Corrections de trois exercices

Énoncé 1

1) Soit X une v.a. à densité de loi $N(0;1)$. Calculer $\mathbb{P}(X>1)$; $\mathbb{P}(X<2)$; $\mathbb{P}(-1<X<2)$ et $\mathbb{P}(|X|>1)$.
2) Soit X une v.a. à densité de loi $N(m;\sigma^2)$. Calculer $\mathbb{E}[X]$ et $Var(X)$.
3) On suppose que le poids en kilogrammes d'un nouveau-né suit une loi normale $N(3.2;0.25)$. Quelle est la probabilité qu'un nouveau-né pèse plus de 4 kg ?

Correction

1) Le terme "calculer" employé dans la question est inapproprié. Ce qu'il s'agirait de faire serait, par exemple, de calculer :

$$\mathbb{P}(X>1)=\frac{1}{\sqrt{2\pi}}\int_1^{\infty}\exp(-\frac{t^2}{2})\ dt.$$

Or on ne peut sans doute pas donner d'expression plus simple de cette intégrale. En effet il a été prouvé qu'il n'est pas possible d'exprimer une fonction primitive de $\exp(-\frac{t^2}{2})$ comme une fraction rationnelle de fonctions ne faisant intervenir que les fonctions polynômes, logarithme, exponentielle et leurs composées.

L'exercice est un exercice de lecture de table. Mais aujourd'hui on n'utilise plus guère de table. Les logarithmes sont donnés par les calculatrices ; les valeurs de la fonction de répartition d'une loi normale centrée réduite aussi (la Ti 82 stats par exemple). Sur mon ordinateur, la commande du tableur pour calculer la valeur de la fonction de répartition d'une loi normale en un point x est "=LOI.NORMALE.STANDARD(x)" (les explications données dans l'aide de ce tableur sont d'ailleurs très confuses).

Cet exercice est donc un exercice archaïque. Ce qu'il s'agit de faire c'est montrer comment on peut se ramener aux valeurs données par une table donnant la fonction de répartition d'une loi normale centrée réduite. Appelons Π cette fonction de répartition. Les tables ne donnaient que les valeurs de $\Pi(x)$ pour $x\geq 0$ (voir plus bas). Pour s'y ramener on utilisait les symétries de la fonction $\exp(-\frac{t^2}{2})$. Par exemple :

$$\begin{aligned}
\mathbb{P}(X>1) &= 1-\Pi(1)\\
\mathbb{P}(X<2) &= \Pi(2)\\
\mathbb{P}(-1<X<2) &= \Pi(2)-\Pi(-1)=\Pi(2)-(1-\Pi(1))=\Pi(2)+\Pi(1)-1\\
\mathbb{P}(|X|>1) &= \mathbb{P}(X>1)+\mathbb{P}(X<-1)=2\mathbb{P}(X>1)=2(1-\Pi(1))
\end{aligned}$$

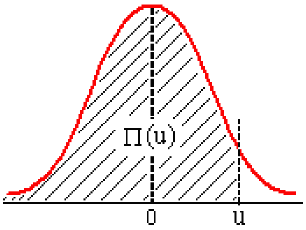
Les quantités apparaissant à droite dans ces égalités pouvaient être lues dans une table. Par exemple ici $\Pi(1)=0,84134$.

Table de Loi Normale

Fonction de répartition Π de la loi normale centrée réduite.

Probabilité de trouver une valeur inférieure à u .

$$\Pi\,(-u) = 1 - \Pi\,(u)$$



u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

2) La réponse est bien sûr $\mathbb{E}[X] = m$ et $Var(X) = \sigma^2$.

Comme je l'ai déjà dit il est inutile de chercher à calculer une primitive de $\exp(-\frac{t^2}{2})$. Mais on sait que

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{t^2}{2})\ dt = 1, \tag{1}$$

Un changement de variables permet ainsi de montrer

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt = 1.$$

Il faut se ramener à cette égalité. L'espérance de X existe si l'intégrale généralisée suivante converge

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} t \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt.$$

La fonction à intégrer est continue, donc localement intégrable. La seule chose à vérifier est donc la convergence en $+\infty$ et $-\infty$. Les propriétés de croissances comparées des fonctions assurent qu'il existe des nombres $A > 0$ et $C > 0$ tel que pour tout $|t| > A$ on ait $t \exp(-\frac{(t-m)^2}{2\sigma^2}) \leq \frac{C}{t^2}$. On en déduit la convergence de l'intégrale en $+\infty$ et $-\infty$. De même, comme $t^2 \exp(-\frac{(t-m)^2}{2\sigma^2}) \leq \frac{C}{t^2}$ si $|t|$ est assez grand, l'intégrale

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} t^2 \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt.$$

converge donc et X a une variance. Une fois la convergence des intégrales assurée le calcul peut se faire rapidement. Je choisis plutôt de revenir à la définition de la convergence d'une intégrale généralisée (les remarques précédentes sur la convergence des intégrales assurent que les limites écrites ci-dessous sont bien définies ; on pourrait aussi montrer la convergence et calculer les valeurs des intégrales en même temps en modifiant un peu la rédaction).

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} t \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt &= \lim_{A\to-\infty, B\to+\infty} \frac{1}{\sqrt{2\pi}\sigma} \int_{A}^{B} t \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt \\
&= \lim_{A\to-\infty, B\to+\infty} (\frac{1}{\sqrt{2\pi}\sigma} \int_{A}^{B} (t-m) \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt \\
&\quad + \frac{1}{\sqrt{2\pi}\sigma} \int_{A}^{B} m \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt)
\end{aligned}$$

Les deux intégrales apparaissant dans la limite convergent, on a donc :

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} t \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt &= \lim_{A\to-\infty, B\to+\infty} \frac{1}{\sqrt{2\pi}\sigma} \int_{A}^{B} (t-m) \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt \\
&\quad + \lim_{A\to-\infty, B\to+\infty} \frac{1}{\sqrt{2\pi}\sigma} \int_{A}^{B} m \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt
\end{aligned}$$

On ne peut dire plus d'une primitive de $\exp(-t^2/2)$ qu'elle est une primitive de $\exp(-t^2/2)$. En revanche, on remarque que la fonction $t\exp(-t^2/2)$ est la dérivée de $-\exp(-t^2/2)$. On a donc

$$\int_{A}^{B} (t-m) \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt = \sigma^2 \exp(-(A-m)^2/2) - \sigma^2 \exp(-(B-m)^2/2).$$

Lorsque A et B tendent respectivement vers $-\infty$ et $+\infty$ cette intégrale tend donc vers 0. D'autre part on a :

$$\lim_{A\to-\infty, B\to+\infty} \frac{1}{\sqrt{2\pi}\sigma} \int_{A}^{B} m \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt = m \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt = m.$$

Finalement on a obtenu

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} t \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt = m.$$

Pour calculer la variance de X il est préférable ici d'utiliser la formule

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[(X-m)^2] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (t-m)^2 \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt.$$

Revenons encore à la définition de l'intégrale généralisée :

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (t-m)^2 \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt = \lim_{A \to -\infty, B \to +\infty} \frac{1}{\sqrt{2\pi}\sigma} \int_{A}^{B} (t-m)^2 \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt.$$

Intégrons par parties. En posant $u = (t-m)$, $v' = (t-m)\text{exp}(-\frac{(t-m)^2}{2\sigma^2})$, $u' = 1$, $v = -\sigma^2\text{exp}(-\frac{(t-m)^2}{2\sigma^2})$, on obtient :

$$\int_{A}^{B} (t-m)^2 \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt = [-(t-m)\sigma^2 \text{exp}(-\frac{(t-m)^2}{2\sigma^2})]_A^B + \sigma^2 \int_{A}^{B} \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt.$$

Lorsque A et B tendent respectivement vers $-\infty$ et $+\infty$ la partie tout intégrée tend vers 0. On en déduit que :

$$\lim_{A \to -\infty, B \to +\infty} \frac{1}{\sqrt{2\pi}\sigma} \int_{A}^{B} (t-m)^2 \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt = \sigma^2 \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp(-\frac{(t-m)^2}{2\sigma^2})\ dt = \sigma^2.$$

Remarques :
a) L'égalité (1) peut s'obtenir de différentes manières. L'une des méthodes est de calculer le carré de cette intégrale en l'écrivant comme une intégrale double sur $\mathbb{R}^2$ puis de passer en coordonnées polaires (on a alors à intégrer une fonction de la forme $t\text{exp}(-t^2)$ dont on connaît une primitive).
b) Il est possible d'aller plus vite en utilisant les propriétés de parité de certaines fonctions et en ne revenant pas aux définitions de la convergence des intégrales généralisées. Cela dit, il faut être prudent dans la manipulation des intégrales généralisées (attention aux intégrations par parties en particulier).

3) Je ne donne pas de correction détaillée de la troisième question. C'est encore une question archaïque. Pour répondre au moyen de la seule table il faut se ramener à une loi normale centrée réduite. Pour le faire on remarque que si X suit une loi $N(3.2; 0.25)$ alors $(X - 3.2)/0.5$ suit une loi $N(0; 1)$.

Énoncé 2

On suppose que la durée de fonctionnement d'une ampoule électrique suit une loi exponentielle de paramètre $1/2$ *(l'unité de mesure est l'année).*
1) Calculer la probabilité pour que l'ampoule dure au moins 3 ans.
2) Donner la durée de vie moyenne (espérance) d'une telle ampoule.
3) On suppose que l'ampoule fonctionne depuis 2 ans. Quelle est la probabilité qu'elle fonctionne encore dans 3 ans ?

Correction

1) Notons X la durée de vie de l'ampoule. On a :

$$\mathbb{P}(X > 3) = \int_{3}^{\infty} 1/2\text{exp}(-t/2)\ dt = \exp(-3/2).$$

2) La durée de vie moyenne est l'espérance de la variable aléatoire

$$(X) = \int_{0}^{\infty} t1/2\text{exp}(-t/2)\ dt.$$

On montre que cette quantité est bien définie (l'intégrale converge) et qu'elle vaut 2.

3) Ce qu'il est demandé de calculer est la probabilité conditionnelle $\mathbb{P}_{(X>2)}(X > 5)$. Comme $(X > 5)$ est inclus dans $(X > 2)$ on a :

$$\mathbb{P}_{(X>2)}(X > 5) = \frac{\mathbb{P}((X > 2) \cap (X > 5))}{\mathbb{P}(X > 2)} = \frac{\mathbb{P}(X > 5)}{\mathbb{P}(X > 2)} = \frac{\exp(-5/2)}{\exp(-2/2)} = \exp(-3/2) = \mathbb{P}(X > 3).$$

La variable X est sans mémoire (autrement dit l'ampoule ne vieillit pas).

Énoncé 3

L'intendant d'un lycée doit nourrir 500 élèves. Il propose pâtes ou légumes ; chaque élève choisit l'un ou l'autre avec probabilité 0,5. L'intendant souhaite préparer le moins de portions possibles mais en nombre tel qu'au maximum 1/20è des élèves n'ait pas le choix. Combien doit-il prévoir de portions de pâtes ?

Correction

Le moins que l'on puisse dire est que l'énoncé de l'exercice n'est pas très clair.
Bien sûr en faisant préparer 500 portions de chaque plat l'intendant est certain que chaque élève aura le choix. Mais il est certain aussi de devoir jeter 500 portions. En faisant préparer 475 portions de chaque sorte il est sûr qu'au plus 1/20ème des élèves (25), n'aura pas le choix. Mais c'est encore un gaspillage de 450 parts.
On peut raisonner autrement et chercher combien il faut préparer de parts de chaque plat pour qu'avec grande probabilité tous les élèves aient le choix. En effet, si les élèves choisissent avec probabilité $1/2$, $1/2$ leurs plats de manière indépendante, on peut penser qu'à peu près la moitié d'entre eux choisiront pâtes, l'autre moitié légumes. Cela peut ne pas se produire. Par exemple tous les élèves pourraient choisir le même plat. Mais c'est très très peu probable.
On peut donc reformuler la question : combien faut-il prévoir de portions pour qu'avec probabilité supérieure à 1-1/20=0,95 tous les élèves aient le plat de leur choix ?
Introduisons des variables aléatoires de Bernoulli : $X_i = 1$ si le ième élève choisit légumes, $X_i = 0$ si le ième élève choisit pâtes. Les choix des élèves sont supposés indépendants ; autrement dit les X_i sont supposées indépendantes. La somme $X_1 + \ldots + X_{500}$ représente le nombre d'élèves voulant des légumes.
Notons $N + 250$ le nombre de parts préparées de chaque sorte de plat (il n'y a pas de raison d'en préparer plus d'une catégorie).
L'événement $X_1 + \ldots + X_{500} > 250 + N$ est l'événement : "un élève (au moins) souhaitant des légumes n'a pas pu en avoir".
L'événement $500 - (X_1 + \ldots + X_{500}) > 250 + N$ est l'événement : "un élève (au moins) souhaitant des pâtes n'a pas pu en avoir".
L'événement "tous les élèves ont eu satisfaction" s'écrit donc

$$250 - N \leq X_1 + \ldots + X_{500} \leq 250 + N$$

ou encore

$$|X_1 + \ldots + X_{500} - 250| \leq N.$$

Mais 250 est l'espérance de $X_1 + \ldots + X_{500}$. On cherche donc N pour que

$$\mathbb{P}(|X_1 + \ldots + X_{500} - \mathbb{E}(X_1 + \ldots + X_{500})| \geq N) \leq 0,05.$$

Or l'inégalité de Bienaymé-Tchebychev s'écrit :

$$\mathbb{P}(|X_1 + \ldots + X_{500} - \mathbb{E}(X_1 + \ldots + X_{500})| \geq N) \leq \frac{Var(X_1 + \ldots + X_{500})}{N^2} = \frac{500.0,25}{N^2},$$

(l'égalité $Var(X_1 + \ldots + X_{500}) = 500.0,25$ est une conséquence de l'indépendance des variables et de $Var(X_1) = 0,25$).

Donc pour $N \geq \sqrt{500.0,25/0,05} = \sqrt{2500} = 50$, c'est-à-dire si on prépare 300 portions de chaque plat, la probabilité que tous les étudiants aient satisfaction est supérieure à 0,95.
On peut aussi utiliser l'approximation donnée par le théorème limite central. La variable aléatoire

$$\frac{X_1 + \ldots + X_{500} - \mathbb{E}(X_1 + \ldots + X_{500})}{\sqrt{0,25.500}}$$

suit approximativement une loi normale centrée réduite (la taille de l'échantillon (500) est suffisante pour que l'approximation soit bonne). On introduit donc cette variable dans notre expression :

$$\begin{aligned}
&\mathbb{P}(|X_1 + \ldots + X_{500} - \mathbb{E}(X_1 + \ldots + X_{500})| \geq N) \\
&\quad = \mathbb{P}(\frac{|X_1 + \ldots + X_{500} - \mathbb{E}(X_1 + \ldots + X_{500})|}{\sqrt{0,25.500}} \geq \frac{N}{\sqrt{0,25.500}}) \\
&\quad \simeq \mathbb{P}(|Y| \geq \frac{N}{\sqrt{0,25.500}}),
\end{aligned}$$

où Y suit une loi normale centrée réduite. Notons Π la fonction de répartition de Y. Nous avons

$$\mathbb{P}(|Y| > \alpha) = \mathbb{P}(X > \alpha) + \mathbb{P}(X < -\alpha) = 2\mathbb{P}(X > \alpha) = 2(1 - \Pi(\alpha)).$$

Ici nous cherchons donc N tel que $2(1 - \Pi(\frac{N}{\sqrt{0,25.500}})) \leq 0,05$ soit $\Pi(\frac{N}{\sqrt{0,25.500}}) \geq 0,975$ ou encore (lecture dans la table) $\frac{N}{\sqrt{0,25.500}} \geq 1,96$ donc $N \geq 22$. Pour $N \geq 22$, c'est-à-dire si on prépare 272 portions de chaque plat, la probabilité que tous les étudiants aient satisfaction est supérieure à 0,95.

Dans la deuxième méthode (théorème limite central) on remplace une probabilité par une valeur approchée. Il faut donc préciser la valeur de cette approximation. Elle est bonne pour les grands échantillons. Vous pourrez trouver par exemple dans la littérature qu'on considère que l'approximation est valable pour un échantillon de taille supérieure à 30 (pour des variables de Bernoulli de paramètre $1/2$).

Dans la première méthode (Bienaymé-Tchebychev) on majore la probabilité. Cette majoration est assez grossière ; on prend beaucoup de marge. Mais elle est valable pour toute taille de l'échantillon.