

## RAPPORT : NOMBRES DE CATALAN (CENTRALE PC 2021)

### Présentation du sujet

Le sujet de cette épreuve propose le dénombrement des chemins de Dyck, c'est-à-dire des trajectoires sur un axe gradué par  $\mathbb{Z}$ , de longueur donnée, partant et aboutissant à 0 tout en restant de signe constant. Ce dénombrement fait naturellement apparaître les nombres de Catalan (qu'on retrouve aussi dans le dénombrement des « bons parenthésages »), qu'on utilise pour deux calculs de déterminant présentant une symétrie particulière (déterminants de Hankel).

La partie I consiste en l'étude du problème de dénombrement des chemins de Dyck, passant par la fonction génératrice de la variable aléatoire  $T$  désignant le temps de premier retour à 0 pour une particule effectuant des sauts de  $\pm 1$ , indépendants et de même loi, sur un axe gradué par  $\mathbb{Z}$ . Ce calcul mène au calcul explicite des nombres de Catalan  $C_n$  et à leur étude asymptotique.

La partie II s'attelle à l'étude des systèmes orthogonaux sur  $\mathbb{R}_n[X]$ , c'est-à-dire des bases orthogonales de polynômes unitaires. Étant donné un produit scalaire défini sur  $\mathbb{R}_n[X]$ , on y montre l'unicité d'un tel système orthogonal et on y effectue le calcul du déterminant de la matrice  $(X^{i-1} | X^{j-1})_{1 \leq i, j \leq n}$ . Ce résultat se révélera décisif en fin de partie III.

La partie III propose enfin d'appliquer les résultats de la partie II au produit scalaire

$$(P, Q) \mapsto \int_0^1 P(4t)Q(4t) \frac{\sqrt{1-t}}{\sqrt{t}} dt$$

permettant ainsi de faire apparaître les nombres de Catalan sous la forme  $C_n = (X^n | 1)$ . Cela aboutit au calcul des déterminants de Hankel détaillés en fin de sujet.

### Analyse globale des résultats

Sur les 3381 copies corrigées, la moyenne constatée, en pourcentage du barème, est de 23,7%, pour un écart-type de 13,7%, ce qui permet de considérer le sujet comme de longueur raisonnable. Il a par ailleurs permis une bonne discrimination parmi les candidats. Comme nous le verrons plus loin, la sélection des meilleurs candidats s'est essentiellement faite sur le soin apporté aux réponses et sur la solidité des connaissances, bien plus que sur le volume traité ou l'originalité des idées. La meilleure copie a obtenu 81,5% des points du barème total.

Ce sujet se caractérise par une grande diversité dans les parties du programme qu'il concerne : des probabilités, du dénombrement et des séries entières en partie I, de l'algèbre bilinéaire en parties II et III, et des intégrales généralisées en partie III. Les candidats ont su profiter de cette diversité, chacune de ces parties ayant été abordée par plus de 90% des copies. Les questions relatives aux déterminants, en fin de sujet, ont été en revanche très peu abordées, car plus difficiles et demandant un effort de synthèse plus important.

La familiarisation avec la situation décrite en partie I (chemins de Dyck, nombres de Catalan) est plutôt réussie. La différence entre les copies se fait essentiellement sur les points suivants, indicatifs du niveau de soin et de discipline pratiqué par les candidats dans leurs raisonnements :

- la connaissance précise du cours ;
- le niveau de rigueur dans la connaissance et le maniement des notions ;
- l'efficacité la ténacité et la précision dans les calculs.

### Commentaires sur les réponses apportées et conseils aux futurs candidats

#### Remarques générales

Commençons par illustrer les points qui sont apparus comme éléments essentiels de différenciation entre les copies.

- La connaissance précise du cours. La toute première question du sujet est un bon indicateur de la disparité des candidats sur ce plan. Un tiers des copies ne parvient pas à reconnaître, même sans justification, une loi binomiale pour la variable aléatoire  $Y_n$ . À l'opposé, seulement un sixième des copies parvient à reconnaître cette loi tout en citant les arguments attendus pour l'établir.

De même, plus loin dans le sujet, la question Q9 est le théâtre de nombreuses confusions entre convergence normale et convergence absolue en tout point.

La question Q28 est aussi le théâtre de nombreuses approximations quant aux propriétés à vérifier pour un produit scalaire, pourtant proche de l'exemple classique  $(P, Q) \mapsto \int_a^b P(t)Q(t) dt$  sur  $\mathbb{R}[X]$  (avec  $a < b$ ).

- Le niveau de rigueur dans la connaissance et le maniement des notions. Dans des questions parfois élémentaires : dans les questions Q24 à Q26, largement abordées, la nature de la matrice  $Q_n$  est un facteur discriminant entre les copies. Environ la moitié des candidats considèrent, à tort, et sans doute troublés par l'écriture  $Q_n^T G_n Q_n$ , que la matrice  $Q_n$  est orthogonale alors que rien ne le laisse entendre dans le sujet.

La question Q27 met également en évidence une grande disparité entre les candidats : certains établissent la convergence absolue de l'intégrale avec organisation et rigueur, là où d'autres se précipitent vers l'étude d'impropriétés plus ou moins réelles (il n'y avait aucun problème en 1), sans parler de continuité par morceaux de la fonction intégrée, ou encore confondant la notion d'intégrabilité avec celle d'intégrale convergente.

- L'efficacité, la ténacité et la précision dans les calculs. Les questions Q17 (développement en série entière de  $x \mapsto \sqrt{1+x}$  sur  $] -1, 1[$ ), Q19 (détermination d'un équivalent de  $\binom{2n}{n}$  à partir de l'équivalent de Stirling) et Q31 (primitivation d'un produit de deux fonctions circulaires) auront permis de mettre en évidence de grandes différences entre les candidats, tant sur l'attention portée à la correction des calculs, que sur leur ténacité et leur efficacité à les faire aboutir.

Le jury a par ailleurs relevé un certain nombre de points généraux lors de la correction des copies et en tire les recommandations suivantes.

*Les études asymptotiques*, mettant en jeu équivalents (Q19) ou études d'intégrales impropres (Q27), posent de grandes difficultés aux candidats, qui déclarent des équivalents à des quantités potentiellement égales à la fonction nulle, ou invoquent des propriétés fausses, comme l'intégrabilité du produit de deux fonctions intégrables.

*Les variables utilisées sont loin d'être systématiquement déclarées*. Il n'est pas rare de voir apparaître des  $x, y, t, u, P, Q$  au milieu d'un raisonnement sans en avoir vu la déclaration au préalable, laissant au lecteur le soin de comprendre dans quel ensemble ces variables se trouvent, ou ce qu'elles désignent. De telles pratiques nuisent à la clarté du discours et rendent le raisonnement confus. Le jury attend davantage de rigueur de la part des candidats sur ce plan.

*Confusions entre réels et fonctions*. Les confusions entre  $f$  et  $f(t)$  sont malheureusement fréquentes et répandues. La continuité, la dérivabilité, l'intégrabilité s'appliquent à une fonction et non à son évaluation en un réel  $t$  (souvent non déclaré, cf. point précédent).

*Impératif d'intégrité*. Le jury rappelle qu'il est vain, dans une question formulée de manière fermée, d'arranger les calculs de manière malhonnête en vue de trouver le résultat attendu. Par exemple, en question Q18, on compte un nombre non négligeable de candidats qui, partant d'une expression fautive concernant  $a_n$ , parviennent tout de même à la bonne expression de  $C_n$  en effectuant des arrangements d'une intégrité douteuse au milieu des calculs. En plus de n'en tirer aucun avantage comptable, un candidat adoptant une telle démarche laisse une impression négative au correcteur.

Le jury recommande aux candidats une posture d'humilité, et notamment de bannir de leur vocabulaire des mots comme « clairement », « trivialement », « évidemment ». Ceux-ci n'apportent rien au contenu mathématique de la copie et ne peuvent jouer qu'en défaveur du candidat, surtout lorsqu'ils sont suivis d'erreurs manifestes ou lorsqu'ils servent à passer rapidement sur des points essentiels à la résolution de la question. Par exemple, en question Q28, il est peu pertinent de qualifier la bilinéarité ou la symétrie de l'application étudiée d'« évidentes » ou de « triviales », comme cela est écrit dans certaines copies.

Le jury rappelle également que les fautes de français, malheureusement nombreuses, nuisent à la copie et laissent au lecteur une impression négative qui peut se répercuter, consciemment ou non, sur la note finale (en plus de faire l'objet d'un malus). La validité d'un raisonnement passe aussi par la correction de la langue employée pour l'exprimer.

### Remarques par question

**Q1.** Les arguments justifiant la loi binomiale pour  $Y_n$  (car on attend justification, comme dans toute réponse, sauf mention explicite du contraire) sont très rarement tous énoncés : les  $X_k$  sont indépendants et de même loi et  $Y_n$  décompte le nombre de  $X_k$  égaux à 1. La connaissance de l'espérance et de la variance de cette loi n'est également pas générale.

**Q2.** Le jury aura beaucoup rencontré la notation «  $\overline{Y}_n$  », censée implicitement désigner le nombre de  $X_k$  égaux à  $-1$ , ou l'écriture «  $S_n = \sum_{k=0}^n [X_k = 1] - \sum_{k=0}^n [X_k = -1]$  ». Ni le complémentaire d'une variable aléatoire, ni la somme d'événements ne possèdent de sens mathématique.

**Q3.** Beaucoup de bonnes réponses dans cette question de compréhension immédiate des notations de l'énoncé.

**Q4.** La maximalité de  $r$  est rarement comprise et bien utilisée par les candidats, pour prouver que  $\beta$  est bien un chemin de Dyck.

**Q5.** On note des confusions entre « indépendance mutuelle » et « indépendance deux à deux » et des erreurs plus lourdes écrivant la probabilité d'une intersection d'événements mutuellement indépendants comme la somme des probabilités de ces événements. Le jury rappelle que l'indépendance mutuelle des événements est un argument capital dans les questions de ce type.

- Q6.** La parité de  $T$  est globalement bien justifiée. En revanche, on trouve peu d'explications solides quant à l'expression de la loi de  $T$ .
- Q7.** Le lien avec la question Q4 doit être convenablement établi, ainsi que les différents procédés utilisés (somme et produit), par exemple en évoquant une partition du décompte selon l'entier  $r$  décrit en Q4. Le jury relève peu de réponses complètement satisfaisantes sur cette question.
- Q8.** Beaucoup de candidats cherchent à étudier le terme général de la série, alors que l'expression ou le comportement asymptotique de  $C_n$  sont complètement inconnus à ce stade du sujet. C'est la propriété de  $\sigma$ -additivité de la mesure de probabilité  $P$  qui se révèle décisive, peu l'auront vu. Le jury note de nombreuses erreurs de majoration dans les propositions des candidats.
- Q9.** La notion de convergence normale d'une série de fonctions sur l'intervalle  $I$  est globalement mal connue des candidats, souvent confondue avec la convergence absolue pour tout  $t \in I$ . On rappelle que la convergence normale d'une série entière a lieu sur tout segment de l'intervalle ouvert de convergence : ici, la série entière ayant un rayon de convergence de  $1/4$ , la convergence normale est acquise sur tout segment de l'intervalle  $] -1/4, 1/4[$  et pas a priori sur le segment  $[-1/4, 1/4]$ .
- Q10.** Il s'agit d'un simple calcul, plutôt bien réussi par les candidats. Il est important de relever que  $p(1-p)t^2 \in I$  lorsque  $t \in [-1, 1]$ .
- Q11.** Question difficile et peu traitée, qui demandait du soin dans l'utilisation des résultats du cours. Il s'agit de s'appuyer sur le fait que  $T$  admet une espérance si et seulement si  $G_T$  est dérivable en 1 et établir soigneusement la dérivabilité de la fonction génératrice  $G_T$  en 1.
- Q12.** L'identification d'un produit de Cauchy est au cœur de cette question. Le jury rappelle que le produit de Cauchy n'est pas qu'une formule et qu'il a un cadre théorique de validité : la convergence absolue des deux séries numériques concernées. Cette justification est peu présente dans les copies.
- Q13.** Question globalement bien réussie par les candidats, par résolution d'une équation du second degré à discriminant positif. Le jury relève certaines démarches proposant de montrer que la forme proposée satisfait la même équation fonctionnelle que  $g$  : cela constitue une sérieuse erreur de logique, car rien n'assure l'unicité de la solution à une telle équation fonctionnelle (loin s'en faut).
- Q14.** Question subtile et rarement bien traitée par les candidats. Le jury regrette la très rare utilisation du théorème des valeurs intermédiaires pour justifier le caractère constant de la fonction  $\varepsilon$  sur l'intervalle  $I \setminus \{1/4\}$ . Il est également important de préciser le domaine sur lequel on établit une continuité : cette question demande un examen particulier du comportement de  $f, g$  et  $\varepsilon$  en  $\pm 1/4$ , il faut donc redoubler de précision sur ce plan.
- Q15.** Le cas  $p = 1/2$  est largement commenté, plus rarement avec les bons termes : il s'agit de rester sur un registre « presque sûr » plutôt que « sûr » ou « nécessaire ». La particule revient presque sûrement à 0 et non forcément à 0.
- Q16.** Question de synthèse plutôt difficile, peu traitée par les candidats.
- Q17.** La plupart des candidats connaît l'existence d'un développement en série entière pour la fonction  $x \mapsto \sqrt{1+x}$  (valorisé par le jury), mais peu savent en déterminer précisément les coefficients selon les règles de l'énoncé (utilisation d'un coefficient binomial). Le jury rappelle également qu'une fonction  $\mathcal{C}^\infty$  sur un voisinage de 0 n'est pas nécessairement développable en série entière au voisinage de 0 (argument rencontré dans les copies). La notation  $\binom{1/2}{n+1}$  est hors programme. Une expression comme « pour tout  $x \in ]-1, 1[$ ,  $f(x)$  est développable en série entière » est impropre. On lui préférera « la fonction  $f$  est développable en série entière sur  $]-1, 1[$  ».
- Q18.** Question peu traitée par les candidats. Les versions vues ont correctement invoqué le théorème d'unicité des coefficients d'un développement en série entière.
- Q19.** De nombreux candidats ne connaissent pas l'équivalent de Stirling. On voit beaucoup une version inversant le quotient  $\left(\frac{n}{e}\right)^n$  en  $\left(\frac{e}{n}\right)^n$  faisant de  $n!$  une quantité tendant vers 0.
- Q20.** Question très peu traitée par les candidats.
- Q21.** Une famille orthogonale n'est pas nécessairement libre, il faut aussi qu'elle soit constituée de vecteurs tous non nuls. Ce point est négligé dans de nombreuses copies.
- Q22.** Question globalement réussie, où l'essentiel est d'utiliser la linéarité à droite du produit scalaire sur une décomposition de  $P$  sur la base  $(V_0, \dots, V_{n-1})$ .
- Q23.** Question très rarement bien traitée.
- Q24.** Question globalement réussie. Il ne faut toutefois pas penser que les coefficients de la matrice  $Q$  sont de la forme  $(V_j | X^i)$ , car la famille  $(1, X, \dots, X^n)$  n'est pas nécessairement orthonormale pour le produit scalaire utilisé.
- Q25.** Trop de candidats évoquent un changement de base, pensant que la matrice  $Q_n$  est une matrice orthogonale, ce qui n'est pas nécessairement le cas (et  $\det(Q_n) = 1$  obtenue à la question Q24 n'est certainement pas un argument suffisant pour cela). Il s'agit plutôt de vérifier l'égalité en effectuant le produit matriciel  $Q_n^T G_n Q_n$ .

**Q26.** Question globalement bien traitée pour les copies qui ne tombent pas dans l'erreur selon laquelle  $G_n$  et  $G'_n$  sont des matrices semblables.

**Q27.** Cette question est globalement mal traitée dans les copies. On note des problèmes d'organisation dans l'étude de la nature de l'intégrale impropre : beaucoup de candidats se précipitent sur l'impropreté en 0 sans prendre le temps de parler de la continuité par morceaux de la fonction intégrée, établissent un équivalent de la fonction intégrée à  $\frac{P(0)Q(0)}{\sqrt{x}}$  sans prendre garde à l'éventuelle nullité de  $P(0)Q(0)$ , etc. Le qualificatif « intégrable » est souvent confondu, à tort, avec l'expression « d'intégrale convergente ». Par ailleurs, le produit de fonctions intégrables n'est pas nécessairement intégrable, erreur sérieuse rencontrée dans beaucoup de copies.

**Q28.** La connaissance des propriétés à vérifier pour un produit scalaire est un point de cours mal maîtrisé par de nombreux candidats. Le jury rappelle néanmoins que l'adjectif « linéaire » utilisé seul, est inadapté pour qualifier un potentiel produit scalaire : on lui préférera « linéaire à gauche / à droite » ou « bilinéaire ».

**Q29.** Le jury attire l'attention des candidats sur le soin à apporter à la rédaction de récurrences dites « doubles » (avec double initialisation et hérédité supposant la propriété vérifiées aux rangs  $n$  et  $n + 1$ ). On rencontre encore beaucoup de versions insuffisantes.

**Q30.** Question peu traitée.

**Q31.** Il est impératif de distinguer les cas  $m = n$  et  $m \neq n$ , en raison d'une division par  $m - n$  apparaissant au milieu du calcul.

**Q32.** Le jury rappelle aux candidats l'importance de bien justifier les changements de variable dans des intégrales impropres (caractère  $\mathcal{C}^1$  et strictement monotone de la fonction de changement, avec précision des domaines concernés par ces propriétés).

**Q33.** Question peu traitée par les candidats.

**Q34 à Q37.** Questions très rarement traitées par les candidats. Les réponses partielles basées sur des idées constructives ont été valorisées.

## Conclusion

Il est absolument primordial de se présenter à une épreuve de ce niveau avec une connaissance précise des éléments de cours et une capacité à les manier avec précision et rigueur. Il est également important d'apporter une attention particulière à ce qui semble être considéré par de nombreux candidats — à tort — comme des détails : déclaration des variables, utilisation pertinente des liens logiques (implications, équivalences) et des mots de liaison. Il importe également que le candidat vérifie la totalité des hypothèses nécessaires avant utilisation d'un résultat précédemment établi, cela est très loin d'être systématique parmi les copies. Le correcteur, à l'écrit (contrairement à l'examineur, à l'oral), ne peut interroger le candidat afin de lui demander d'étayer ses affirmations ou de les compléter ; il faut donc que tout soit exprimé sur la copie. Ce manque de rigueur explique que de nombreux candidats risquent de se retrouver déçus par leur note, ayant eu l'impression de traiter de nombreuses questions du sujet, alors que la plupart des réponses auront été incomplètes ou insuffisamment précises.

Le jury tient également à rappeler la plus-value importante qu'apportent une rédaction soignée et une copie bien présentée. Il en aura été tenu compte dans la notation. Les désagréments impliqués par un manquement à ces règles sont doubles :

- sur le fond, un certain manque de soin ou une rédaction précipitée fait perdre des points importants de la question ou certaines étapes cruciales d'un raisonnement ;
- sur la forme, l'impression laissée au correcteur par une copie négligée est forcément négative. Pour éviter tout désagrément, le jury recommande aux candidats de soigner leur écriture, de limiter les ratures, d'éviter de multiplier les inserts plus ou moins lisibles ou les renvois vers une autre page et d'écrire dans un français correct.

Enfin, il n'est pas nécessaire de se précipiter et de traiter un nombre impressionnant de questions pour obtenir un très bon total : il suffit de procéder avec soin, dans un esprit scientifique empreint de rigueur et de précision. Nous encourageons les futurs candidats à prendre ces bonnes habitudes dans leur préparation. Les bonnes et très bonnes copies auront, presque sans exception, été de cette sorte.