Probabilités et Statistique pour SIC:
Problèmes 2011

Problème 1 Un mot de passe est consitué de six lettres, sans tenir compte de leur casse. Combien de mots de passe existe-il? Combien de mots de passe ne contiennent que des lettres distinctes?

Problème 2 Pour constituer une équipe de football (11 joueurs), on a le choix entre 20 postulants. En supposant que chaque joueur est polyvalent (il peut jouer à tous les postes), combien peut-on constituer d'équipes différentes? Si parmi les 20 postulants, 17 sont joueurs de champ et 3 sont gardiens, combien d'équipes distinctes peut-on alors constituer?

Problème 3 Combien de nombres de 3 chiffres distincts peut-on former à l'aide des six chiffres 2, 3 ,5, 6, 7, 9? Combien de ces nombres sont:
1) inférieurs à 500?
2) impairs?
3) pairs?
4) multiples de 5?

Problème 4 Cinq serveurs sont connectés par une ligne de communication. A chaque instant, un serveur donné est soit prêt à transmettre, soit déjà occupé.
1) Combien d'états du système formé des cinq serveurs y a-t-il?
2) Combien y a-t-il de possibilités que 3 serveurs exactement soient prêts à transmettre à un instant donné?

Problème 5 Combien de façons y-a-t-il de mettre 4 tours sur un échiquier de sorte qu'aucune tour n'en menace une autre ?

Problème 6 On souhaite souder 8 composants électroniques sur un circuit imprimé, 4 composants du type A et 4 composants du type B. Le circuit imprimé ne dispose que de 8 places libres alignées. Combien y-a-t-il de façons de procéder si:
1) Il n'y a pas d'autres restrictions.
2) Les composants A doivent rester ensemble et les composants B également.
3) Seuls les composants A doivent rester ensemble.
4) Les composants A et B doivent être alternés.

Problème 7 Raymond a oublié son mot de passe pour accéder à sa messagerie électronique. Il se souvient néanmoins qu'il n'a utilisé que les lettres "a", "b" et "c" et que la lettre "c" apparait r fois, dont en dernière position.
Combien de mots de passe de longueur n pourraient être le mot de passe de Raymond?

Problème 8 Afin de pouvoir mieux organiser les nombreux fichiers associés à un logiciel (fichiers de données, de code, ...), on impose un nouveau système pour nommer ceux-ci. Dans un tel système, un nom de fichier se compose d'un identificateur et d'une extension reliés par un point. Un identificateur de fichier comporte de 1 à 5 caractères dont le premier est une lettre, les caractères suivants sont soit une lettre soit un chiffre. Une extension comporte de 1 à 3 caractères, le premier est une lettre, les suivants une lettre ou un chiffre. De façon à ne pas avoir des noms trop imprononçables, on n'utilise que 20 lettres de A à T, toutes en minuscule.

1) Combien d'identificateurs de fichiers peut-on avoir?

2) Combien de noms de fichier existe-t-il?

3) Les fichiers dont l'extension commencent par la lettre "d" sont des fichiers de données. Combien de possibilités y a-t-il pour nommer les fichiers de données?

Problème 9 Calculer le nombre de diagonales d'un polygone à $n \geq 3$ côtés. Existe-t-il un polygone où le nombre de côtés soit égal au nombre de diagonales?

Problème 10 4 composants éléctroniques de type A, 3 composants de type B et 5 composants de type B doivent être montés en série. La seule contrainte au fonctionnement du système et que les composants de même type doivent rester groupés. Combien de dispositions peut-on imaginer ?

Problème 11 Il faut répartir 14 ordinateurs dans deux bureaux de 7 personnes chacun. Combien y a-t-il de répartitions possibles entre les deux bureaux?

Problème 12 Quatre bits sont transmis à travers un canal. Chaque bit est soit correctement reçu, soit reçu avec erreur de transmission. On note d un bit reçu sans erreur et e un bit reçu avec erreur.

1) Quel est l'ensemble fondamental Ω?

2) Donner l'espace des événements $\mathcal{F}$ tels que au plus un bit soit reçu avec erreur.

3) Dénombrer les événements tels que plus de la moitié des bits soient transmis avec erreur.

Problème 13 Un contrôle de qualité est effectué chaque heure sur des cartes à puce. A chaque test, la carte à puce testée est classée selon son bon ou mauvais fonctionnement. Le contrôle s'arrête lorsque deux cartes à puce consécutives sont défectueuses, ou lorsque quatre cartes à puces ont été testées. Quel est l'ensemble fondamental Ω de cette expérience? Donner l'espace des événements $\mathcal{F}$ tels que quatre cartes soient testées.

Problème 14 Dans une fabrique de processeurs, on prélève toutes les heures les trois derniers processeurs produits. Ceux-ci sont classés dans deux categories: fonctionnel, codé 1, et défectueux, codé 0.

1) Décrivez l'espace fondamental associé a cette expérience aléatoire.

2) Décrivez (en termes ensemblistes) les événements suivants:

a) le premier processeur est défectueux,

b) le dernier processeur est fonctionnel,

c) deux processeurs sont défectueux,

d) au moins deux processeurs sont fonctionnels.

Problème 15 Un disque dur a 1% de probabilité de tomber en panne. Par conséquent, il est muni de deux backups, chacun ayant 2% de chance de tomber en panne. L'information est perdue uniquement lorsque les trois composants ont en panne. En supposant que les trois composants fonctionnent indépendament les uns des autres, quelle est la probabilité que l'information soit sauvée?

Problème 16 Trois terminaux sont reliés à un ordinateur via 2 lignes de communication. Le terminal 1 a sa propre ligne alors que les terminaux 2 et 3 partagent une ligne de telle sorte qu'à chaque instant seuls deux des terminaux peuvent être utilisés. Durant une journée de travail, le terminal 1 est utilisé 30 minutes chaque heure, le terminal 2 est utilisé 10 minutes chaque heure et terminal 3 est utilisé 5 minutes chaque heure. En supposant que les lignes de communication opèrent de manière indépendantes, quelle est la probabilité qu'au moins un terminal soit opératif à un moment quelconque de la journée de travail?

Problème 17 A un test de fin d'année, trois élèves, Alain, Boris et Charles, ont oublié de mettre leur nom sur les copies. Le professeur sait quels sont ces trois élèves mais ne sait pas quelle copie appartient à chacun d'entre eux. Il estime néanmoins que Alain a 80% de chance d'avoir réussi l'examen, Boris 70% et Charles 60%. Après correction des trois copies, il s'avère que deux ont réussi l'examen et un l'a échoué. En supposant que les étudiants ont travaillé indépendamment, quelle est la probabilité que ce soit Charles qui ait échoué à l'examen?

Problème 18 Quelle est la probabilité pour que l'écriture décimale d'un nombre entier tiré au hasard entre 0 et 9999 comporte au moins une fois le chiffre 1?
Indication: utiliser la formule d'inclusion-exclusion.

Problème 19 Au jeu de loto, sous contrôle d'un huissier, 6 numéros parmi 49 numéros de 1 à 49 sont tirés au hasard. Un joueur de loto doit jouer 6 numéros distincts (on ignore ici la notion de "numéro complémentaire"). On dit qu'on a gagné la cagnotte si les 6 numéros joués correspondent aux 6 numéros tirés.

1) Détailler l'espace probabilisé.
2) Quelle est la probabilité de gagner la cagnotte?
3) La société organisatrice du loto envisage de passer de 49 numéros à 60 numéros avec un tirage de 8 numéros plutôt que 6. Aura-t-on plus de chances de gagner la cagnotte?

Problème 20 Sophie veut parier à une course de chevaux mais n'a aucune idée des performances des jockeys et chevaux. Elle décide donc de parier au hasard.

1) Sophie a-t-elle plus de chance de trouver les 3 chevaux de tête dans l'ordre lors d'une course de 17 chevaux, ou les 4 premiers chevaux dans l'ordre lors d'une course de 18 chevaux?
2) Même question que précédemment mais dans le désordre.

Problème 21 Des bits sont envoyés à la suite sur un canal. Un bit donné prend la valeur 1 avec la probabilité p. Les bits envoyés sont indépendants.

1) On enregistre deux bits à la suite b_0 et b_1.
   a) Quelle est la probabilité que les deux bits aient des valeurs différentes?
   b) Quelle est la probabilité que le second bit soit égal à 1 si le premier est égal à 0?
2) On note maintenant la valeur de quatre bits envoyés à la suite.
   a) Trouver la probabilité d'avoir au moins deux bits égaux à 1 parmi les quatre.
   b) Trouver la probabilité d'avoir au moins deux bits égaux à 1 et le quatrième bit à 0.

Problème 22 Un logiciel libre programmé par un étudiant de l'EPFL contient 5 bugs, répartis uniformément dans le code. Le code comporte 10 000 lignes dont 3000 de commentaire. Une certaine exécution du logiciel passe par 1/3 des lignes exécutables. Quelle est la probabilité que le programme s'exécute correctement?

Problème 23 n résistances sont installées en série, dont les résistances A et B.

1) Décrire l'ensemble fondamental associé à cette expérience, et en donner le nombre d'éléments.
2) Quelle est la probabilité qu'il y ait k résistances placées entre A et B $(0 \leq k \leq n-2)$?
3) Vérifier votre résultat pour $n = 3$ en explicitant tous les cas possibles.

Problème 24 Le Daily Mail du 13 octobre 2010 rapporte que la famille Allali (Angleterre) vient juste d'avoir son troisième enfant et que, fait surprenant, ses trois enfants sont tous nés le même jour (un 7 octobre, en 2005, 2007 et 2010). Le journaliste affirme qu'il y a moins d'une chance sur 48 millions que trois enfants d'une famille (ni jumeaux, ni triplets) naissent un même jour. Que pensez-vous de cette affirmation?

Problème 25 Vous venez d'installer un module de détection de courriers indésirables dans votre client de courrier électronique. Le module réussit à identifier les courriers indésirables dans 99% des cas.

Néanmoins le module annonce qu'un message est indésirable alors qu'il ne l'est pas dans 2% des cas. Les statistiques officielles indiquent que 10% des courriers électroniques reçus sont indésirables. Quelle est la probabilité qu'un message soit effectivement indésirable lorsque le module indique que c'est le cas?

Problème 26 L'envoi d'un paquet du serveur S_1 au serveur S_2 sur internet passe par deux routeurs intermédiaires R_1 et R_2. La probabilité que le paquet se perde au niveau S_1, R_1 ou R_2 est p. On constate, au niveau du serveur S_2 la perte du paquet. Quelle est la probabilité qu'il ait été perdu au niveau de S_1? de R_1? de R_2?

Problème 27 Un routeur reçoit l'essentiel des paquets qu'il fait transiter depuis deux autres routeurs R_1 et R_2, en proportion équilibrée. On constate qu'environ un paquet sur 200 est endommagé lorsque ceux-ci proviennent de R_1, alors que seulement un paquet sur 1000 est endommagé lorsqu'ils proviennent de R_2. On considère un paquet reçu de l'un de ces deux routeurs R_1 ou R_2. Le paquet inspecté n'est pas endommagé. Quelle est la probabilité qu'il provienne de R_1?

Problème 28 100 programmes informatiques sont testés contre diverses sources d'erreurs. Il est trouvé que 20 d'entre eux ont des erreurs de syntaxe, 10 des erreurs d'entrée/sortie qui ne sont pas des erreurs de syntaxe, 5 ont d'autres types d'erreurs, 6 ont à la fois des erreurs de sytaxe et des erreurs d'entrée/sortie, 3 ont des erreurs de syntaxe ainsi que d'autres types d'erreur et un seul a les trois types d'erreur. Un programme est sélectionné aléatoirement parmi ces 100 programmes.

1) Exprimer sous forme de probabilités les quantités données dans l'énoncé.

2) Quelle est la probabilité que le programme sélectionné ait des erreurs de syntaxe ou des erreurs d'entrée/sortie ou les deux?

4) Quelle est la probabilité que le programme soit bugué?

5) On sait que le programme sélectionné a une erreur de syntaxe. Quelle est la probabilité qu'il ait également une erreur d'entrée/sortie?

Problème 29 Des requêtes sont effectuées sur des serveurs via 5 lignes de communication. Les lignes de communication ne sont pas utilisées égalitairement: les pourcentages des requêtes envoyées sur les lignes 1, 2, 3, 4 et 5 sont respectivement de 20, 30, 10, 15 et 25. De plus la ligne sur laquelle est envoyée la requête dépend de la longueur de la requête: la probabilité qu'une requête excède 100 caractères sur les lignes 1 à 5 est respectivement de 0.4, 0.5, 0.6, 0.2 et 0.8. Quelle est la probabilité qu'une requête quelconque excède 100 caractères?

Problème 30 On tire au hasard deux chiffres entre 0 et 9. Quelle est la probabilité que le premier chiffre soit 6, sachant que la somme des deux est i ? La calculer pour tous les i possibles.

Problème 31 Un magasin d'électronique a remarqué que la probabilité p_n que n clients entrent dans le magasin un jour donné est $p^n(1-p)$, $n = 0, 1, \ldots$. De plus, lorsqu'un client visite le magasin, il en ressort avec un achat deux fois sur trois. Néanmoins, le client n'est pas satisfait de cet achat et le rapporte au magasin une fois sur quatre. On suppose que le fait d'acheter un produit et le fait d'en être satisfait sont des événements indépendants.

1) Quelle est la probabilité qu'un client achète un produit et en soit satisfait?

2) Si on sait que k produits vendus n'ont pas été rapporté au magasin, montrer que la probabilité conditionnelle que le magasin ait reçu la visite de n clients est donnée par

$$C_n^k p^{n-k}(2-p)^{k+1}/2^{n+1}, \quad n \geq k.$$

Problème 32 Deux pièces A et B sont reliées entre elles par une porte ouverte. Seule la pièce B possède une issue vers l'extérieur. Une guêpe initialement dans la pièce A voudrait sortir à l'air libre. Son trajet obéit aux règles suivantes :

- Lorsqu'elle est en A au temps $t = n$, alors au temps $t = n + 1$, elle reste en A avec une probabilité égale à $1/3$ ou elle passe en B avec une probabilité égale à $2/3$,

- Lorsqu'elle est en B au temps $t = n$, alors au temps $t = n+1$, elle retourne en A avec une probabilité égale à $1/4$, ou elle reste en B avec une probabilité égale à $1/2$, ou elle sort à l'air libre (noté S) avec une probabilité égale à $1/4$.

Au temps $t = 0$, la guêpe est dans la pièce A. Lorsqu'elle est sortie, elle ne revient plus. On note X_t l'endroit où se trouve la guêpe au temps t.
1) Calculer explicitement les distributions de probabilités des variables X_0 et X_1.
2) Exprimer $P(X_{n+1} = A)$ et $P(X_{n+1} = B)$ en fonction de $P(X_n = A)$ et $P(X_n = B)$.
3) Vérifier que la suite $\frac{6}{10}P(X_n = A) - \frac{3}{10}P(X_n = B)$ est constante.
4) Vérifier que la suite $\frac{4}{10}P(X_n = A) + \frac{3}{10}P(X_n = B)$ est géométrique de raison $\frac{5}{6}$.
5) En déduire l'expression de $P(X_n = A)$ et $P(X_n = B)$.
6) Justifier que, pour $n \geq 2$, $P(X_n = S) = \frac{1}{4}P(X_{n-1} = B)$ et en déduire $P(X_n = S)$.

Problème 33 La probabilité pour que la durée de vie d'un composant dépasse 10 000 heures est de 80% et on admet que les pannes des différents composants sont indépendantes.
1) Avec 3 composants, on réalise un montage "en série". Il suffit que l'un des composants tombe en panne pour que le système ne fonctionne plus. Quelle est la probabilité pour que celui-ci fonctionne au moins 10000 heures?
2) Avec les mêmes composants que précédemment, on réalise un deuxième montage "en parallèle". Dans ce cas pour que le système ne fonctionne plus il faut que tous les trois composants soient en panne.
a) Quelle est la probabilité pour que le système fonctionne au moins 10000 heures?
b) Sachant que le système, qui a été sous tension 10000 heures, est en état de fonctionnement, quelle est la probabilité qu'un composant au moins soit défectueux?

Problème 34 Dans une assemblée de n personnes, on doit constituer un comité de s personnes puis désigner un président parmi celles-ci. Les n personnes n'arrivant pas à se mettre d'accord pour former le comité, elles décident après des heures de débat de procéder par tirage au sort, aussi bien pour désigner les s personnes du comité que son président. M. Truc, membre de l'assemblé, n'est pas le président du comité. Quelle est est la probabilité que celui-ci fasse néanmoins partie du comité?

Problème 35 On installe un répartisseur de charges ("load-balancer") pour distribuer le travail entre trois serveurs, notés A, B et C. Une connexion donnée a une probabilité 1/3 d'être dirigée vers le serveur A, 1/3 vers le serveur B et 1/3 vers le serveur C. On impose néanmoins que deux connexions successives ne soient jamais dirigées vers le même serveur. On note α_n la probabilité que la $n$ième connexion soit dirigée vers A, β_n vers B et δ_n vers C.
On suppose que le serveur A a reçu la connexion 0.
1) Donner les valeurs de α_1, γ_1 et δ_1 puis de α_2, γ_2 et δ_2.
2) Calculer α_{n+1} en fonction de γ_n et δ_n puis γ_{n+1} en fonction de α_n et δ_n et δ_{n+1} en fonction de α_n et γ_n.
3) Donner α_{n+1} en fonction de α_n, γ_{n+1} en fonction de γ_n et δ_{n+1} en fonction de δ_n
4) Montrer que la suite $\alpha^* = \alpha_n - 1/3$ est une suite géométrique et préciser sa raison.
5) En déduire α_n, γ_n et δ_n en fonction de n puis leur limite pour n infini. Le résultat obtenu était-il prévisible?

Problème 36 On met au point un algorithme pour détecter si une page web écrite en anglais est rédigée par un natif (quelqu'un dont l'anglais est la langue maternelle). On évalue que 55% des pages web en anglais sont écrites par des natifs. L'algorithme réussit à détecter correctement que la page est écrite par un natif dans 95% des cas lorsque la page est effectivement écrite par un natif. Elle affirme cependant incorrectement que la page est écrite par un natif alors que ce n'est pas le cas avec probabilité 1%.
Quelle est la probabilité qu'une page soit écrite par un natif lorsque l'algorithme l'affirme?

Problème 37 Un étudiant passe un test sous la forme d'un questionnaire à choix multiple; pour chacune des questions posées, il y a 5 réponses proposées, dont une seule réponse juste. Si l'étudiant connaît la réponse juste, il la choisit. Dans le cas contraire, il choisit une réponse au hasard parmi les 5 réponses proposées. L'étudiant connait la réponse juste à 70% des questions.

1) Quelle chance a l'étudiant de répondre correctement à une question donnée?

2) Si pour une question l'étudiant a répondu correctement, quelle est la probabilité qu'il ait répondu en connaissance de cause (c'est-à-dire pas au hasard)?

Problème 38 Une requête peut être envoyée de manière équiprobable sur six serveurs, quatre étant de type X et deux serveurs étant de type Y. Deux requêtes consécutives ne peuvent pas être envoyées sur le même serveur. On considère deux requêtes consécutives.

Définissons les événements suivants $B_1 =$ "la première requête est envoyée sur un serveur de type X" et $B_2 =$ "la seconde requête est envoyée sur un serveur de type X".

1) Décrire l'ensemble fondamental Ω de cette expérience stochastique.

2) Calculer les probabilités $P(B_1)$, $P(B_2|B_1)$, $P(B_2|B_1^c)$ et $P(B_2)$.

3) Vérifier que $P(B_1 \cap B_2) + P(B_1^c \cap B_2) + P(B_1 \cap B_2^c) + P(B_1^c \cap B_2^c) = 1$.

Problème 39 Vous êtes directeur de cabinet du ministre de la santé. Une maladie est présente dans la population, dans la proportion d'une personne malade sur 10000. Un responsable d'un grand laboratoire pharmaceutique vient vous vanter son nouveau test de dépistage : si une personne est malade, le test est positif à 99%. Si une personne n'est pas malade, le test est positif à 0.1%. Autorisez-vous la commercialisation de ce test?

Problème 40 Un programme consiste en deux modules. Le nombre d'erreurs X_1 dans le premier module a la fonction de masse P_1 et le nombre d'erreurs X_2 dans le deuxième module a la fonction de masse P_2 où

x	0	1	2	3
$P_1(x)$	0.5	0.2	0.2	0.1
$P_2(x)$	0.7	0.2	0.1	0

1) Vérifier que P_1 et P_2 sont bien des fonctions de masse.

2) Tracer les fonctions de répartition de X_1 et X_2.

3) Calculer la fonction de masse de $Y = X_1 + X_2$, le nombre total d'erreurs dans le programme.

4) Tracer sa fonction de répartition.

Problème 41 On considère le jeu de dé suivant: pour participer au jeu, on mise 3 CHF. Puis on lance le dé et on récupère 15 CHF si le nombre obtenu est 6. Le jeu est-il en moyenne rentable? A partir de quel gain pour l'obtention d'un 6 devriendrait-il rentable?

Problème 42 On considère l'algorithme naïf ci-dessous pour trouver le plus grand élément, noté M, dans une liste non vide L de n éléments :

1. On initialise le maximum M avec la valeur du premier élément de la liste.
2. On parcourt ensuite la liste, du 2ième au dernier élément, et on affecte cet élément à M si celui-ci est plus grand que la valeur courante de M.

Lors du traitement d'une liste de taille n :

1) Quel est le nombre de comparaisons effectuées?

2) Quel est le nombre d'affectations à M dans le cas le plus favorable ? le plus défavorable ?

Dans la suite, on suppose que les éléments de la liste sont deux-à-deux distincts et que les $n!$ permutations possibles ont la même probabilité d'apparaître dans la liste. Soit $p(n,k)$ la probabilité pour que, lors du déroulement de l'algorithme sur une liste (aléatoire) L de taille n, k affectations soient nécessaires.

3) Montrer la récurrence suivante, pour $n \geq 2$ et $k \geq 1$:

$$p(n,k) = \frac{1}{n}p(n-1,k-1) + \frac{n-1}{n}p(n-1,k)$$

Indication: Raisonner selon que le $n$ième élément est maximal ou non.

4) Soit E_n l'espérance mathématique du nombre d'affectations effectuées lors du traitement d'une liste de taille n. Déduire de la récurrence ci-dessus la récurrence sur E_n:

$$E_n = \frac{1}{n} + E_{n-1}$$

et donner une approximation de E_n (en fonction de n) lorsque $n \to +\infty$.

Problème 43 Soit X une variable aléatoire prenant ses valeurs dans les nombres naturels. Montrer que $E(X) = \sum_{n=1}^{\infty} P(X \geq n)$.

Problème 44 Un assistant de mathématiques essaie désespérément de passer son permis de conduire. On suppose qu'il a une probabilité $p \in [0,1]$ de réussir son permis à chaque essai, et on note T le nombre d'essais nécessaires pour qu'il réussisse son permis. Donner la distribution de T et calculer sa moyenne et sa variance.

Problème 45 On désire estimer le nombre N de poissons dans un lac sans avoir à tous les compter. Une méthode pour cela est la méthode dite de la *capture-recapture*: dans un premier temps, M poissons sont capturés, marqués d'une certaine façon pour les distinguer, puis relachés dans le lac. Quelques minutes plus tard (le temps que les poissons aient pu se mélanger), n poissons sont capturés, et on note X le nombre de poissons qui sont marqués, parmi ces n poissons.
(*Cette méthode a été introduite par Laplace en 1768 pour estimer la population de la France.*)

1) Calculer la fonction de masse de X.

2) L'application de cette méthode au lac Léman donne $X = k$. Déterminer alors le nombre de poissons du lac Léman, N, le plus probable, i.e. trouver $\hat{N}$ maximisant la probabilité $P(X = k)$ (on parle de *maximum de vraisemblance*).

3) Montrer que l'espérance de $\hat{N}$ est infinie.

Problème 46 On a calculé qu'un certain logiciel produit en moyenne 25 erreurs toutes les 1000 heures d'utilisation, avec un écart-type de 2. Donner un majorant de la probabilité d'avoir plus de 30 erreurs en 1000 heures d'utilisation. Qu'en est-il si l'écart-type passe à 4?

Problème 47 Un serveur reçoit des requêtes de 8 lignes de communication. On sait que le nombre de requêtes issu d'une quelconque de ces lignes est en moyenne de 20 requêtes par seconde. On suppose de plus que le nombre de requêtes est indépendant de la ligne. On souhaite savoir si le nombre de requêtes total reçu par le serveur n'excède pas le seuil de 500 requêtes par seconde, qui est un seuil critique pour le serveur.

1) Utiliser l'inégalité de Markov pour obtenir un majorant de cette probabilité.

2) Qu'obtenez-vous en utilisant l'inégalité de Bienyamé–Chebychev?

Problème 48 Un joueur tourne une roue. A chaque essai, il a une probabilité $1-p$ de tomber sur une case lui rapportant 1 CHF et p de tomber sur la case banqueroute qui lui fera tout perdre et l'éliminera du jeu. Il débute avec une somme initiale nulle et souhaite atteindre un gain G_0, après quoi il quittera le jeu – s'il n'y a pas déjà été contraint avant. Les résultats successifs obtenus en tournant la roue sont supposés indépendants.

1) Quelle est la probabilité que le joueur atteigne le gain G_0 qu'il s'était fixé?

2) En déduire l'éspérance du gain final.

Problème 49 Pour la nouvelle année, 20% des nouveaux utilisateurs d'un certain fournisseur internet reçoivent une promotion spéciale. 20 nouvels utilisateurs s'inscrivent à ce fournisseur internet. Par quelle loi peut-on modéliser le nombre d'entre eux qui bénéficient d'une offre spéciale de nouvelle année? Calculer la probabilité qu'au moins le quart d'entre eux en bénéficient.

Problème 50 Une version avancée d'un jeu sur console sort. 60% des joueurs ont passé tous les niveaux de la version précédente. 30% d'entre eux décident d'acheter la version avancée alors que seulement 10% des joueurs n'ayant pas passé tous les niveaux décident de l'acheter.

1) Quelle est la probabilité qu'un joueur achète la version avancée du jeu?
2) Sur 10 joueurs, quel est le nombre moyen de personnes achetant la version avancée?
3) Quelle est la probabilité qu'au moins trois joueurs l'achètent?

Problème 51 On considère le jeu suivant, qui peut être joué autant de fois que voulu. Chaque tour est indépendant et à chaque fois, la probabilité de gagner est p. A chaque tour, le joueur parie une certaine quantité x, quelconque. En cas de réussite, le joueur récupère le double de sa mise. En cas d'échec, le joueur perd sa mise. Arnaud applique une stratégie simple: il joue tant qu'il n'a pas gagné. Au premier tour, il parie 100CHF. A chaque tour perdu, il parie le double de sa mise au tour précédent. Dès qu'il a gagné une fois, il s'arrête.

1) Quel est la quantité misée par Arnaud au tour n?
2) Lorsque Arnaud joue au tour n, combien a-t-il perdu aux tours précédents?

Indication: on pourra utiliser la relation $\sum_{i=0}^{n} 2^i = 2^{n+1} - 1$.

3) Quel est le gain (absolu) d'Arnaud lorsqu'il s'arrête?
4) Quelle est la loi du nombre de tours joués par Arnaud avant le premier succès?
5) Montrer que la valeur moyenne de la mise d'Arnaud au tour où il s'arrête est

$$\begin{cases} \frac{100p}{2(1-p)} & \text{si } p > 1/2 \\ \infty & \text{si } p \leq 1/2 \end{cases}$$

En donner une interprétation.

Problème 52 La production d'un certain composant donne 5% de composants défectueux. On a besoin de 10 composants non défecteux pour 10 nouveaux ordinateurs. Les composants sont testés jusqu'à ce que 10 composants non défectueux soient trouvés.

1) Quelle est la loi du nombre X de composants à tester avant d'en trouver 10 fonctionnels?
2) Quelle est la loi du nombre Y de composants fonctionels parmi 12 testés?
3) En déduire la probabilité que plus de 12 composants doivent être testés pour en trouver 10 fonctionnels?

Problème 53 On considère une classe composée de $N \geq 10$ élèves.

1) Donner la loi du nombre de paires d'élèves ayant leur anniversaire le même jour (on supposera qu'une année est composée de 365 jours pour simplifier).
2) Calculer, en fonction de N, la probabilité qu'au moins deux des élèves aient leur anniversaire le même jour .

Indication: utiliser l'approximation par la loi de Poisson (en la justifiant).

3) Quel doit être le nombre minimal d'étudiants dans la classe pour que cette probabilité soit supérieure à 60%?

Problème 54 On sait que la fonction caractéristique d'une variable aléatoire X est $\phi_X(t) = \left(\frac{e^{it}+2}{3}\right)^5$.

1) Montrer que X suit une loi binomiale dont on précisera les paramètres.
2) Calculer la probabilité que $X = 1$.

Problème 55 A un instant t donné, n serveurs cherchent à envoyer n réponses via n canaux de transmission indépendants. Ces canaux étant partagés par d'autres serveurs, il y a une probabilité p qu'un canal soit déjà occupé à cet instant. Lorsqu'un serveur essaie d'envoyer une réponse sur un canal occupé, l'envoi échoue et la réponse est mise en attente jusqu'au pas de temps suivant $(t + 1)$ lorsque le serveur retente l'envoi, et ainsi de suite jusqu'à ce que les n réponses aient été envoyées.

1) On considère une réponse particulière à envoyer,
   a) Quelle est la loi du nombre de pas de temps nécessaires à l'envoi de cette réponse?

b) Quelle la probabilité que moins de k pas de temps soient nécessaires?

2) On considère les n réponses à envoyer,

a) Quelle est la probabilité que moins de k pas de temps soient nécessaires à l'envoi de ces n réponses?

b) En déduire la probabilité qu'exactement k pas de temps soient nésessaires.

Problème 56 n serveurs sont connectés via une ligne de communication. On estime le réseau tel que, chaque minute, un serveur reçoit avec une probabilité p une requête sur cette ligne.

1) Quelle est la loi du nombre de requêtes reçues par minute par les n serveurs?

2) En déduire l'espérance et la variance de la proportion de requêtes reçues par serveur sur une minute.

Problème 57 Une enquête réalisée auprès de clients fréquentant un magasin révèle que, une fois entrés dans le magasin, 40% d'entre eux font un achat. Sur 1000 clients entrés dans le magasin, on désigne par X le nombre de ceux qui ont réellement acheté quelque chose.

1) Quelle est la distribution de X? Calculer son espérance et sa variance.

2) Calculer la probabilité que moins de 300 personnes aient fait des achats.

3) Calculer la probabilité que plus de 500 personnes aient fait des achats.

Problème 58 Un serveur Web fournit par programme 1000 pages Web par jour. En phase de tests, on constate globalement 1500 erreurs d'affichage sur ces 1000 pages. On suppose que les erreurs sont indépendantes. Quelle est la loi du nombre d'erreurs sur une page donnée? Calculer sa moyenne et son écart-type.

Problème 59 Une requête est envoyée sur un réseau. On suppose qu'il a une probabilité $p \in [0,1]$ que cette requête reçoive une réponse. Dans le cas contraire, la requête est renvoyée jusqu'à ce qu'une réponse soit reçue. On note T le nombre d'essais nécessaires à cela.

1) Donner la distribution de T et calculer sa moyenne et sa variance.

2) Montrer que la variable aléatoire T n'a pas de mémoire, c'est-à-dire:

$$P(T \geq n_0 + n | T > n_0) = P(T \geq n) \qquad (n \geq 1).$$

Problème 60 Une variable aléatoire X a pour densité

$$f(x) = \begin{cases} 2x^{-3}, & x \geq 1 \\ 0, & x < 1 \end{cases}.$$

Calculer sa moyenne et sa variance.

Problème 61 Si X admet la densité de probabilité $f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$, on dit que X est une variable aléatoire de Cauchy. Montrer que X n'admet pas d'espérance mathématique.

Problème 62 Soit $\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy$ la fonction gamma. Montrer que pour $\alpha > 0$, $\beta > 0$, la fonction

$$f(x) = \begin{cases} \frac{\beta^\alpha e^{-\beta x} x^{\alpha-1}}{\Gamma(\alpha)} & \text{si} \quad x \geq 0 \\ 0 & \text{si} \quad x < 0 \end{cases}$$

est une densité de probabilité. Soit X une variable aléatoire qui admet cette $f(x)$ pour densité de probabilité.

1) Décrire la densité $f(x)$ de X lorsque $\alpha = 1$, $\alpha < 1$ et $\alpha > 1$.

2) En employant une intégration par parties, montrer que $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$. Pouvez-vous trouver une expression pour $\Gamma(n)$ quand n est un nombre entier ?

3) Montrer que l'espérance de X est α/β.

Problème 63 La durée de vie d'un composant est une variable aléatoire continue de fonction de densité

$$f(x) = \begin{cases} \frac{k}{x^4} & \text{si} \quad x \geq 1 \\ 0 & \text{si} \quad x < 1 \end{cases}$$

1) Trouver k.
2) Calculer la fonction de répartition de la durée de vie du composant et la tracer.
3) Calculer la probabilité que la durée de vie du composant excède 3 ans.
4) Quelle est la durée de vie moyenne du composant?
5) Quel est l'écart-type de durée de vie?

Problème 64 La fonction de répartition de Pareto à deux paramètres $\theta > 0$ et $\sigma > 0$ est

$$F(x) = \begin{cases} 1 - \left(\frac{x}{\alpha}\right)^{-\theta}, & x > \alpha \\ 0, \ \ x \leq \alpha \end{cases}.$$

1) Vérifier qu'il s'agit bien d'une fonction de répartition.
2) Calculer les moments d'ordre 1 et 2. A quelle condition sont-ils finis?

Problème 65 On suppose que le temps d'attente entre deux "jobs" successifs d'une imprimante est en moyenne de 1/4 d'heure.

1) Quelle loi utiliser pour le temps d'attente (en heure) entre deux jobs?
2) Quelle est la probabilité que le job suivant soit envoyé dans les 5 minutes?

Problème 66 La durée de vie d'un processeur a une moyenne $\mu = 12$ ans et un écart-type $\sigma = 4$ ans.

1) En déduire la valeur des paramètres de la loi (à préciser) décrivant la durée de vie du processeur.
2) Calculer la probabilité que le processeur ait une durée de vie entre 10 et 15 ans.

Problème 67 On admet que la quantité de précipitations en été suit une loi normale de moyenne de 10 litres/m^2 et d'écart-type 4 litres/m^2. Pour un été de précipitations inférieures à 4 litres/m^2, l'état de sécheresse est décrété. On suppose les quantités de précipitations indépendantes d'un été à l'autre.

1) Quelle est la probabilité que l'état de sécheresse soit décrété un été donné?
2) Quelle est la probabilité que l'état de sécheresse ne soit pas décrété sur une période de 10 ans?
3) Donner la loi du nombre d'années entre deux sécheresses consécutives.
4) Quel est le nombre moyen d'années entre deux sécheresses?

Problème 68 Le téléchargement d'un logiciel se fait via deux miroirs possibles, l'un situé en Europe et l'autre aux Etats-Unis. On estime que, selon l'état du réseau, le téléchargement via l'un quelconque de ces deux miroirs met entre a et b secondes, toute durée entre ces deux valeurs étant équiprobable. Les temps de téléchargements via les deux miroirs sont indépendants.

1) Quelle est la loi du temps nécessaire au téléchargement en passant par l'un des deux miroirs?
2) Adrien demande simultanément le téléchargement du logiciel par les deux miroirs.
   a) Quelle est la loi du temps au bout le premier téléchargement du logiciel sera terminé?
   b) Calculer son espérance et sa variance.

Problème 69 Un calcul sur ordinateur nécessite l'appel à n fonctions, dont les temps de calculs sont indépendants et de loi normales $\mathcal{N}(\mu_i, \sigma_i)$. On suppose que le programme principal appelant ces fonctions ajoute un temps b, fixe, au tant total de calcul. Quelle est la loi du temps total de calcul?

Problème 70 On suppose que la durée de vie d'un individu est une variable aléatoire de densité

$$f(t) = \begin{cases} ct^2(1 - t^2) \text{ si } t \in [0, 100] \\ 0 \text{ sinon} \end{cases}.$$

1) Calculer la constante c.

2) Quelle est l'espérance de vie moyenne d'un individu?

3) Quelle est la probabilité qu'un individu ait une durée de vie entre 50 et 80 ans?

Problème 71 On dispose de deux composants électroniques dont les durées de vie (temps avant la panne), exprimées en années, sont modélisées par des variables aléatoires X et Y indépendantes. On suppose que la durée de vie moyenne du premier composant est un an, alors que le deuxième composant a une durée de vie moyenne de 6 mois.

1) Donner les lois de X et Y.

2) Calculer leurs fonctions de répartition F_X et F_Y.

On considère deux montages possibles pour ces composants : en série ou en parallèle. Le système monté en série tombe en panne dès que l'un des composants tombe en panne. Le système monté en parallèle tombe en panne lorsque les deux composants sont tombés en panne. On note S la durée de vie du système en série et T celle du système en parallèle.

3) Calculer la fonction de répartition F_T de T.

4) Calculer la fonction de répartition F_S de S.

5) Calculer la probabilité que le composant de durée de vie X fonctionne toujours à l'instant x ($x > 0$) sachant que le système en série est tombé en panne avant cet instant.

6) Sachant que le composant de durée de vie X est tombé en panne avant l'instant x, quelle est la probabilité que le système en parallèle fonctionne toujours à un instant t (avec $t > x$)?

Problème 72 Un rabais est promis à tous les 13ièmes clients qui arrivent dans un magasin. Si les arrivées des clients forment un processus de Poisson de paramètre $\lambda = 1$ client/minute, trouver la densité de probabilité des intervalles de temps séparant deux arrivées chanceuses.

Problème 73 Un système a une durée de vie aléatoire X de densité donnée (en mois) par:

$$f(x) = \begin{cases} cxe^{-x/2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Quelle est la probabilité que le système fonctionne pendant au moins 5 mois ?

Problème 74 Sur une route à sens unique, l'écoulement des voitures (par seconde) peut être décrit par un processus de Poisson de paramètre $\lambda = 1/6$. Un piéton qui veut traverser la route a besoin d'un intervalle d'au moins $4\,s$ entre deux voitures successives.
Calculer

1) la probabilité qu'il soit obligé attendre (on suppose qu'il arrive juste au moment où passe une voiture),

2) le nombre moyen de voitures qu'il voit passer avant de pouvoir traverser la route (y compris la voiture qu'il voit passer lorsqu'il arrive),

3) la durée moyenne des intervalles qui lui permettent de traverser la route.

Problème 75 Adèle attend l'ascenseur, dont la capacité maximale est de 750 kilos. L'ascenseur arrive avec 10 adultes dedans. On suppose que le poids moyen d'un adulte suit une loi normale de moyenne 70 kilos et d'écart-type de 12 kilos.

1) Quelle est la loi du poids des 10 personnes dans l'ascenseur?

2) Est-il raisonnable qu'Adèle, 55 kilos, entre dans l'ascenseur?

Problème 76 Des ingénieurs civils ayant participé à la construction du pont de Millau ont estimé que le poids W (en milliers de kilos) que la travée du pont peut supporter sans subir de dommage au niveau de sa structure, suit une loi normale, de moyenne 400 et d'écart-type 40. Supposons que le poids (également en milliers de kilos) d'une voiture est une variable aléatoire normale de moyenne 3 et d'écart-type 0.3. Combien de voitures devraient se trouver sur cette travée pour que la probabilité de rupture soit supérieure à 0.1 ?

Problème 77 Les timbres produits par la poste sont théoriquement de forme carré de côté $l = 1.5$ cm. Néanmoins, la machine étant mal calibrée, les timbres produits ne sont pas toujours exactement de côté l. On suppose que l'erreur de production est de +/- 0.2 cm et que toutes les valeurs entre ces deux bornes sont équiprobables.
1) Par quelle loi peut on modéliser les longueurs mesurées? Donner sa fonction de répartition.
2) Quelle est la loi de l'aire du timbre?

Problème 78 Soit Y une variable aléatoire de loi Gamma de paramètres 1 et 1/2. On rappelle que sa densité est

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{\pi y}} e^{-y}, & y > 0 \\ 0, & y \leq 0 \end{cases}.$$

Soit X une autre variable aléatoire dont on ne connait que la loi conditionnellement à Y: la loi de X sachant Y est la loi normale $\mathcal{N}(0, \frac{1}{2Y})$.
1) Calculer la loi du couple (X, Y).
2) Caculer la loi X.
3) Calculer et reconnaître la loi conditionnelle de Y sachant X.

Problème 79 Montrer que si X et Y sont deux variables indépendantes suivant les lois de Poisson de paramètres respectifs λ_1 et λ_2 alors $X + Y$ suit la loi de Poisson de paramètre $\lambda_1 + \lambda_2$:
1) Par calcul direct de la fonction de masse.
2) En utilisant la transformée de Laplace.

Problème 80 On suppose que X et Y sont des variables aléatoires indépendantes suivant chacune la loi uniforme sur $[0, 1]$. Montrer que $Z = X + Y$ admet pour densité

$$f_Z(z) = \begin{cases} z \text{ si } 0 \leq z \leq 1 \\ 2 - z \text{ si } 1 \leq z \leq 2 \\ 0 \text{ sinon} \end{cases}.$$

Problème 81 Soit X une variable gaussienne standard et ϵ une variable indépendante de X telle que $P(\epsilon = 1) = P(\epsilon = -1) = 1/2$. On note $Y = \epsilon X$.
1) Calculer $Cov(X, Y)$.
2) Le vecteur (X, Y) est-il gaussien?
3) Les variables X et Y sont-elles indépendantes?

Problème 82 Soit (X, Y) un vecteur aléatoire de densité conjointe

$$f_{X,Y}(x, y) = \begin{cases} c(x^2 + \frac{xy}{2}) \text{ si } 0 \leq x \leq 1, 0 \leq y \leq 2 \\ 0 \text{ sinon} \end{cases}$$

1) Déterminer la constante c.
2) Déterminer la densité marginale de X et celle de Y. Les variables aléatoires X et Y sont-elles indépendantes?
3) Calculer $P(X > Y)$.

Problème 83 Un disque dur a 350 gigabytes de disponibles. Cela est-il vraisemblablement suffisant pour stocker 300 films de taille moyenne 1.1 gigabyte et d'écart-type 0.4?

Problème 84 La mise à jour d'un logiciel nécessite l'installation de 85 fichiers, installés séquentiellement. Le temps d'installation est aléatoire, mais prend en moyenne 15 secondes par fichier, avec un écart-type de 4 secondes. Quelle est la probabilité que le logiciel soit mis à jour en moins de 20 minutes?

Problème 85 280 messages indépendants sont envoyés d'un centre de transmission. Les messages sont traités séquentiellement et le durée de transmission d'un message suit une loi exponentielle de paramètre

$\lambda = 5$ min^{-1}. Quelle est la probabilité que 200 messages soient transmis en moins d'une heure?

Problème 86 Un serveur reçoit 1000 requêtes par minute. Néanmoins, suite à des problèmes sur les lignes de communication, on estime que la probabilité qu'un requête échoue est de 4%. Ces événements étant supposés indépendants, quelle est la probabilité pour que plus de 30 requêtes échouent en une minute?

Problème 87 N touristes veulent se rendre à Zermatt et doivent donc prendre le train à Täsch. Or le train ne comporte que deux wagons de n places chacuns. On admet que les choix des touristes de monter dans la voiture de tête ou de queue sont indépendants les uns des autres et qu'un touriste quelconque a une chance sur deux de monter dans le premier wagon.

1) Quelle est la probabilité p que tous les touristes ne puissent pas monter dans la voiture qu'ils ont choisi ?

2) Combien aurait-il fallu de places dans le train si on sait que $N = 1000$ et si on veut que p soit inférieur à 0,01 ?

Problème 88 Un serveur dessert 1000 postes via 50 lignes à haut débit. Aux heures de pointe, chaque poste est occupé en moyenne pendant $2,5$ secondes par minute. On suppose les postes indépendants. Quelle est la probabilité de saturation du réseau pendant une durée moyenne d'une minute de pointe ?

Problème 89 L'utilisation d'un logiciel de reconnaissance d'écriture a produit, à partir de 1000 pages de texte, 1500 erreurs réparties au hasard.

1) Donner une valeur exacte de la probabilité qu'une page retranscrite par le logiciel contienne moins de 2 erreurs.

2) Calculer la probabilité de cet événement si on remplace la loi du nombre d'erreurs par une loi de Poisson bien choisie.

Problème 90 Une entreprise assemble 25 ordinateurs par jour. Les contrôles de qualité montrent que 95% des ordinateurs assemblés fonctionnent correctement.
Donner une bonne approximation de la probabilité qu'au moins 600 ordinateurs sans défaut aient été assemblés à l'issue de 25 journées de travail.

Problème 91 Un local doit être éclairé en permanence; lorsque l'ampoule tombe en panne, elle est immédiatement remplacée par une nouvelle ampoule. Il en existe de deux qualités : une qualité A de durée de vie (en heure) exponentielle de paramètre $\lambda_A = 0.01$ et une qualité B de durée de vie (en heures) exponentielle de paramètre $\lambda_B = 0.02$. On a stocké 40 ampoules de qualités A et 60 de qualité B. Quelle est la probabilité que cette réserve soit suffisante pour une illumination de 6500 heures du local ?

Problème 92 Le nombre de clients entrant dans un magasin un jour donné suit une loi de Poisson de paramètre 12. Quelle est la probabilité de ne pas tomber en-dessous de 250 clients sur un mois de 22 jours ouvrables ? On fera les hypothèses d'indépendance qui s'imposent.

Problème 93 A partir d'un échantillon aléatoire de 500 temps de réponse d'un serveur, on a trouvé un temps de réponse moyen de 5 millisecondes avec un écart-type 2 millisecondes.

1) Donner un intervalle de confiance à 95% du temps de réponse moyen.

2) Si l'on souhaite que la longueur de l'intervalle de confiance n'excède pas 0.3 ms, quelle devrait être la taille de l'échantillon testé?

Problème 94 Une étude effectuée sur 300 employés d'une entreprise a révélé que le nombre moyen, X, de cafés bus annuellement à la cafétéria de l'entreprise par un employé suit une loi normale de moyenne

500 et d'écart-type 100. Donner un intervalle de confiance à 95% de la moyenne et de la variance de X.

Problème 95 On admet que la durée de vie d'un composant (en mois) est de distribution normale $\mathcal{N}(\mu, 9)$. On teste 20 de ces composants pour trouver une durée de vie moyenne $\bar{y} = 100.9$. Donner un intervalle de confiance de μ d'indice 99%.

Problème 96 Les connexions internet sont parfois ralenties dû à la latence de transmission du réseau. On souhaite quantifier cette latence. On envoie pour cela 500 paquets sur le réseau. On obtient une latence moyenne de 0.8 seconde, avec un écart-type de 0.1 seconde. Donner un intervalle de confiance au seuil 99.5% de la latence moyenne dans le réseau.

Problème 97 Une entreprise souhaite recruter un ingénieur informaticien mais n'a pas encore décidé quel salaire annuel proposer pour l'embauche. Ne voulant pas proposer un salaire trop éloigné des salaires moyens proposés dans les autres entreprises, le recruteur décide d'interroger 1000 ingénieurs sur leurs salaires annuels. Il ressort de l'enquête que le salaire moyen des 1000 ingénieurs est de 48000 CHF avec un écart-type de 12000 CHF. Donner l'intervalle de confiance au seuil 90% du salaire moyen.

Problème 98 L'utilisation d'un ordinateur consomme de l'électricité. Sur 9 ordinateurs on a obtenu les coûts journaliers suivantes exprimés en CHF :

$$1,2 \qquad 0,8 \qquad 0,6 \qquad 1,1 \qquad 1,2 \qquad 0,9 \qquad 1,5 \qquad 0,9 \qquad 1,0$$

1) Calculer la moyenne et la variance de cet échantillon.
2) Déterminer un intervalle de confiance à 95% du coût journalier moyen de l'utilisation d'un ordinateur (ce coût est supposée suivre une loi normale).

Problème 99 On dit qu'une variable aléatoire Y suit une loi log-normale de paramètres (m, σ) si $X = \ln Y$ suit une loi normale $N(m, \sigma^2)$. On dispose d'un échantillon de neuf observations d'une loi log-normale de paramètres (m, σ) inconnus :

$$11,3 \qquad 2,1 \qquad 1,1 \qquad 8,9 \qquad 4,6 \qquad 5,7 \qquad 13,5 \qquad 24,5 \qquad 16,4$$

1) Construire un intervalle de confiance bilatère pour m (au niveau $\alpha = 0,05$).
2) Même question pour σ.

Problème 100 Le nombre de messages éléctroniques envoyés par 500 personnes sur la journée du 1er janvier est donné dans le tableau suivant:

nbre messages	0	1	2	3	4	5	6	7	8	9	10	> 10
nbre personnes	4	17	42	66	86	101	69	52	24	19	11	9

Charles affirme que le nombre de messages envoyé par ces 500 personnes le 1er janvier suit une loi de Poisson de paramètre 5. Appliquer le test du chi-deux pour déterminer si Charles semble avoir raison.

Problème 101 L'étude de 320 familles ayant 5 enfants s'est traduite par la distribution suivante :

Nombre de garçons	5	4	3	2	1	0
Nombre de filles	0	1	2	3	4	5
Nombre de familles	18	56	107	91	40	8

On veut comparer cette distribution à la distribution théorique qui correspond à l'équiprobabilité de la naissance d'un garçon et de la naissance d'une fille. Appliquer le test de Kolmogorov-Smirnov. Que peut-on en conclure au seuil de 5%?

Problème 102 On considère les données d'un essai visant à déterminer la solidité d'une corde d'escalade. Un morceau de 1 m de corde est mis sous tension jusqu'à cassure. On se demande si la corde peut casser à n'importe quel endroit ou si certaines parties de la corde sont plus fragiles. On mesure la distance entre la cassure et un bout de la corde et on obtient les résultats suivants (en mètre): 0.1, 0.4, 0.4, 0.6, 0.7, 0.7, 0.8, 0.9, 0.9, 0.9. Qu'en conclure en utilisant le test de Kolmogorov-Smirnov au risque 5%?

Problème 103 On mesure le temps (en millisecondes) entre deux requêtes successives envoyées à un serveur . Voici les 30 valeurs obtenues :
12.6 12.0 20.9 14.2 16.2 15.3 10.4 22.1 19.8 15 12.8 20 11.8 20.6 21.3 11.7 18 9.1 15 15.2 15.1 14.7 13.3 21.7 15.4 16.7 15.6 17.1 7.2 12.6
Selon le test du chi-deux, cette variable suit-elle une loi normale ?

Problème 104 Tester l'adéquation à la loi normale $\mathcal{N}(5,2)$ de l'échantillon suivant :

4.42	6.17	5.74	3.39	4.65	3.91	6.52	5.31	7.49	5.06	4.87	3.03	5.46	3.63	6.82
6.27	5.19	4.67	7.38	4.49	6.37	4.23	4.90	4.70	6.45	4.79	6.77	4.28	4.31	5.19

1) En appliquant le test de Kolmogorv-Smirnov
2) En appliquant le test du chi-deux.

Problème 105 On a comptabilisé le nombre de connexions reçues par un serveur sur une période de 248 secondes. Ces connexions sont supposées indépendantes les unes des autres. On a obtenu le tableau suivant donnant le nombre n_i de secondes où x_i connexions ont été reçues:

x_i	0	1	2	3	4	5
n_i	70	60	20	20	8	70

Le nombre de connexions par seconde suit-il une distribution binomiale?

Problème 106 Dans un laboratoire on a répété 18 fois une expérience pour mesurer la constante de gravitation g. La moyenne et l'écart-type empiriques sont : $\overline{x} = 9,72$ et $s = 0,06$. On supposera le modèle gaussien. L'expérience est-elle significative (au niveau $\alpha = 0,01$) pour rejeter l'hypothèse Ho : "$g = 9,81$".

Problème 107 Une entreprise est dotée de 10 imprimantes. On a comptabilisé pour chacune des imprimantes l'intervalle de temps (arrondi en minutes) entre deux impressions successives au cours d'une journée de travail (8h-18h):

Imprimante	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}
Temps (minutes)	4	15	3	17	5	7	16	6	13	14

Tracer le quantile-quantile plot des données pour déterminer si l'intervalle de temps entre deux impressions successives lancées dans cette entreprise provient d'une distribution normale.

Problème 108 On s'intéresse aux performances de deux pares-feu installés sur une machine. Le premier pare-feu est installé durant les deux premières semaines de l'étude. Le nombre d'attaques bloquées chaque jour est:

56, 47, 49, 37, 38, 60, 50, 43, 43, 59, 50, 56, 54, 58.

Le deuxième pare-feu est installé durant les 20 jours suivants. Le nombre d'attaques bloquées est alors

53, 21, 32, 49, 45, 38, 44, 33, 32, 43, 53, 46, 36, 48, 39, 35, 37, 36, 39, 45.

Construire les boxplots correspondant aux deux pares-feu et conclure quant à leur efficacité respective.

Problème 109 Un fournisseur d'accès internet étudie la charge de son réseau. Le nombre d'utilisateurs simultanés (en milliers de personnes) est comptabilisé durant 50 jours:

17.2	22.1	18.5	17.2	18.6	14.8	21.7	15.8	16.3	22.8
24.1	13.3	16.2	17.5	19	23.9	14.8	22.2	21.7	20.7
13.5	15.8	13.1	16.1	21.9	23.9	19.3	12	19.9	19.4
5.4	16.7	19.5	16.2	16.9	17.1	20.2	13.4	19.8	17.7
19.7	18.7	17.6	15.9	15.2	17.1	15	18.8	21.6	11.9

1) Résumer les données à l'aide d'un boxplot. Y a-t-il des valeurs aberrantes?

2) Le fournisseur réseau affirme que le nombre d'utilisateurs simultanés suit une loi normale. L'histogramme des données va-t-il dans ce sens?

Problème 110 Le tableau suivant présente les données du PIB par habitant en 2010 selon le fond monétaire international pour 15 pays d'Asie:

Pays	PIB en 2010 ($ US)
Afghanistan	560
Arabie Saoudite	16641
Arménie	2676
Chine	4283
Corée du Sud	20165
Inde	1176
Iran	4484
Israël	27085
Japon	42325
Koweit	32530
Pakistan	1049
Philippines	2011
Quatar	74422
Turquie	10206
Vietnam	1155

Construire le boxplot des données et l'interpréter.

Problème 111 Soit $X_1, \ldots, X_n$ un échantillon aléatoire issu de la loi uniforme $\mathcal{U}(0, b)$. Trouver l'estimateur de maximum de vraisemblance de b.

Problème 112 Considérons un échantillon aléatoire $X_1, \ldots, X_n$ issu de la loi normale dont la moyenne μ est inconnue mais l'écart-type σ est connu.

1) Déterminer l'estimateur de maximum de vraisemblance de μ. Est-il biaisé? Calculer sa variance.

2) Calculer l'information de Fisher.

Problème 113 Soit $X_1, \ldots, X_n$ un échantillon aléatoire issu d'une population de densité

$$f_{\theta,\gamma}(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{1}{\theta}(x-\gamma)} \text{ si } x > \gamma \\ 0 \text{ sinon} \end{cases}$$

où $\theta > 0$. Déterminer les estimateurs de maximum de vraisemblance de θ et γ.

Problème 114 On considère un $n-$échantillon $X_1, \ldots, X_n$ de loi $p\mathcal{U}[0, a] + (1-p)\mathcal{U}[0, b]$, i.e. que X_i suit la loi uniforme sur $[0, a]$ avec la probabilité p et la loi uniforme sur $[0, a]$ avec la probabilité $1 - p$. On suppose $a < b$ avec a et b fixés et connus.

1) Déterminer la fonction de répartition et la densité de X_1.

2) Considérons N_a la variable aléatoire égale au nombre d'individus X_i compris entre 0 et a. Quelle est la loi de N_a? En déduire son espérance et sa variance.

3) Déterminer l'estimateur de maximum de vraisemblance de p.

Problème 115 Chaque année, la hauteur maximale H de la crue d'un fleuve est mesurée. On estime qu'une crue supérieure à 6 mètres serait catastrophique. On a modélisé la loi de la variable aléatoire H comme étant de Rayleigh, i.e la densité de H est

$$f_H(x) = \frac{x}{a} e^{-\frac{x^2}{2a}} \quad , \quad x > 0,$$

où a est un paramètre inconnu. Pendant 8 ans on a observé les hauteurs de crue (en m)

$$2,5 \qquad 2,9 \qquad 1,8 \qquad 0,9 \qquad 1,7 \qquad 2,1 \qquad 2,2 \qquad 2,8$$

1) Donner l'estimateur du maximum de vraisemblance de a.

2) Une compagnie d'assurance estime qu'une catastrophe n'arrive en moyenne qu'au plus une fois tous les mille ans. Ceci peut-il être justifié par les observations?

Problème 116 On considère un n-échantillon $X_1, \ldots, X_n$ i.i.d de loi de densité de paramètre $\alpha > 0$:

$$f(x) = \left\{ \begin{array}{l} \frac{1}{\alpha} x^{-\frac{1}{\alpha}-1} \text{ si } x \geq 1 \\ 0 \text{ si } x < 1 \end{array} \right. .$$

1) Vérifier que f est une densité de probabilité.

2) Déterminer l'estimateur de maximum de vraisemblance de α.

Problème 117 Sur 100 ampoules électriques choisies au hasard, une durée totale de 88 912 heures a été trouvée.

1) Quelle loi proposez-vous pour modéliser la durée de vie d'une ampoule? Donner l'estimateur de maximum de vraisemblance de son paramètre.

2) Quelle est la probabilité que la durée de vie d'une ampoule soit entre 500 et 1000h?

Problème 118 On considère la variable aléatoire "durée d'attente à un feu rouge". On observe le temps d'attente de n personnes, temps notés $t_1, \ldots, t_n$. On fait l'hypothèse que les temps d'attente sont indépendants et de même loi uniforme $\mathcal{U}[0, \theta]$. La durée d'attente maximale à ce feu rouge est donc θ, paramètre inconnu et strictement positif.

1. Représenter le graphe de la densité de la loi uniforme $\mathcal{U}[0, \theta]$ et calculer sa moyenne et sa variance.

2) Monter que $\theta_1 = 2\bar{T}$ où $T = \frac{1}{n} \sum_{i=1}^{N} T_i$ est un estimateur sans biais et que sa variance converge vers 0 quand $n \to \infty$.

3) Calculer la fonction de répartition et la densité de la variable aléatoire $M_n = \max_{i=1}^{n} T_i$. Calculer son espérance et sa variance.

4) En déduire un deuxième estimateur sans biais $\hat{\theta}_2$. Montrer que sa variance converge également vers 0 quand $n \to \infty$.

5) Lequel des deux estimateurs θ_1 et θ_2 choisiriez-vous pour estimer θ?

6) Montrer que M_n et θ_2 convergent en probabilité vers θ.

Problème 119 Une chaîne de fabrication de composants électroniques produit un pourcentage $100 \times p$ de produits défectueux. p est inconnu mais on sait que $p \leq 0.08$. En utilisant l'estimateur du maximum de vraisemblance de p, quelle doit être la taille de l'échantillon pour que le risque soit inférieur ou égal à 0,002?

Problème 120 Le nombre de requêtes par unité de temps à un serveur particulier suit une loi de Poisson de paramètre inconnu $\theta > 0$. On sait alors que sous cette hypothèse, la durée T séparant deux arrivées successives de requêtes suit une loi exponentielle de paramètre θ et de densité:

$$f(t) = \theta e^{-\theta t} \text{ pour } t > 0.$$

On désire estimer le paramètre θ inconnu à l'aide de l'observation de n durées t_i, $i = 1, \ldots, n$ séparant des arrivées successives de requêtes.

1) Dans un premier temps, on suppose qu'aucune information a priori n'est disponible sur θ. Déterminer à partir des observations t_i l'estimateur du maximum de vraisemblance de θ.

2) Dans un second temps, on suppose que le paramètre θ suit une loi exponentielle de paramètre λ connu, de densité

$$g(\theta) = \lambda e^{\lambda\theta}$$

La densité $g(\theta)$ représente donc la loi a priori sur le paramètre θ.

a) Montrer que la loi a posteriori de θ est proportionnelle à $\theta e^{\theta(t+\lambda)}$.

b) Déterminer à partir des observations t_i les estimateurs du maximum a posteriori de θ.

c) Montrer que les deux estimateurs trouvés en 1) et 2 b) sont équivalents lorsque n est grand et interpréter ce résultat.

Problème 121 On considère n variables aléatoires réelles $X_1, X_2, ..., X_n$, indépendantes et identiquement distribuées, ayant pour fonction de densité:

$$f(x) = \begin{cases} 0 \text{ si } x < 0 \\ \alpha x \text{ si } 0 \leq x \leq \theta \\ 0 \text{ si } x > \theta \end{cases}$$

(θ est un paramètre réel > 0).

1) Déterminer la valeur de α pour laquelle f_θ est bien une fonction de densité.

2) Donner la fonction de répartition F_{M_n} du maximum M_n des variables $X_1, X_2, ..., X_n$ et calculer ses deux premiers moments.

3) Montrer que M_n converge en probabilité vers θ.

Problème 122 Les temps $T_1, T_2, \ldots, T_n$ entre deux pannes successives d'un ordinateur sont des variables aléatoires indépendantes distribuées identiquement selon la loi Γ de paramètres 2 et $\frac{1}{\theta}$, qui a pour fonction de densité:

$$f(y) = \frac{y}{\theta^2} e^{-\frac{y}{\theta}} \quad \text{si } y > 0, \quad 0 \quad \text{sinon.}$$

θ est un paramètre > 0 inconnu et à estimer.

1) On observe les n premiers temps $t_1, t_2, ..., t_n$ entre deux pannes successives; estimer θ à l'aide de $t_1, t_2, ..., t_n$ par la méthode du maximum de vraisemblance (on appellera $\hat{\theta}_n$ l'estimateur obtenu).

2) Donner le biais B_n et le risque R_n de l'estimateur $\hat{\theta}_n$ de θ.

3) $\hat{\theta}_n$ est-il consistant?

4) Que pouvez-vous dire de la distribution de $[\frac{(\hat{\theta}_n - \theta)}{\sqrt{R_n}}]$ lorsque n est grand?

Problème 123 Une entreprise produit des composants électroniques. Le responsable de la chaîne de production affirme que seul 5% des composants produits sont défectueux. Un inspecteur de qualité affirme quant à lui que la proportion de composants défectueux est plus élevée et l'estime à 10%. Afin de déterminer lequel des deux est le plus près de la réalité, un échantillon de 20 composants est prélevé aléatoirement et testé. Le test révèle 3 composants défectueux.
On se place dans un cadre bayésien. On note p la variable aléatoire égale à la probabilité qu'un composant soit défectueux et X la variable aléatoire égale au nombre de composants défectueux parmi les 20 composants. En l'absence d'information supplémentaires, on suppose a priori que les taux de défaillance de 5% et 10% sont équiprobables.

1) Traduire l'énoncé en termes probabilistes sur p et X.

2) Calculer la probabilité marginale que $X = 3$.

3) Calculer la distribution a posteriori de p et interpréter le résultat.

4) Calculer la moyenne a posteriori de p et sa variance a posteriori.

5) Conclure.

Problème 124 On s'intéresse au nombre moyen de pannes réseau hebdomadaires sur un certain réseau. On se place dans un cadre bayésien et on considère donc que ce nombre moyen, θ, est une variable aléatoire.

1) On note X la variable aléatoire égale au nombre de pannes réseau sur une semaine donnée. On suppose que la répartition des pannes sur la semaine est aléatoire. Quelle loi proposez-vous pour X, conditionnellement à θ?

2) On a trouvé sur des réseaux similaires un nombre moyen de 4 pannes par semaine avec un écart-type de 2 pannes par semaine. Proposer un choix simple de loi a priori pour θ et préciser ses paramètres. *Indication:* pensez à la densité conjuguée de la loi proposée en 1).

3) Deux pannes se sont produites sur le réseau la semaine dernière, et aucune panne cette semaine. Donner la loi a posteriori de θ, son espérance a posteriori et sa variance a posteriori.

Problème 125 Un opérateur téléphonique désire mettre à jour la statistique des appels qu'il traite. Les dernières statistiques donnaient une moyenne de 1000 appels par heure, avec un écart-type de 200 appels. Du fait de l'extension du réseau, le nombre moyen θ d'appels traités peut avoir évolué. On propose de donner une estimation bayésienne de θ. Une nouvelle enquête est donc menée. Elle révèle que, au cours de 10 heures choisies aléatoirement, 7265 appels ont été traités. On note θ la fréquence des appels téléphoniques et X le nombre d'appels durant une heure donnée.

1) Préciser la loi de X conditionnellement à θ et proposer une loi a priori pour θ.

2) Déterminer la loi a posteriori de θ.

3) Donner la moyenne et l'écart-type a posteriori de θ.

4) Déterminer l'intervalle de crédibilité au niveau 95% de θ.

Problème 126 Lors du mondial 2010 "Paul le Poulpe" prédit avec succès le vainqueur des sept matchs de l'Allemagne et de la finale. On suppose qu'à chaque tentative, Paul a une probabilité q de trouver le bon vainqueur (on ne considère pas le cas des matchs nuls) et que chaque pari est indépendant.

1) A quelles valeurs de q correspondent le pari "chanceux" (Paul parie au hasard) et le pari "visionnaire" (Paul connait à l'avance le résultat du match)?

2) Donner la loi du nombre de prédictions réussies par Paul, pour q fixé. Donner la probabilité que Paul ait réussi les 8 prédictions du mondial dans les deux types de pari du 1).

3) On se place dans un cadre bayésien pour l'estimation q sachant que Paul a prédit avec succès huit matchs lors du mondial 2010. On suppose que q peut prendre les deux valeurs trouvées en 1) et qu'il est a priori peu probable (1% de chance) qu'un poulpe soit visionnaire. Calculer la loi a posteriori de q et son espérance a posteriori. Conclure quant aux capacités de Paul.

4) On considère un troisième état correspondant au pari "instinctif" selon lequel il y a une chance sur 9 de trouver les huit vainqueurs de huit matchs.

a) En déduire la valeur de q correspondant.

b) En supposant qu'il y a a priori 1% de chance qu'un poulpe soit visionnaire mais 9% de chance qu'il ait de l'instinct, calculer la loi a posteriori de q et son espérance a posteriori. Conclure quant aux capacités de Paul.