UNIVERSITE DE CERGY PONTOISE
LICENCE D'ECONOMIE ET GESTION 2^ème^ année
ANALYSE STATISTIQUE
Examen 1^ère^ session 2013
durée : 2h
Enseignant responsable: ANDRIANASITERA

<u>Exercice1(10pts)</u>

On a prélevé un échantillon de taille 101 dans une population supposée normale de moyenne m et d'écart type σ **inconnus.** On a calculé pour l'échantillon $(x_1, x_2, \ldots x_{101})$ les deux statistiques suivantes : $\sum_{i=1}^{101} x_i = 455$ et : $\sum_{i=1}^{101} \left(x_i - \bar{x}\right)^2 = 413{,}5$

1-

a)Proposer un estimateur sans biais de la moyenne m. Déduire une estimation de m à partir de cet échantillon et construire un intervalle de confiance à 95% pour m.

b) Mêmes questions pour σ^2.

2 –Dans un premier temps on suppose $\sigma^2 = 4$ et on veut réaliser ,à partir de notre échantillon, le test suivant :

$$\begin{cases} H_0 : m = 4 \\ H_1 : m = 4{,}5 \end{cases}$$

a)Donner la statistique Z pour construire la région critique(R) ? Quelle est sa loi de probabilité ?

b)Déterminer la forme de la région critique.

Déterminer cette région critique quand on fixe le risuqe de première espèce α à 5%.

c) Calculer le risque du second espèce β de ce test et en déduire la puissance η.

d)Définir la règle de décision et conclure au vu de l'échantillon.

3-On veut maintenant réaliser , à partir de notre échantillon , le test suivant :

$$\begin{cases} H_0 : \sigma^2 = 4 \\ H_1 : \sigma^2 \neq 4 \end{cases}$$

a)Donner la statistique que Z 'pour construire la région critique(R) ?

b)Déterminer la forme de la région critique.

Déterminer cette région critique quand on fixe le risque de première espèce α à5%.et en déduire la région d'acceptation (A).

c)Définir la règle de décision et conclure.au vu de l'échantillon.

Exercice2(10pts)

On suppose que la durée de vie d'un appareil est une variable aléatoire X continue positive dont la loi de probabilité est définie par la fonction de densité $f(x)$ dépendant d'un paramètre θ telle que :

$$f(x)=\begin{cases}\theta e^{-\theta x} & x\geq 0\\ 0 & x<0\end{cases} \quad avec \quad \theta>0$$

Soit un échantillon $(X_1, X_2, ..., X_n)$ de variable X_i .Ces variables aléatoires, toutes positives , sont indépendantes en probabilité et suivent la même loi définie précédemment.

1) Sachant que la fonction génératrice des moments non centrés de la variable X $G_X(t)=E(e^{Xt})=\dfrac{\theta}{\theta-t}$ $\forall t<\theta$ calculer $E(X)$ et $V(X)$.

2) Déterminer la fonction de vraisemblance $L(x_1, x_2, ..., x_n, \theta)$.

3)Déterminer l'estimateur $\hat{\theta}$ du paramètre θ en utilisant la méthode du maximum de vraisemblance et en déduire un estimateur T de $\dfrac{1}{\theta}$.

Cet estimateur T est –il un estimateur sans biais de $\dfrac{1}{\theta}$? est-il convergent ?

Sachant que $\sum_{i=1}^{100} x_i = 20000$ *heures* et n=100,donner une estimation de θ et de $\dfrac{1}{\theta}$.

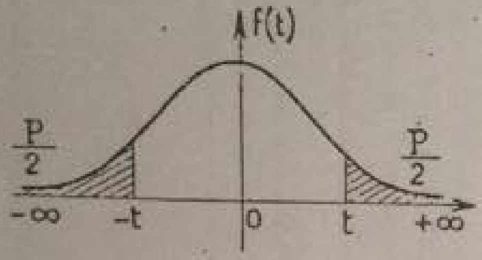
4)L'échantillon étant de grande taille ,on suppose que la variable T suit une loi normale . Enoncer le théorème qui permet de faire cette hypothèse .Quelle sont les paramètres de la loi normale ?

5)Soit U la variable aléatoire normale centrée réduite associée à T . Lire dans la table la valeur u telle que $P\left[|U|<u\right]=0,95$ et en déduire un intervalle de confiance $IC(\theta)$ de θ à 95%.

TABLE DE DISTRIBUTION DE t
(Loi de Student)

Valeurs de t ayant la probabilité P d'être dépassées en valeur absolue

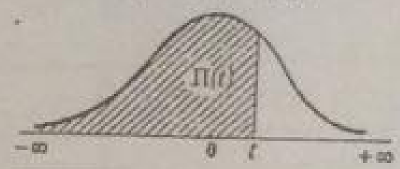
P(|T| ≥ t) = P



ν \ P	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,929
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

EXTRAITS DE LA TABLE DE LA FONCTION INTÉGRALE DE LA LOI NORMALE CENTRÉE, RÉDUITE $\mathcal{N}(0, 1)$

$\Pi(t) = \Pr\{T < t\}$.



t	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500 0	0,504 0	0,508 0	0,512 0	0,516 0	0,519 9	0,523 9	0,527 9	0,531 9	0,535 9
0,1	0,539 8	0,543 8	0,547 8	0,551 7	0,555 7	0,559 6	0,563 6	0,567 5	0,571 4	0,575 3
0,2	0,579 3	0,583 2	0,587 1	0,591 0	0,594 8	0,598 7	0,602 6	0,606 4	0,610 3	0,614 1
0,3	0,617 9	0,621 7	0,625 5	0,629 3	0,633 1	0,636 8	0,640 6	0,644 3	0,648 0	0,651 7
0,4	0,655 4	0,659 1	0,662 8	0,666 4	0,670 0	0,673 6	0,677 2	0,680 8	0,684 4	0,687 9
0,5	0,691 5	0,695 0	0,698 5	0,701 9	0,705 4	0,708 8	0,712 3	0,715 7	0,719 0	0,722 4
0,6	0,725 7	0,729 0	0,732 4	0,735 7	0,738 9	0,742 2	0,745 4	0,748 6	0,751 7	0,754 9
0,7	0,758 0	0,761 1	0,764 2	0,767 3	0,770 4	0,773 4	0,776 4	0,779 4	0,782 3	0,785 2
0,8	0,788 1	0,791 0	0,793 9	0,796 7	0,799 5	0,802 3	0,805 1	0,807 8	0,810 6	0,813 3
0,9	0,815 9	0,818 6	0,821 2	0,823 8	0,826 4	0,828 9	0,831 5	0,834 0	0,836 5	0,838 9
1,0	0,841 3	0,843 8	0,846 1	0,848 5	0,850 8	0,853 1	0,855 4	0,857 7	0,859 9	0,862 1
1,1	0,864 3	0,866 5	0,868 6	0,870 8	0,872 9	0,874 9	0,877 0	0,879 0	0,881 0	0,883 0
1,2	0,884 9	0,886 9	0,888 8	0,890 7	0,892 5	0,894 4	0,896 2	0,898 0	0,899 7	0,901 5
1,3	0,903 2	0,904 9	0,906 6	0,908 2	0,909 9	0,911 5	0,913 1	0,914 7	0,916 2	0,917 7
1,4	0,919 2	0,920 7	0,922 2	0,923 6	0,925 1	0,926 5	0,927 9	0,929 2	0,930 6	0,931 9
1,5	0,933 2	0,934 5	0,935 7	0,937 0	0,938 2	0,939 4	0,940 6	0,941 8	0,942 9	0,944 1
1,6	0,945 2	0,946 3	0,947 4	0,948 4	0,949 5	0,950 5	0,951 5	0,952 5	0,953 5	0,954 5
1,7	0,955 4	0,956 4	0,957 3	0,958 2	0,959 1	0,959 9	0,960 8	0,961 6	0,962 5	0,963 3
1,8	0,964 1	0,964 9	0,965 6	0,966 4	0,967 1	0,967 8	0,968 6	0,969 3	0,969 9	0,970 6
1,9	0,971 3	0,971 9	0,972 6	0,973 2	0,973 8	0,974 4	0,975 0	0,975 6	0,976 1	0,976 7
2,0	0,977 2	0,977 8	0,978 3	0,978 8	0,979 3	0,979 8	0,980 3	0,980 8	0,981 2	0,981 7
2,1	0,982 1	0,982 6	0,983 0	0,983 4	0,983 8	0,984 2	0,984 6	0,985 0	0,985 4	0,985 7
2,2	0,986 1	0,986 4	0,986 8	0,987 1	0,987 5	0,987 8	0,988 1	0,988 4	0,988 7	0,989 0
2,3	0,989 3	0,989 6	0,989 8	0,990 1	0,990 4	0,990 6	0,990 9	0,991 1	0,991 3	0,991 6
2,4	0,991 8	0,992 0	0,992 2	0,992 5	0,992 7	0,992 9	0,993 1	0,993 2	0,993 4	0,993 6
2,5	0,993 8	0,994 0	0,994 1	0,994 3	0,994 5	0,994 6	0,994 8	0,994 9	0,995 1	0,995 2
2,6	0,995 3	0,995 5	0,995 6	0,995 7	0,995 9	0,996 0	0,996 1	0,996 2	0,996 3	0,996 4
2,7	0,996 5	0,996 6	0,996 7	0,996 8	0,996 9	0,997 0	0,997 1	0,997 2	0,997 3	0,997 4
2,8	0,997 4	0,997 5	0,997 6	0,997 7	0,997 7	0,997 8	0,997 9	0,997 9	0,998 0	0,998 1
2,9	0,998 1	0,998 2	0,998 2	0,998 3	0,998 4	0,998 4	0,998 5	0,998 5	0,998 6	0,998 6

TABLE POUR LES GRANDES VALEURS DE t

t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
Π(t)	0,998 65	0,999 04	0,999 31	0,999 52	0,999 66	0,999 76	0,999 841	0,999 928	0,999 968	0,999 997

Nota. — La table donne les valeurs de $\Pi(t)$ pour t positif. Lorsque t est négatif il faut prendre le complément à l'unité de la valeur lue dans la table.

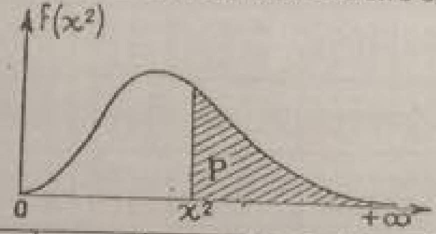
Exemple : pour $t = 1{,}37$ $\quad \Pi(t = 1{,}37) = 0{,}914\,7$

pour $t = -1{,}37$ $\quad \Pi(t = -1{,}37) = 1 - 0{,}9147 = 0{,}085\,3$

TABLES STATISTIQUES

TABLE DE DISTRIBUTION DE χ^2
(Loi de K. Pearson)

Valeurs de χ^2 ayant la probabilité P d'être dépassées



ν \ P	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010	0,001
1	0,0002	0,0010	0,0039	0,0158	2,71	3,84	5,02	6,63	10,83
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	13,82
3	0,12	0,22	0,35	0,58	6,25	7,81	9,35	11,34	16,27
4	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	18,47
5	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	20,52
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	22,46
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,47	24,32
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	26,13
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	27,88
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	29,59
11	3,05	3,82	4,57	5,58	17,27	19,67	21,92	24,72	31,26
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	32,91
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	34,53
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	36,12
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	37,70
16	5,81	6,91	7,96	9,31	23,54	26,30	28,84	32,00	39,25
17	6,41	7,56	8,67	10,08	24,77	27,59	30,19	33,41	40,79
18	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,80	42,31
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	43,82
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	45,32
21	8,90	10,28	11,59	13,24	29,61	32,67	35,48	38,93	46,80
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	48,27
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	49,73
24	10,86	12,40	13,85	15,66	33,20	36,41	39,37	42,98	51,18
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31	52,62
26	12,20	13,84	15,38	17,29	35,56	38,88	41,92	45,64	54,05
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	55,48
28	13,57	15,31	16,93	18,94	37,92	41,34	44,46	48,28	56,89
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59	58,30
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	59,70

Lorsque $\nu > 30$ on peut admettre que la quantité $\sqrt{2\chi^2} - \sqrt{2\nu - 1}$ suit la loi normale réduite.

Exemple

Calculer la valeur de χ^2 correspondant à une probabilité P = 0,10 de dépassement lorsque ν = 41. La Table 2 - 2 donne, pour P = 0,10, u = 1,2816. D'où :

$\chi^2 = \frac{[u + \sqrt{2\nu - 1}]^2}{2} = \frac{1}{2}\left[1,2816 + \sqrt{[illegible]}\right]^2 = \frac{1}{2}$ [illegible]

Analyse statistique
Examen 1resession 2013

Enseignant responsable : ANDRIANASITERA

Durée : 2 heures

Exercice 1 – 12 points

On a prélevé un échantillon de taille 101 dans une population supposée normale, de moyenne m et d'écart type σ **inconnus**. On a calculé pour l'échantillon $(x_1, x_2, \ldots, x_{101})$ les deux statistiques suivantes :

$$\sum_{i=1}^{101} x_i = 455 \qquad \text{et} \qquad \sum_{i=1}^{101}(x_i - \bar{x})^2 = 413,5$$

1. (a) Proposer un estimateur sans biais de la moyenne m. Déduire une estimation de m à partir de cet échantillon et construire un intervalle de confiance à 95% pour m.

On sait, puisque la population est supposée normale, que le bon estimateur de la moyenne est la moyenne empirique

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Cet échantillon donne comme estimation de m :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{455}{101} = 4,505$$

Pour construire l'intervalle de confiance, il faut considéré l'écart type comme inconnu, et on utilise

$$P\left(\frac{\bar{X}_{101} - m}{\frac{S}{\sqrt{101}}} < t_\alpha\right) = 1 - \alpha$$

Et

$$\frac{\bar{X}_{101} - m}{\frac{S}{\sqrt{101}}} \rightsquigarrow T_{100}$$

Avec

$$S^2 = \frac{\sum_{i=1}^{100}(X_i - \bar{X})^2}{100}$$

Étant donné α, on lit dans la table de Student la valeur de t_α telle que

$$P\left(\frac{\bar{X}_{101} - m}{\frac{S}{\sqrt{101}}} < t_\alpha\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(\bar{X} - t_\alpha \frac{S}{\sqrt{101}} < m < \bar{X} + t_\alpha \frac{S}{\sqrt{101}}\right) = 1 - \alpha$$

D'où l'intervalle de confiance

$$IC(m) = \left[\bar{x} - t_\alpha \frac{s}{\sqrt{101}}; \bar{x} + t_\alpha \frac{s}{\sqrt{101}}\right]$$

Ici $1-\alpha = 95\%$ et $n = 101$. On détermine par interpolation linéaire t_α dans la table de Student à 100 degrés de liberté. On procède de la façon suivante :

$$\frac{t_{0,05} - 2\,000}{100-80} = \frac{1980 - 2\,000}{120-80} \quad \Rightarrow \quad t_{0,05} = 2\,000 - \frac{0,02}{2} = 1,99$$

En outre,

$$\bar{x} = \frac{\sum_{i=1}^{101} x_i}{101} = 4,505$$

$$s^2 = \frac{\sum_{i=1}^{100}(x_i - \bar{x})^2}{100} = \frac{413,5}{100} \text{ soit } s \approx 2,033$$

Ce qui donne un intervalle de confiance

$$IC(m) = \left[4,505 - 1,99\frac{2,033}{\sqrt{101}}; 4,505 + 1,99\frac{2,033}{\sqrt{101}}\right] = [4,102; 4,908]$$

(b) Mêmes questions pour σ^2.

Puisque la moyenne est inconnue, on utilise comme estimateur sans biais de σ^2 la variance corrigée de l'échantillon :

$$S^2 = \frac{\sum_{i=1}^{101}(X_i - \bar{X})^2}{100}$$

Et cet échantillon donne comme estimation de σ^2 :

$$s^2 = \frac{\sum_{i=1}^{101}(x_i - \bar{x})^2}{100} = \frac{413,5}{100} = 4,135$$

On sait que $100\frac{S^2}{\sigma^2} \rightsquigarrow \chi^2_{100}$. Étant donné α, on lit dans la table les deux valeurs k_1 et k_2 telles que :

$$P(\chi^2_{100} > k_2) = \frac{\alpha}{2} \quad \text{et } P(\chi^2_{100} < k_1) = \frac{\alpha}{2}$$

$$\Leftrightarrow P\left(k_1 < \frac{100S^2}{\sigma^2} < k_2\right) = 1-\alpha$$

$$\Leftrightarrow P\left(\frac{100S^2}{k_2} < \sigma^2 < \frac{100S^2}{k_1}\right) = 1-\alpha$$

D'où l'intervalle de confiance pour σ^2

$$IC(\sigma^2) = \left[\frac{100s^2}{k_2}; \frac{100s^2}{k_1}\right] = \left[\frac{413,5}{k_2}; \frac{413,5}{k_1}\right]$$

Ici, $\alpha = 0,05$ donc

$$P(\chi^2_{100} > k_2) = \frac{\alpha}{2} = 0,025 \qquad \Rightarrow \quad k_2 = 129,561$$

$$P(\chi^2_{100} > k_1) = 1 - \frac{\alpha}{2} = 0,975 \qquad \Rightarrow \quad k_1 = 74,2219$$

Ce qui donne comme intervalle de confiance à 95% pour σ^2

$$IC(\sigma^2) = [3,192; 5,571]$$

NB : lors du partiel, la table de la loi du χ^2 fournie ne comportait pas de valeurs pour 100 degrés de liberté. Les valeurs pour 30 degrés de liberté ont été utilisées. Voici les résultalts obtenus :

$$k_1 = 16,7 \quad \text{et} \quad k_2 = 45,96$$
$$IC(\sigma^2) = [8,992; 24,76]$$

2. Dans un premier temps, on suppose $\sigma^2 = 4$ et on veut réaliser à partir de notre échantillon le test suivant :

$$\begin{cases} H_0 : m = 4 \\ H_1 : m = 4,5 \end{cases}$$

(a) Donner la statistique Z pour construire la région critique (R). Quelle est sa loi de probabilité ?

C'est un test qui porte sur la moyenne d'une variable suivant une loi $N(m, 2)$, (variance connue) donc on va s'appuyer sur la statistique

$$Z = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

qui est un estimateur sans biais de m. La région critique sera définie par la statistique $\bar{X}_n$, qui suit une loi $N(m, \frac{2}{\sqrt{101}})$

(b) Déterminer la forme de la région critique. Déterminer cette région critique quand on fixe l'erreur de première espèce α à 5%.

La région critique est de la forme $\bar{X}_n \geq k$ car la valeur de m dans l'hypothèse H_1 est supérieure à celle de m dans H_0, et k est tel que

$$\begin{aligned} \alpha &= P_{m=4}(\bar{X}_n \geq k) \\ &= P_{U \leadsto N(0;1)}(U \geq t_\alpha) \text{ avec } U = \frac{\bar{X}_n - 4}{\frac{2}{\sqrt{101}}} \end{aligned}$$

On trouve dans une table de la loi normale centrée réduite la valeur de t_α et on en déduit

$$k = 4 + t_\alpha \frac{2}{\sqrt{101}}$$

et la région critique (R) est définie par

$$\bar{x}_n \geq 4 + t_\alpha \frac{2}{\sqrt{101}}$$

Pour $\alpha = 5\%$, on trouve $t_\alpha = 1,645$ et on en déduit

$$k = 4 + 1,645\frac{2}{\sqrt{101}} = 4,327$$

La région critique est donc définie par $\bar{x}_{101} \geq 4,327$

(c) Calculer l'erreur de seconde espèce β de ce test et en déduire la puissance η.

$$\begin{aligned}\beta &= P_{m=4,5}(\bar{X}_n \leq k)\\ &= P_{U\rightsquigarrow N(0;1)}(U \leq \frac{4,327-4,5}{\frac{2}{\sqrt{101}}})\\ &= \Phi(-0,8693) = 1-\Phi(0,8693)\\ &= 1-0,80785 = 0,192\\ \Rightarrow \eta &= 1-\beta\\ &= 0,80785\end{aligned}$$

(d) Définir la règle de décision et conclure au vu de l'échantillon.

La règle de décision est la suivante : si $\bar{x}_{101} \geq 4,327$, on rejette l'hypothèse $m=4$
Décision : $\bar{x} = 4,505 \in (R)$, donc on rejette l'hypothèse $m=4$.

3. On veut maintenant réaliser, à partir de notre échantillon, le test suivant :

$$\begin{cases} H_0 : \sigma^2 = 4\\ H_1 : \sigma^2 \neq 4\end{cases}$$

(a) Donner la statistique Z' pour construire la région critique (R).

Ce test est un test bilatéral portant sur la variance d'une variable suivant une loi $N(m,\sigma)$, dont la moyenne m est inconnue. La statistique Z' que l'on peut utiliser est la suivante :

$$Z' = \sum_{i=1}^{101}(X_i - \bar{X}_{101})^2$$

(b) Déterminer la forme de la région critique. Déterminer cette région critique quand on fixe l'erreur de première espèce α à 5%, et en déduire la région d'acceptation (A).

La région critique est de la forme suivante :

$$(R) : \left\{\sum_{i=1}^{101}(x_i - \bar{x}_{101})^2 \leq k_1\right\} \cup \left\{\sum_{i=1}^{101}(x_i - \bar{x}_{101})^2 \geq k_2\right\}$$

On sait que $\frac{\sum_{i=1}^{101}(X_i-\bar{X}_n)^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$. Donc sous l'hypothèse H_0,

$$\frac{\sum_{i=1}^{101}(X_i-\bar{X}_{101})^2}{4} \rightsquigarrow \chi^2_{100}$$

Pour un risque de première espèce α on obtient à l'aide de la table de χ^2_{100} les deux valeurs K_1 et K_2 suivantes :

$$\begin{aligned} P(\chi^2_{100} \geq K_1) &= 1-\frac{\alpha}{2} \quad \text{et} \quad P(\chi^2_{100} \geq K_2) = \frac{\alpha}{2}\\ \Leftrightarrow k_1 &= 4K_1 \quad \text{et} \quad k_2 = 4K_2\end{aligned}$$

Pour $\alpha = 5\%$, on lit $K_1 = 74,2219$ et $K_2 = 129,561$ et par suite, $k_1 \approx 296,888$ et $k_2 = 518,244$. La région critique de ce test est donc

$$(R) : \left\{\sum_{i=1}^{101}(x_i - \bar{x}_{101})^2 \leq 296,888\right\} \cup \left\{\sum_{i=1}^{101}(x_i - \bar{x}_{101})^2 \geq 518,244\right\}$$

Ce qui donne comme région d'acceptation

$$(A) :]296,888; 518,244[$$

NB : lors du partiel, les valeurs de la loi χ^2_{100} ont été remplacées par celles de la loi χ^2_{30}. Voici les résultats obtenus dans ce cas :

$$\begin{aligned} K_1 &= 16,7 &&\Rightarrow k_1 = 66,8 \\ K_2 &= 45,96 &&\Rightarrow k_2 = 183,84 \end{aligned}$$

$$(R) : \left\{ \sum_{i=1}^{101}(x_i - \bar{x}_{101})^2 \leq 66,8 \right\} \cup \left\{ \sum_{i=1}^{101}(x_i - \bar{x}_{101})^2 \geq 183,84 \right\}$$

$$(A) :]66,8; 183,84[$$

(c) Définir la règle de décision et en conclure au vu de l'échantillon.

D'où la règle de décision suivante : si $296,888 \leq \sum_{i=1}^{101}(x - \bar{x}_{101})^2 \leq 518,244$ on accepte l'hypothèse $\sigma^2 = 4$.
Décision : $\sum_{i=1}^{101}(x - \bar{x}_{101})^2 = 413,5 \in (A)$ donc on accepte l'hypothèse $\sigma^2 = 4$.

NB : avec les valeurs de la loi χ^2_{30}, la règle de décision est si $66,8 \leq \sum_{i=1}^{101}(x - \bar{x}_{101})^2 \leq 183,84$ on accepte l'hypothèse $\sigma^2 = 4$.
Décision : $\sum_{i=1}^{101}(x - \bar{x}_{101})^2 = 413,5 \notin (A)$ donc on rejette l'hypothèse $\sigma^2 = 4$.

Exercice 2 – 10 points

On suppose que la durée de vie d'un appareil est une variable aléatoire X continue positive dont la loi de probabilité est définie par la fonction de densité $f(x)$ dépendant d'un paramètre θ, avec $\theta > 0$, telle que

$$f(x) = \begin{cases} \theta e^{-\theta x} & \text{pour } x \geq 0 \\ 0 & \text{pour } x < 0 \end{cases}$$

Soit un échantillon $(X_1, X_2, \ldots, X_n)$ de variables X. Ces variables aléatoires, toutes positives, sont indépendantes en probabilité et suivent la même loi définie précédemment.

1. Sachant que la fonction génératrice des moments non centrés de la variables X est

$$\forall t < 0 \quad G_X(t) = E(e^{Xt}) = \frac{\theta}{\theta - t}$$

calculer $E(X)$ et $V(X)$.

D'après le cours $E(X) = G'(0)$ et $E(X^2) = G''(0)$ et donc $V(X) = E(X^2) - E(X)^2 = G''(0) - G'(0)^2$. Avec

$$\begin{aligned} G'_X(t) &= \left(\frac{\theta}{\theta - t}\right)' = \frac{\theta}{(\theta - t)^2} && \Rightarrow E(X) = G'_X(0) = \frac{1}{\theta} \\ G''_X(t) &= \left(\frac{\theta}{\theta - t}\right)'' = \left(\frac{\theta}{(\theta - t)^2}\right)' = \frac{2\theta}{(\theta - t)^3} && \Rightarrow E(X^2) = G''_X(0) = \frac{2}{\theta^2} \\ V(X) &= G''(0) - G'(0)^2 = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2} \end{aligned}$$

2. Déterminer la fonction de vraisemblance $L(x_1, x_2, \ldots, x_n, \theta)$

$$L(x_1, x_2, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$$

3. (a) Déterminer l'estimateur $\hat{\theta}$ du paramètre θ en utilisant la méthode du maximum de vraisemblance et en déduire un estimateur T de $\frac{1}{\theta}$.

L'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ s'obtient en résolvant $\frac{dln(L)}{d\theta} = 0$.

$$\begin{aligned} L(x_1, x_2, \ldots, x_n, \theta) &= ln\left(\theta^n e^{-\theta \sum_{i=1}^{n} x_i}\right) \\ &= nln\theta - \theta \sum_{i=1}^{n} x_i \end{aligned}$$

$$\begin{aligned} \frac{dln(L)}{d\theta} &= \frac{n}{\theta} - \sum_{i=1}^{n} x_i = 0 \\ \Leftrightarrow \qquad \frac{n}{\theta} &= \sum_{i=1}^{n} x_i \\ \Leftrightarrow \qquad \theta &= \frac{n}{\sum_{i=1}^{n} x_i} \\ \Rightarrow \qquad \hat{\theta} &= \frac{n}{\sum_{i=1}^{n} X_i} \end{aligned}$$

Pour obtenir l'estimateur T de $\frac{1}{\theta}$, on part de $\theta = \frac{n}{\sum_{i=1}^{n} x_i}$ et on déduit $\frac{1}{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}$ et par suite

$$T = \frac{\sum_{i=1}^{n} X_i}{n}$$

(b) Cet estimateur T est -il un estimateur sans biais de $\frac{1}{\theta}$? Est-il convergent ?

$$\begin{aligned} E(T) = E(\bar{X}_n) &= E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) \\ &= \frac{\sum_{i=1}^{n} E(X_i)}{n} \\ &= \frac{\sum_{i=1}^{n} \frac{1}{\theta}}{n} \\ &= \frac{1}{\theta} \end{aligned}$$

Donc T est un estimateur sans biais de $\frac{1}{\theta}$.

$$\begin{aligned}
V(T) = V(\bar{X}_n) &= V\left(\frac{\sum_{i=1}^n X_i}{n}\right)\\
&= \frac{\sum_{i=1}^n V(X_i)}{n^2}\\
&= \frac{\sum_{i=1}^n \frac{1}{\theta^2}}{n^2}\\
&= \frac{1}{n\theta^2}\\
\Rightarrow \lim_{n\to\infty} \frac{1}{n\theta^2} &= 0
\end{aligned}$$

T est un estimateur sans biais et $\lim_{n\to\infty} V(T) = 0$ donc il est convergent.

(c) Sachant que $\sum_{i=1}^{100} x_i = 20\ 000$ heures et $n = 100$, donner une estimation de θ et de $\frac{1}{\theta}$.

On obtient comme estimations :

$$\theta = \frac{100}{20\ 000} = \frac{1}{200} \quad \text{et} \quad \frac{1}{\theta} = 200$$

4. L'échantillon étant de grande taille, on suppose que la variable T suit une loi normale. Énoncer le théorème qui permet de faire cette hypothèse. Quels sont les paramètres de la loi normale ?

   D'après le TCL, comme les X_i sont des variables aléatoires indépendantes de même loi d'espérance $\frac{1}{\theta}$ et de variance $\frac{1}{\theta^2}$, la variable aléatoire

   $$T = \frac{\sum_{i=1}^n X_i}{n}$$

   suit une loi normale de moyenne $\frac{1}{\theta}$ et d'écart type $\frac{1}{\sqrt{n}\theta}$

5. Soit U la variable aléatoire normale centrée réduite associée à T. Lire dans la table la valeur de u telle que $P(|U| < u) = 0,95$ et en déduire un intervalle de confiance $IC(\theta)$ de θ à 95%.

$$P(|U| < u) = 0,95 \Leftrightarrow u = 1,96$$

$$\begin{aligned}
P(|U| < u) = P\left(\frac{|T-\frac{1}{\theta}|}{\frac{1}{10\theta}} < u\right) &\Leftrightarrow -u < \frac{T-\frac{1}{\theta}}{\frac{1}{10\theta}} < u\\
&\Leftrightarrow -\frac{u}{10\theta} < T - \frac{1}{\theta} < \frac{u}{10\theta}\\
&\Leftrightarrow \frac{1}{\theta} - \frac{u}{10\theta} < T < \frac{1}{\theta} + \frac{u}{10\theta}\\
&\Leftrightarrow \frac{1}{\theta}(1-\frac{u}{10}) < T < \frac{1}{\theta}(1+\frac{u}{10})\\
&\Leftrightarrow \frac{1}{T}(1-\frac{u}{10}) < \theta < \frac{1}{T}(1+\frac{u}{10})
\end{aligned}$$

Soit pour $u = 1,96$, comme $\bar{x} = \frac{20\ 000}{100} = 200$,

$$IC(\theta) = \left]\frac{0,804}{200}; \frac{1,196}{200}\right[=]0,004; 0,00598[$$