une suite de n épreuves répétées indépendantes avec pour chaque épreuve trois résultats possibles : A avec probabilité a, B avec probabilité b ou C avec probabilité c ($a + b + c = 1$). On notera respectivement A_i, B_i et C_i les événements obtention du résultat A (respectivement, B, C) à la i-ème épreuve.

1) L'événement E consistant en l'obtention dans cet ordre de deux A suivis d'un B et de deux C s'écrit :

$$E = A_1 \cap A_2 \cap B_3 \cap C_4 \cap C_5.$$

En raison de l'indépendance des épreuves, les cinq événements intervenant dans cette intersection sont *mutuellement indépendants*. D'où :

$$P(E) = P(A_1)P(A_2)P(B_3)P(C_4)P(C_5) = a^2bc^2.$$

L'événement F obtention (sans condition d'ordre) de deux A, un B et deux C est la réunion disjointe de tous les sous-événements du type E construits en imposant parmi les 5 épreuves les rangs des deux A et celui du B parmi les trois rangs restants (les deux C occupant alors les deux derniers rangs non utilisés). Chacun de ces sous-événements a même probabilité que E et il y a en tout $C_5^2 \times C_3^1$ dispositions possibles. Ainsi

$$P(F) = C_5^2 C_3^1 a^2bc^2 = \frac{5!}{2!\,3!}\frac{3!}{1!\,2!}a^2bc^2 = \frac{5!}{2!\,1!\,2!}a^2bc^2 = 30a^2bc^2.$$

2) Généralisation : pour $i+j+k=n$, notons $E_{i,j,k}$ l'évènement obtention dans cet ordre de i résultats A, j résultats B et k résultats C :

$$E_{i,j,k} = \Big(\bigcap_{\ell=1}^{i} A_\ell\Big) \cap \Big(\bigcap_{\ell=i+1}^{i+j} B_\ell\Big) \cap \Big(\bigcap_{\ell=i+j+1}^{n} C_\ell\Big).$$

Par indépendance des épreuves,

$$\mathbf{P}(E_{i,j,k}) = a^i b^j c^k.$$

Dans ces écritures, nous avons fait la convention habituelle selon laquelle une intersection indexée par une condition impossible à réaliser (par exemple $i+1 \le \ell \le i+j$ lorsque $j=0$) vaut Ω. Le calcul ci-dessus inclut donc les cas particuliers d'indices i, j ou k nuls. Si l'on supprime la condition d'ordre, on obtient l'évènement $F_{i,j,k}$, réunion disjointe d'événements de même probabilité que $E_{i,j,k}$. Leur dénombrement s'effectue comme ci-dessus, il y en a $C_n^i C_{n-i}^j$: en effet, il y a C_n^i façons de placer les i A, puis C_{n-i}^j façons de placer les j B parmi les $n-i$ positions restant libres, après quoi il ne reste plus que $k=n-i-j$ positions pour les k C, donc une seule façon de les placer. La probabilité d'obtenir au cours des n épreuves (et sans condition d'ordre) i A, j B et k C ($i+j+k=n$) vaut donc :

$$\mathbf{P}(F_{i,j,k}) = C_n^i C_{n-j}^j a^i b^j c^k = \frac{n!}{i!\,(n-i)!}\frac{(n-i)!}{j!\,(n-i-j)!} a^i b^j c^k = \frac{n!}{i!\,j!\,k!} a^i b^j c^k.$$

Ex 4. *Code de la Route III.*

On reprend les notations et les hypothèses de l'exercice 3. Ce que peut réellement observer l'examinateur, c'est la valeur prise par U (ou ce qui revient au même par V). Les quantités intéressantes pour tirer des conclusions sur le niveau réel du candidat sont les $\mathbf{P}(S=i \mid U=m)$ pour $i \le m \le 40$. Par exemple si le candidat a obtenu 38 bonnes réponses, on aimerait évaluer $\mathbf{P}(S \ge 36 \mid U=38)$.

1) Notons $E=\{S=i, U=m\}$ et $F=\{S=i, T=m-i, V=40-m\}$. Nous allons vérifier que $E=F$.

Soit $\omega \in E$. Cela signifie que $S(\omega)=i$ et $U(\omega)=m$. Comme $T=U-S$ et $V=40-U$, on a $T(\omega)=m-i$ et $V(\omega)=40-m$. Donc ω vérifie les trois conditions d'appartenance à F. Donc tout $\omega \in E$ est aussi élément de F (autrement dit, $E \subset F$).

Réciproquement, soit $\omega \in F$. Alors $S(\omega)=i$, $T(\omega)=m-i$ et $V(\omega)=40-m$. Comme $U=T+S$, $U(\omega)=T(\omega)+S(\omega)=(m-i)+i=m$. Ainsi $S(\omega)=i$ et $U(\omega)=m$, donc ω est dans E. On a vérifié que tout ω de F est aussi dans E, autrement dit que $F \subset E$.

Les inclusions $E \subset F$ *et* $F \subset E$ montrent que $E=F$, d'où l'égalité des probabilités :

$$\mathbf{P}(S=i, U=m) = \mathbf{P}(S=i, T=m-i, V=40-m), \quad i \le m \le 40.$$

2) On peut alors utiliser cette égalité pour le calcul de $\mathbf{P}(S=i \mid U=m)$:

$$\mathbf{P}(S=i \mid U=m) = \frac{\mathbf{P}(S=i, U=m)}{\mathbf{P}(U=m)} = \frac{\mathbf{P}(S=i, T=m-i, V=40-m)}{\mathbf{P}(U=m)}.$$

Comme la loi de U nous est connue, il suffit de calculer $\mathbf{P}(S=i, T=m-i, V=40-m)$. Pour ce faire, on utilise le résultat de l'exercice 4 en considérant que l'on a une série de 40 épreuves répétées indépendantes avec pour chacune, trois résultats possibles :

A réponse connue du candidat avec probabilité $a = p$;
B bonne réponse chanceuse avec probabilité $b = (1-p)/4$;
C réponse fausse avec probabilité $c = (3-3p)/4$.

Dans ces conditions, la probabilité d'avoir i réponses connues, $m-i$ bonnes réponses chanceuses et $40-m$ réponses fausses est

$$\mathbf{P}(S=i, T=m-i, V=40-m) = \frac{40!}{i!(m-i)!(40-m)!} p^i \Big(\frac{1-p}{4}\Big)^{m-i} \Big(\frac{3-3p}{4}\Big)^{40-m}.$$

Comme U suit la loi $\mathrm{Bin}\Big(40; \frac{1+3p}{4}\Big)$,

$$\mathbf{P}(U=m) = \frac{40!}{m!(40-m)!} \Big(\frac{1+3p}{4}\Big)^m \Big(\frac{3-3p}{4}\Big)^{40-m}.$$

En reportant dans le calcul de $\mathbf{P}(S=i \mid U=m)$, il vient après simplifications

$$\mathbf{P}(S=i \mid U=m) = \frac{m!}{i!(m-i)!} p^i \Big(\frac{1-p}{4}\Big)^{m-i} \Big(\frac{4}{1+3p}\Big)^m = C_m^i \Big(\frac{4p}{1+3p}\Big)^i \Big(\frac{1-p}{1+3p}\Big)^{m-i}.$$

Rappelons que ce résultat est valide pour $0 \le i \le m \le 40$. On remarque que

$$0 \le \frac{4p}{1+3p} \le 1 \quad \text{et} \quad \frac{4p}{1+3p} + \frac{1-p}{1+3p} = 1.$$

On peut donc énoncer le résultat sous la forme suivante : *conditionnellement à l'évènement* $\{U=m\}$, S suit la loi binomiale de paramètres m et $4p/(1+3p)$.

3) En remarquant que si $i > m$, $\mathbf{P}(S=i \mid U=m) = 0$ (ceci revient à considérer que le candidat n'est pas distrait et ne peut donner moins de bonnes réponses qu'il n'en connaît !), on en déduit

$$\mathbf{P}(S \ge 36 \mid U=38) = \mathbf{P}(S=36 \mid U=38) + \mathbf{P}(S=37 \mid U=38) + \mathbf{P}(S=38 \mid U=38).$$

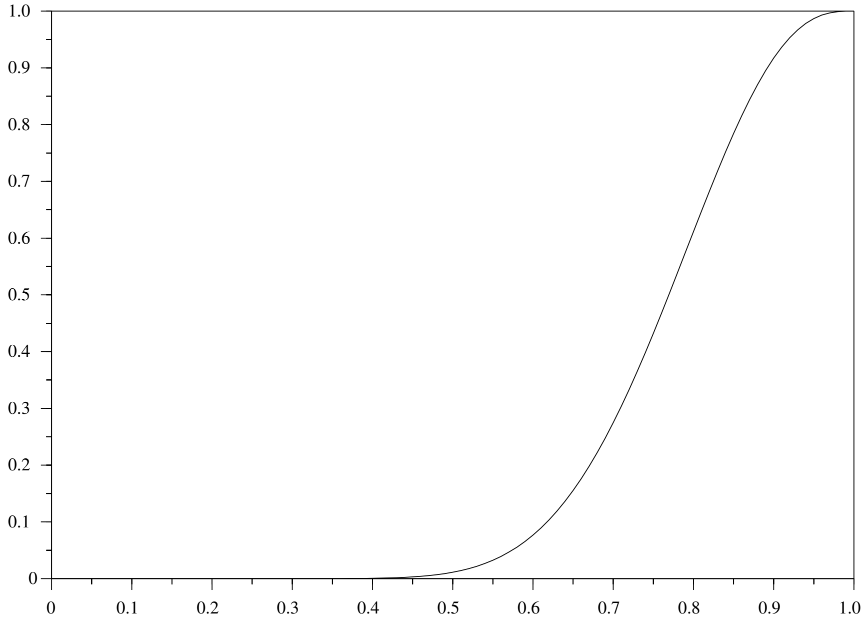
Posons pour alléger les écritures $r = 4p/(1+3p)$.

$$\begin{aligned} \mathbf{P}(S \ge 36 \mid U=38) &= C_{38}^{36} r^{36}(1-r)^2 + C_{38}^{37} r^{37}(1-r) + C_{38}^{38} r^{38} \\ &= r^{36}\Big(703(1-r)^2 + 38r(1-r) + r^2\Big) \\ &= r^{36}\Big(666r^2 - 1368r + 703\Big). \end{aligned}$$

En remplaçant r par sa valeur, on aboutit après calcul à

$$\mathbf{P}(S \ge 36 \mid U=38) = \frac{(4p)^{36}}{(1+3p)^{38}}(567p^2 - 1254p + 703) =: f(p).$$

Voici la représentation graphique de f :


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0