T.D. de Probabilités et Statistiques

Licence Informatique

Gilles HUNAULT, 1998

T.D. 4 : Lois classiques et v.a. continues

1. Conditions sur n sachant $m \geq 10$ pour $\mathcal{B}(n, p)$
2. Loi de u erreurs dans un livre de v pages
3. Saturation d'un serveur multiposte
4. Retard annuel d'une montre et hypothése "déraisonnable"
5. Salles de cinéma
6. Etude de la loi trapézoidale $\mathcal{T}(a, \alpha, h, \beta, b)$
7. Addition de deux lois uniformes continues sur $[0, 1]$

1. Conditions sur n sachant $m > 10$ pour $\mathcal{B}(n,p)$

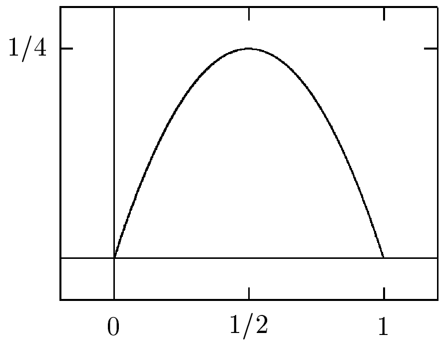
Enoncé

Montrer que $m \geq 10 \Rightarrow n \geq 40$ pour $\mathcal{B}(n,p)$.

Solution

On sait que la moyenne m de $\mathcal{B}(n,p)$ est $np(1-p)$. L'étude de la fonction $f : x \rightarrow x(1-x)$ sur $[0,1]$ montre que $f(x) \leq 1/4$.

x	0		1/2		1
f'(x)		+	0	−	
f(x)			1/4		
		$\nearrow$		$\searrow$	
	0				0



D'où : $n < 40$ et $p(1-p) < 1/4 \Rightarrow np(1-p) < 10$. *Cqfd.*

2. Loi de u erreurs dans un livre de v pages

Enoncé

Un livre de 1000 pages contient 1500 erreurs. On appelle X la v.a. "nombre d'erreurs pour une page donnée". Quelle est la loi de X ? sa moyenne ? son écart-type ?

Solution

En l'absence d'hypothèse, nous déciderons que les erreurs sont idépendantes les unes des autres. Chaque erreur apparait ou n'apparait pas et suit donc une loi binaire (*bernoulli*) de paramètre $p = 1/1000$. X est la loi "compte" du nombre d'erreurs, c'est à dire la somme de $n = 1500$ lois $b(p)$. X est donc la loi binomiale $\mathcal{B}(1500, 0.001)$. Sa moyenne $m = np$ vaut donc 1.5 et sa variance $V = np(1-p)$ vaut 1.4985 d'où un écart-type $\sigma = \sqrt{V}$ de 1.2241.

3. Saturation d'un serveur multiposte

Enoncé

Un serveur/concentrateur 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. Quelle est la probabilité de saturation du réseau pendant une durée moyenne d'une minute de pointe ?

Solution

L'énoncé ne propose aucune v.a., donc c'est à nous de tout inventer ! Prenons comme évènement élémentaire "un poste est occupé". Cet évènement est binaire : il se produit ou ne se produit pas, en terminologie probabiliste "il est réalisé" ou il ne l'est pas, et on peut donc le modéliser par une loi $b(p)$. L'énoncé indique que l'on doit prendre $p = 2,5/60$ si on ramène tout en secondes. Maintenant, introduisons la v.a. X définie par X est "le nombre de postes occupés désirant communiquer avec le serveur". X correspond alors à la somme des $n = 1000$ évènements élémentaires que l'on considérera comme indépendants. On peut représenter X par $\mathcal{B}(n, p) = \mathcal{B}(1000, 0.0417)$ La saturation correspond à l'évènement $X > 50$. Sa probabilité est

$$p(X \geq 51) = \sum_{k=51}^{1000} C_{1000}^{k}\, p^k (1-p)^{1000-k}$$

Elle vaut en gros 0,084.

On peut la calculer directement, sous Maple par exemple avec les expressions :

```
Digits := 15 ;
p := 2.5/60 ; q := 1 - p
v := Sum(binomial(n,k)*p^k*q^(n-k),k=51..n) ;
n := 1000 ;
evalf(v) ;
```

ce qui donne comme résultat numérique .08413259 en quelques secondes.

On verra dans le T.D. 8 comment réaliser une approximation de cette probabilité en utilisant la fonction de répartition de la loi normale $\mathcal{N}(0,1)$.

4. Retard annuel d'une montre

Enoncé

Une montre fait une erreur d'au plus 30 secondes par jour (dans un sens ou dans un autre). Quelle est la probabilité que l'erreur soit inférieure à 15 minutes au bout d'un an ? En quoi cet énoncé est-il "déraisonnable", "irréaliste" ?

Solution

Soit X_i l'erreur en minute au jour i. X_i peut être modélisée par une loi réelle uniformément répartie sur $[-1/2, 1/2]$. Supposant les jours indépendants, l'erreur S au bout d'un an est la somme des X_i pour i de 1 à 365 (on ignore "poliment" les années bisextiles) : $S = \sum_{i=1}^{365} X_i$. Compte-tenu de nos connaissances actuelles, il n'est pas possible de résoudre la question posée, à savoir calculer $p(|S| \leq 15)$.

On verra, là encore dans le T.D. 8, comment réaliser une approximation de cette probabilité en utilisant la fonction de répartition de la loi normale $\mathcal{N}(0,1)$.

Les hypothèses "déraisonnables", "irréalistes" sont de supposer "les jours indépendants", ce qui ne veut rien dire. Il serait plus "naturel" de penser que si une montre retarde ou avance, c'est qu'il s'agit d'un problème mécanique et non pas d'une volonté ou d'un caprice. En conséquence, il y a de fortes chances que ce problème reste constant ou s'aggrave, ce qui implique que le retard (ou l'avance) reste le même ou augmente régulièrement, ce qui signifie donc que les X_i ne sont pas indépendantes.

5. Salles de cinéma

Enoncé

Un cinéma comporte 2 salles de cinéma de n places chacune. On admettra que les choix des spectateurs sont indépendants, c'est à dire que si N personnes se présentent, un spectateur a 1 chance sur 2 d'aller dans la première salle et 1 chance sur 2 d'aller dans la deuxième salle. Comment choisir n pour que la probabilité qu'un spectateur ne puisse voir le film qu'il a choisi soit inférieur à ε ? Application : $N = 1000$, $\varepsilon = 0.001$.

Solution

Représentons l'évènement "le spectateur i veut aller dans la salle 1" S_i. Clairement, $S_i = b(1/2)$. Soit S le nombre de spectateurs ayant choisi la salle 1. Alors $S = \sum_{i=1}^{N} S_i = \mathcal{B}(N, 1/2)$. Soit de la même façon T le nombre de spectateurs ayant choisi la salle 2. Alors $T = N - S$. L'évènement "un spectateur ne peut pas voir le film qu'il a choisi" correspond aux évènements "$S > n$" ou "$T > n$" si chaque salle comporte n places. On peut rendre symétrique l'évènement "$S > n \sqcup T > n$" en l'écrivant "$S - N/2 > n - N/2 \sqcup N/2 - T < N/2 - n$" soit encore "$S - N/2 > n - N/2 \sqcup S - N/2 < N/2 - n$" c'est à dire "$|S - N/2| > n - N/2$" bien sûr dans le cas où $N \leq 2n$.

Donc, de deux choses l'une ! Soit $N > 2n$ auquel cas notre évènement est certain et sa probabilité est 1 (2000 personnes se présentent pour 2 fois 200 places par exemple donc forcément "des spectateurs ne peuvent pas voir le film qu'ils (elles) ont choisi".

Sinon, là encore nos connaissances actuelles ne nous permettent pas de résoudre la question posée, à savoir calculer n pour que $p(|S - N/2| > n - N/2| \leq \varepsilon)$. Signalons seulement qu'avec "la bonne approximation", pour $N = 1000$, $\varepsilon = 0.001$, on trouve $n \geq 540{,}8$ ce qui un peu plus que $N/2$=500.

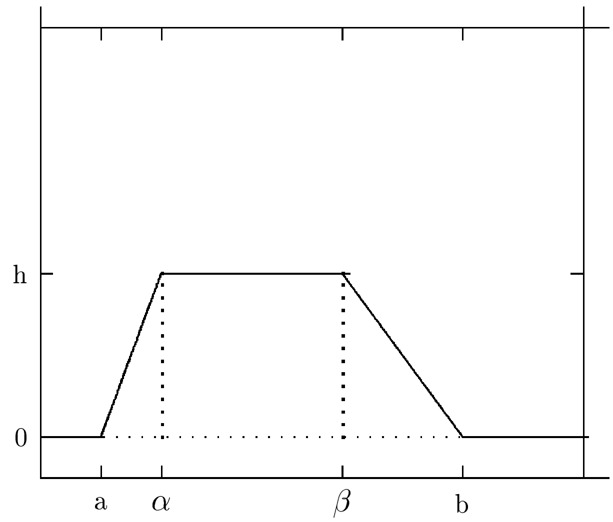
6. Etude la loi trapèze

Enoncé

Une variable aléatoire continue trapézoidale notée $\mathcal{T}(a, \alpha, h, \beta, b)$ est composée d'un segment droite oblique joignant $f(a) = 0$ à $f(\alpha) = h$ puis d'un segment de droite horizontal entre $x = \alpha$ et $x = \beta$ à la hauteur h puis d'un segment droite oblique joignant $f(\beta) = h$ à $f(b) = 0$. Donner dans le cas général l'équation des droites du trapèze et la valeur de la moyenne de $T(a, \alpha, h, \beta, b)$. *Application* : on prend $a = 1$, $\alpha = 2$, $\beta = 5$, $b = 7$. Donnez la valeur numérique de h et de la moyenne. Notant $1_{[u,v]}$ la fonction caractéristique de l'intervalle $[u, v]$ c'est à la dire la fonction qui vaut 1 pour $x \in [u, v]$ et 0 ailleurs, écrire la densite de la variable trapézoidale comme combinaison linéaire de trois fonction caractéristiques bien choisies.

Solution

Commençons par donner une représentation graphique de la densité de $\mathcal{T}(a, \alpha, h, \beta, b)$.



La partie entre a et α est un segment de droite qui s'annule en $x = a$ donc $D_1(x) = K_1.(x-a)$ comme $D_1(\alpha)$ vaut h on en déduit $K_1 = h/(\alpha - a)$. De même, la partie entre β et b est un segment de droite qui s'annule en $x = b$ donc $D_3(x) = K_3.(x-a)$ comme $D_3(\beta)$ vaut h on en déduit $K_3 = h/(\beta - b)$. Enfin, la partie entre α et β est constante et vaut h donc $D_2(x) = h$. On peut résumer ces informations avec les fonctions caractéristiques d'ensemble :

$$f(x) = K_1.1_{[a,\alpha]} + h.1_{[\alpha,\beta]} + K_3.1_{[\beta,b]}$$

La moyenne m vaut $\displaystyle\int_0^\alpha IdD_3 + \int_\alpha^\beta IdD_2 + \int_\beta^b IdD_3$ soit encore

$$\frac{h}{\alpha - a}\left[\frac{x^3}{3} - a\frac{x^2}{2}\right]_a^\alpha + h\left[\frac{x^2}{2}\right]_\alpha^\beta \frac{h}{\beta - b}\left[\frac{x^3}{3} - b\frac{x^2}{2}\right]_\beta^b$$

d'où finalement

$$m = \frac{h}{\alpha - a}\left(\frac{\alpha^3}{3} - a\frac{\alpha^2}{2} - \frac{a^3}{2} + \frac{a^3}{3}\right) + h\left(\frac{\beta^2}{2} - \frac{\alpha^2}{2}\right) + \frac{h}{b-a}\left(\frac{b^3}{3} - \frac{b^3}{2} - \frac{\beta^3}{3} + b\frac{\beta^2}{2}\right)$$

Numériquement, si a=1, α=2, β=5 et b=7 alors pour avoir une densité il faut (géométriquement) $h(\alpha - a)/2 + h(\beta - \alpha) + h(b - \beta)/2$=1 soit $h(9/2)$=1 donc $h = 2/9$. On trouve alors m=34/9 soit à peu près 3,778.

7. Addition de deux lois uniformes continues

Enoncé

Soient X et Y deux lois uniformes continues sur [0,1] et indépendantes. Quelle est la loi de $Z = X + Y$?

Solution

Soit f_X la densité de X et f_Y la densité de Y. Puisque X et Y sont indépendantes, la densité g de $Z = X + Y$ est définie par

$$g(s) = \int_{-\infty}^{+\infty} f_1(x) f_2(s-x)dx$$

qui est l'équivalent de la décomposition

$$p("Z = s") = \sum_{x=0}^{s} p("X = x")p("Y = s - x")$$

pour la somme de deux variables aléatoires discrètes indépendantes.

Puisque f_1 est nulle hors de $[0, 1]$, $g(s)$ se réduit à $g(s) = \int_0^1 f_2(s - x)dx$. Le changement de variable $x = s - u$ permet de remplacer $g(s)$ par $g(s) = \int_{s-1}^{s} f_2(u)du$ et il reste à distinguer différents cas pour s avant de conclure.

Si $s < 0$ alors les bornes d'intégration sont $s - 1$ et 0 donc $g(s) = 0$ ce qui est "évident" car la somme de deux fonctions positives ne peut donner de valeur négative. De même, pour $s > 2$, les bornes sont toutes deux supérieures à 1 et f_2 étant nulle hors $[0, 1]$, $g(s) = 0$, résultat là aussi "évident" car la somme de deux fonctions inférieures à 1 ne peut donner de valeur supérieure à 2.

Reste à envisager $0 \leq s \leq 2$. Séparons en deux cas, par rapport à 1.

Si $0 \leq s \leq 1$ alors

$$g(s) = \int_{s-1}^{s} f_2(u)du = \int_0^s 1du = s$$

Si $1 \leq s \leq 2$ alors

$$g(s) = \int_{s-1}^{s} f_2(u)du = \int_{s-1}^{1} 1du = 1 - (s - 1) = 2 - s$$

Avec les notations du cours, Z est donc la loi triangle $\mathcal{T}([0, 2])$.

On laisse au lecteur le soin de démontrer que la somme Z de n lois uniformes continues sur [0,1] indépendantes de densité f_n définie par

$$f_n(x) = \int_{-\infty}^{+\infty} f_1(x - y)f_{n-1}(y)dy = \int_{x-1}^{x} f_{n-1}(y)dy$$

vaut, par récurrence

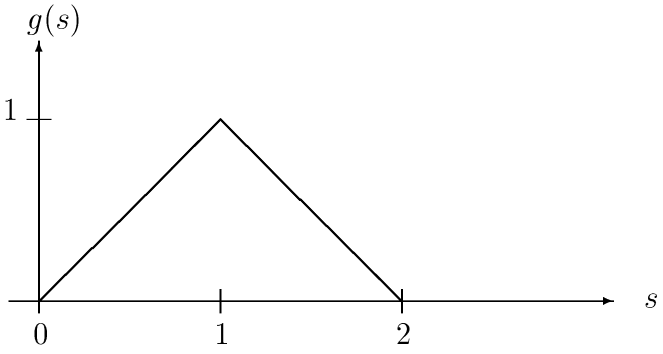
$$f_n(x) = \frac{1}{(n-1)!}\left[x^{n-1} - C_n^1(x - 1)^{n-1} + C_n^2(x - 2)^{n-1} + ...\right]$$

soit encore

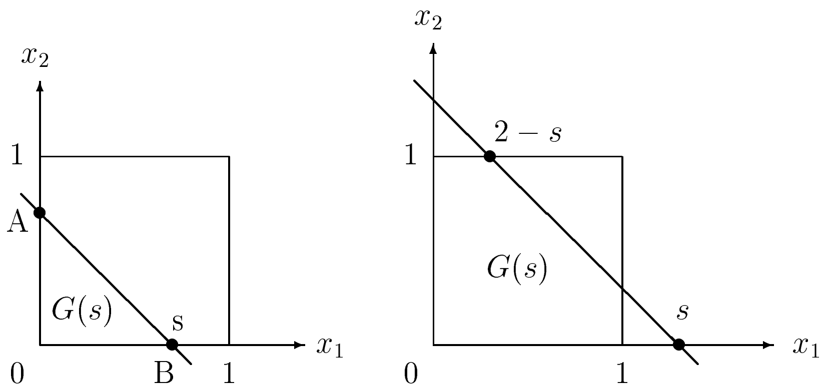
$$f_n(x) = \frac{1}{(n-1)!}\sum C_n^k(x - k)^{n-1}$$

où la sommation est étendure aussi longtemps que les termes x,$x-1$,$x-2$... sont positifs, c'est à dire jusqu'au terme en $(x-p)^{n-1}$ sur l'intervalle $[p, p+1]$ avec bien sûr $f_n(x) = 0$ pour $x \geq n$.

La représentation graphique de $\mathcal{T}([0,2])$ est



Une représentation graphique de la fonction de réparition G et un peu de calcul géométrique élémentaire aideront à comprendre les calculs (puisque $g = G'$) :



Pour la figure de gauche, s est inférieur à 1 et $G(s)$ correspond à la surface du triangle isocèle OAB où A et B sont sur la droite $x_1 + x_2 = s$. A a pour coordonnées $(0, s)$ et B $(s, 0)$. Donc $G(s) = s^2$ d'où $g(s) = s$.

Pour la figure de droite, $G(s)$ est la surface du trapèze obtenu en otant au carre $[0,1] \times [0,1]$ le triangle isocèle de coté $2-s$. Donc $G(s) = 1-(2-s)^2/2$ d'où $g(s) = 2 - s$.Cqfd.