**Université Mohammed Premier** **Échantillonnage et Estimation**
**Faculté de Droit** **S., El Melhaoui & M., Faizi**
**ECG/Grp : A & B** **Semestre 3, 2014/2015**

**Série d'exercices n°3**
**Échantillonnage et Estimation**

**Exercice 1**
Une compagnie d'assurance constate que chaque année 0.4% de ces assurés meurent à la suite d'une maladie. Posons $X$ le nombre de sinistre à payer au cours d'une année dans laquelle la compagnie gère 100 000 dossiers.

1. Déterminez la loi exacte de la variable aléatoire $X$ et déduisez la probabilité pour que la compagnie n'ait pas de sinistre à payer.
2. En utilisant une approximation adéquate, déterminez la probabilité que la compagnie payeras plus de 440 risques assurés.

**Exercice 2**
Un appareil contient un composant essentiel dont l'espérance de la durée de vie est $\mu = 10$ h, avec un écart type $\sigma = 2$ h. L'appareil doit fonctionner pendant 2200 heures.

1. Combien de composant faut il avoir en réserve pour garantir une fiabilité de 97% ?

**Exercice 3**
Selon la magazine *USA Today*, le nombre moyen des jours par an passés sur les routes pour un représentant commercial est égal à 115. L'écart type est de 60 jours par an. Supposez que ces résultats soient associés à la population des représentants commerciaux et qu'un échantillon de 50 représentants soit sélectionné.

1. Quelle est la valeur de l'écart type de la moyenne ?
2. Quelle est la probabilité que la moyenne d'échantillon soit supérieur à 115 jours par an ?
3. Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de $\pm 5$ jours de la moyenne de la population ?

4. Quelle serait la probabilité de la question (3) si la taille d'échantillon était 100 ?

**Exercice 4**

Trois firmes ont des inventaires différents par leur taille. La firme A a une population de 2000 pièces, La firme B a une population de 5000 pièces et la firme C a une population de 10 000 pièces. L'écart type de la population pour le coût d'une pièce est de $\sigma = 144$.

1. En utilisant le facteur d'exhaustivité, calculez l'erreur type pour chacune des trois firmes, étant donné que les échantillons sélectionnés sont de type ESSR et de taille 50.
2. Quelle est la probabilité que pour chaque firme, la moyenne d'échantillon $\bar{X}$ s'écarte au plus de $\pm 25$ de la moyenne de la population $\mu$.

**Exercice 5**

Une série de fabrications ne peut pas être expédiée si un échantillon de 100 pièces contient plus de 5% de pièces défectueuses. Si une série de fabrication a une proportion de pièces défectueuses égale à 10%, quelle est la probabilité que cette série de fabrication soit expédiée ?

**Exercice 6**

Soit un échantillon aléatoire simple de taille n, $(X1, ..., Xn)$, distribué suivant une loi exponentiel de paramètre $\theta > 0$ :

$$f(x, \theta) := \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & x \geq 0\,; \\ 0, & \text{sinon.} \end{cases}$$

1. Déterminez l'estimateur de $\theta$ obtenus par la méthode des moments
2. Déterminez l'estimateur de $\theta$ obtenus par la méthode du maximum de vraisemblance.
3. Déterminez les propriétés de ces estimateurs.

**Exercice 7**

Soit une population normale de moyenne $\mu$ inconnue et d'écart-type $\sigma = 1.5$ On y a prélevé un échantillon aléatoire simple d'effectif $n = 9$. Les observations sont les suivantes : 24.3 ; 24.7 ; 23.2 ; 23.8 ; 24.5 ; 26.0 ; 25.4 ; 24.8 ; 23.8.
Trouvez un intervalle de confiance pour $\mu$ au niveau de confiance 95%.

**Exercice 8**

Un comptable pense que les problèmes de liquidité d'une entreprise sont une conséquence directe de l'encaissement lent des comptes fournisseurs. Le comptable prétend qu'au moins 70% des comptes fournisseurs datent de plus de deux mois. Un échantillon de 120 comptes fournisseurs a révélé que 88 dataient de plus de deux mois.

1. Quelle est l'estimation ponctuelle de la proportion $p$ de comptes fournisseurs datant de plus de deux mois.
2. Donnez un intervalle de confiance à 99% pour la vraie proportion $p$ de comptes fournisseurs datant de plus de deux mois.
3. Déterminez la taille minimum $n$ de l'échantillon à prélevé pour que l'incertitude absolue de l'estimation (au niveau de confiance 99%) soit inférieur à 0.03.

**Exercice 9**

Une entreprise mène une politique de formation du personnel. Au cours de l'année l'ensemble de ces 200 employés ont suivi une même formation intitulée «FC». Le directeur s'interesse à l'ancienneté de ces employés ainsi qu'a leur satisfaction de la FC.

1. Le directeur a demandé à un nombre $n$ de ces employés choisis au hasard avec remise (ESAR) leurs ancienneté dans l'entreprise : la moyenne échantillon est $\bar{x} = 68$ mois et l'écart-type échantillon $s = 24$ mois. Determiner les intervalles de confiance à 90% de l'ancienneté moyenne des employés dans les cas suivants $n = 50$ et $n = 9$.
2. Le directeur a demandé à 50 de ces employés s'ils étaient satisfait de la FC ; 40 se sont déclarés satisfaits.

   a) Supposons que les employés sont choisis au hasard et avec remise. Determiner l'intervalle de confiance à 90% de la proportion des employés satisfaits de la formation.

   b) Supposons maintenant que les employés sont choisis au hasard et sans remise.
      i – Dans ces conditions peut on affirmer que L'ESSR donneras approximativement les mêmes résultats que l'ESAR ? pourquoi ?

      ii – Proposer un intervalle de confiance à 90% de la proportion des employés satisfaits de la formation.

# Exercices corrigés du TD 3

### Solution de l'Exercice 1

Posons $X$ le nombre de sinistres à payer au cours d'une année où la compagnie gère 100 000 dossiers.

1. La loi exacte la variable aléatoire $X$ est la loi binomiale $\mathcal{B}(100\,000,\ 0.004)$. Ainsi, la probabilité pour que la compagnie n'ait pas de sinistre à payer est
$$P(X = 0) = \mathrm{C}^0_{100000} 0.004^0 \times 0.996^{100\,000} \approx \mathbf{0}.$$
2. Comme $n \geq 20$, $np = 400 \geq 10$ et $nq = 99\,600 \geq 10$, on peut approximer la loi binomiale standardisée par la loi normale par application du *Théorème de Moivre-Laplace*. Ainsi,
$$\frac{X - 400}{\sqrt{398.4}} \approx \mathcal{N}(0, 1),$$

et par conséquent,

$$\begin{aligned} P(X>440) &= 1-P(X\leq 440) \\ &\approx 1-P\left(\frac{X-400}{\sqrt{398.4}}\leq\frac{440-400}{19.96}\right) \\ &\approx 1-F_Z(2) \\ &\approx 1-0.9772 \\ &\approx \mathbf{0.0028}. \end{aligned}$$

**Conclusion.** *La probabilité pour que la compagnie paye plus de* 440 *risques assurés est* 0.28%.

**Solution de l'Exercice 2**

Soit $X_i$ la durée de vie du ième composant ; il est clair que $X_i \sim Loi(10, 2^2)$ et que les $X_i$ sont indépendantes. Soit $X_T = \sum_i^n X_i$ la durée de vie de l'appareil. Supposons que $n > 30$, d'après le *Théorème Central Limite* (*TCL*)

$$\frac{X_T - n\times 10}{2\sqrt{n}} \approx \mathcal{N}(0,1).$$

Pour assurer une fiabilité de 97% il faut que

$$\begin{aligned} & P(X_T > 2200) = 0.97 \\ \Leftrightarrow\ & P(Z > \frac{2200-10n}{2\sqrt{n}}) = 0.97 \\ \Leftrightarrow\ & P(Z < \frac{2200-10n}{2\sqrt{n}}) = 0.03 \\ \Leftrightarrow\ & \frac{2200-10n}{2\sqrt{n}} = z_{0.03} \\ \Leftrightarrow\ & \frac{2200-10n}{2\sqrt{n}} = -1.88 \\ \Leftrightarrow\ & 10n - 2\times 1.88\sqrt{n} - 2200 = 0 \quad (\text{ équation du second degré}) \\ \Leftrightarrow\ & \sqrt{n} = 15.02 \\ \Leftrightarrow\ & n \approx \mathbf{226} \end{aligned}$$

**Conclusion.** *Il faut avoir en réserve* 226 *composants afin de garantir une fiabilité de* 97%.

**Solution de l'Exercice 3**

1. La valeur de l'écart type de la moyenne est
$$\sigma_{\bar{X}} = \sigma/\sqrt{n} = 60/\sqrt{50} \approx \mathbf{8.485}.$$

2. Comme $n = 50 \geq 30$, il s'agit du cas des grands nombres donc
$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - 115}{8.485} \approx \mathcal{N}(0, 1).$$
Ainsi, la probabilité que la moyenne d'échantillon soit supérieur à 115 jours par an est
$$P(\bar{X} \geq 115) = P(\frac{\bar{X} - 115}{8.485} \geq \frac{115 - 115}{8.485}) = 1 - P(Z \leq 0) = \mathbf{50}\%.$$

3. La probabilité que la moyenne d'échantillon s'écarte au plus de $\pm 5$ jours de la moyenne de la population est
$$\begin{aligned} P(|\bar{X} - 115| \leq 5) &= P(|Z| \leq 5/8.485) \\ &= P(|Z| \leq 0.59) \\ &= 2F_Z(0.59) - 1 = \mathbf{44.48}\%. \end{aligned}$$

4. Si $n = 100$ alors $\sigma_{\bar{X}} = 60/\sqrt{100} \approx \mathbf{6}$ ainsi La probabilité que la moyenne d'échantillon s'écarte au plus de $\pm 5$ jours de la moyenne de la population est
$$\begin{aligned} P(|\bar{X} - 115| < 5) &= P(|Z| < \frac{5}{6}) = P(|Z| \leq 0.83) \\ &= 2F_Z(0.83) - 1 = \mathbf{79.67}\%. \end{aligned}$$

**Solution de l'Exercice 4**

1. L'erreur type pour la firme A, étant donné que les échantillons sélectionnés sont de type ESSR et de taille 50 est :
$$\begin{aligned} \sigma_{\bar{X}}^{(A)} &= \sqrt{\frac{N - n}{N - 1} . \frac{\sigma^2}{n}} \\ &= \sqrt{\frac{2000 - 50}{2000 - 1} . \frac{144^2}{50}} \\ &\simeq \mathbf{20.11} \end{aligned}$$

Par la même façon on trouve,

$$\begin{aligned} \sigma_{\bar{X}}^{(B)} &\simeq \mathbf{20.26} \\ \sigma_{\bar{X}}^{(C)} &\simeq \mathbf{20.31} \end{aligned}$$

2. Pour la firme A, la probabilité que la moyenne d'échantillon $\bar{X}$ s'écarte au plus de $\pm 25$ de la moyenne de la population $\mu$ est :

$$\begin{aligned} P(|\bar{X}-\mu| \leq 25) &= P(|\bar{X}-\mu|/20.11 \leq 25/20.11) \\ &= P(|Z| \leq 1.24) \\ &= 2F_Z(1.24) - 1 \simeq \mathbf{78.50\%}. \end{aligned}$$

Par la même façon on trouve, pour les firme B et C, respectivement :

$$\begin{aligned} 2F_Z(1.2339) - 1 &\simeq \mathbf{78.27\%} \\ 2F_Z(1.2309) - 1 &\simeq \mathbf{78.16\%} \end{aligned}$$

---

**Conclusion.** *En conclusion, même pour des tailles de population différentes, l'ESSR donne pour une taille d'échantillon fixé approximativement les mêmes performances.*

---

**Solution de l'Exercice 5**

Dans un processus continu de fabrication, on peut supposer toujours que la taille population est assez grande par rapport à la taille échantillon, ainsi on suppose qu'on a un ESAR avec $n = 100 > 30$. Par conséquent, La distribution d'échantillonnage de $\bar{p}$ est :

$$\frac{\bar{p} - 0.1}{\sqrt{0.1(1-0.1)/100}} = \frac{\bar{p} - 0.1}{0.03} \approx \mathcal{N}(0,1).$$

Ainsi,

$$\begin{aligned} P(\bar{p} > 0.05) &\approx P(\frac{\bar{p}-0.1}{0.03} > \frac{\bar{0}.05 - 0.1}{0.03}) \\ &\approx P(Z > -1.67) \\ &= 1 - P(Z < -1.67) \\ &= 1 - F_Z(-1.67) \\ &= F_Z(1.67) \\ &= \mathbf{0.9525}. \end{aligned}$$

**Conclusion.** *Une série de fabrication qui contient* $10\%$ *de pièces défectueuses à* $1 - 0.9525 = 4.75\%$ *de chance d'échapper à ce test de contrôle.*

**Solution de l'Exercice 6**

1. Déterminons l'estimateur de $\theta$ obtenus par la méthode des moments

$$\begin{aligned} E(X) &= \frac{1}{\theta}\int_0^{+\infty} x e^{-x/\theta} dx \\ &= [-xe^{-x/\theta}]_0^{+\infty} + \int_0^{+\infty} e^{-x/\theta} dx \text{ (int. par parties)} \\ &= 0 + [-\theta\, e^{-x/\theta}]_0^{+\infty} \\ &= -\theta\,(e^{-\infty/\theta} - e^{-0/\theta}) = -\theta\,(0-1) \\ &= \theta \end{aligned}$$

ainsi

$$E(X) = \theta \Rightarrow \hat{\theta} = \bar{\mathbf{X}}.$$

2. Pour un échantillon où tous les $X_i$ sont strictement positifs la vraisemblance du paramètre $\theta$ est

$$\begin{aligned} L(X_1, ..., X_n; \theta) &= \prod_{i=1}^{n} f(X_i) \\ &= \frac{1}{\theta^n} * \exp\left(-\sum_{i=1}^{n} X_i/\theta\right). \end{aligned}$$

La log-vraisemblance du paramètre $\theta$ est

$$L^*(\theta) = \ln L(X_1, ..., X_n; \theta) = -n \ln\theta - \sum_{i=1}^{n} (X_i/\theta).$$

La fonction $L^*$ ; et par suite la fonction $L$ ; admet un maximum au point $\hat{\theta}$ si et seulement

$$\begin{cases} \dfrac{\partial L^*}{\partial \theta}(\hat{\theta}) = 0, \\ \dfrac{\partial^2 L^*}{\partial \theta^2}(\hat{\theta}) < 0 \end{cases}$$

$$\frac{\partial L^*}{\partial \theta}(\hat\theta) = 0 \Leftrightarrow -n/\hat\theta + \sum_{i=1}^{n}(X_i/\hat\theta^2) = 0 \Leftrightarrow \hat\theta = \bar X.$$

$$\frac{\partial^2 L*}{\partial \theta^2}(\hat\theta) = n/\hat\theta^2 - 2\,n\,\bar X/\hat\theta^3 = -n/\hat\theta^2 < 0.$$

Par conséquent l'estimateur de maximum de vraisemblance de $\theta$ est égale à l'estimateur obtenu par la méthode des moments :

$$\hat\theta^{(\mathbf{n})} = \bar{\mathbf{X}}.$$

3. L'estimateur $\hat\theta$ est **sans biais** puisque $E(\hat\theta) = \theta$ en effet :

$$E(\hat\theta) = E(\bar X) = E(X) = \theta.$$

L'estimateur $\hat\theta^{(n)}$ est **convergent** puisque :

$$V(\hat\theta) = V(\bar X) = V(\bar X) = \frac{V(X)}{n} = \frac{\theta^2}{n} \overset{n\mapsto+\infty}{\longrightarrow} 0.$$

**N. B.**

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{1}{\theta}\int_0^{+\infty} x^2 e^{-x/\theta} dx - \theta^2 \\ &= 2\theta^2 - \theta^2 = \theta^2. \end{aligned}$$

**Solution de l'Exercice 7**

Déterminons un intervalle de confiance pour $\mu$ au niveau de confiance 95%. On sait bien que $X \sim N(\mu, \sigma^2)$ donc

$$\bar X \sim N(\mu, \sigma^2/n),$$

par suite,

$$\begin{aligned} P[-z_{1-\alpha/2} \le \frac{\bar X - \mu}{\sigma/\sqrt{n}} \le z_{1-\alpha/2}] &= 1-\alpha \\ P[-1.96\frac{\sigma}{\sqrt{n}} \le \bar X - \mu \le 1.96\frac{\sigma}{\sqrt{n}}] &= 0.95 \\ P[-\bar X - 1.96\frac{\sigma}{\sqrt{n}} \le -\mu \le -\bar X + 1.96\frac{\sigma}{\sqrt{n}}] &= 0.95 \\ P[\bar X - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar x + 1.96\frac{\sigma}{\sqrt{n}}] &= 0.95 \end{aligned}$$

Donc,

$$\begin{aligned} I.C(0.95) &= [\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}; \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}] \\ &= [24.5 - 1.96 * 1.5/3; 24.5 + 1.96 * 1.5/3] \\ &= \mathbf{[23.52\ ;\ 25.48]} \end{aligned}$$

**Solution de l'Exercice 8**

1. L'estimation ponctuelle de la proportion $p$ de comptes fournisseurs datant de plus de deux mois est $\hat{p}$ la proportion de comptes fournisseurs datant de plus de deux mois dans l'échantillon :

$$\bar{p} = 88/120 = 0.73.$$

2. Puisque $n = 120 \geq> 30$, Le TCL donne :

$$\begin{aligned} &\frac{\bar{p} - p}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \approx N(0,1) \\ \Rightarrow\ &P[-z_{1-\alpha/2} \leq \frac{\bar{p} - p}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \leq z_{1-\alpha/2}] = 1 - \alpha \\ \Rightarrow\ &P[\hat{p} - \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}2.58 \leq p \leq \bar{p} + \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}2.58] = 0.99 \end{aligned}$$

Donc

$$\begin{aligned} I_{0.99}(p) &= [\bar{p} - \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}2.58; \bar{p} + \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}2.58] \\ &= [0.73 - 2.58 * 0.04; 0.73 + 2.58 * 0.04] = [\mathbf{0.63}; \mathbf{0.84}]. \end{aligned}$$

Soit $I_a$ l'incertitude absolue (le rayon de l'intervalle de confiance).

$$\begin{aligned} I_a &\leq 0.03 \\ \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}z_{1-\alpha/2} &\leq 0.03 \\ n &\geq \frac{z^2_{1-\alpha/2}}{(0.03)^2} * 0.1971 \\ n &\geq \mathbf{1}458. \end{aligned}$$

**N. B.** *Il faut faire la différence entre le $\bar{X}$ comme variable aléatoire (estimateur) dépendant de l'échantillon ( expérience aléatoire = échantillonnage) et $\bar{x}$ comme valeur (estimation) déterminée après échantillonnage.*

**Solution de l'Exercice 9**

1. Soit $X$ la v.a égale à l'ancienneté d'un employé.
   • Le cas $n = 50$ est le cas des grands échantillons alors
   $$\mathrm{I}_{0.9}(\mu) = \left[\bar{x} \pm z_{0.95}\frac{s}{\sqrt{n}}\right] = [\mathbf{62.43}, \mathbf{73.57}];$$
   où $z_{0.95} = 1.64$ est le quantile d'ordre 0.95 d'une normale standard.
   • Le cas $n = 9$ est le cas des petits échantillons alors en supposant que $X \sim \mathcal{N}(\mu, \sigma^2)$ on a
   $$\mathrm{I}_{0.9}(\mu) = \left[\bar{x} \pm t_{8;0.95}\frac{s}{\sqrt{n}}\right] = [\mathbf{53.12}, \mathbf{82.88}];$$
   où $t_{8;0.95} = 1.86$ est le quantile d'ordre 0.95 de la loi de Student à 8 dll.
2. La fréquence empirique des employés satisfaits de la formation est $\bar{p} = 40/50 = 0.8$
   a) On a un grand ESAR de taille $n = 50 \geq 30$, alors
   $$\mathrm{I}_{0.9}(p) = \left[\bar{p} \pm z_{0.95}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right] = [\mathbf{70.73}\%, \mathbf{89.27}\%].$$
   b) Supposons maintenant que les employés sont choisis au hasard et sans remise.
      i – Non car le taux de sondage est $f = 50/200 > 5\%$.
      ii – L'erreur type dans le cas d'un ESSR est
      $$\sqrt{\frac{200-50}{200-1}} * \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.8681 * 0.0566 = \mathbf{0.0491}.$$
      L'intervalle de confiance à 90% de la proportion des employés satisfaits de la formation est
      $$\mathrm{I}_{0.9}(p) \approx [\bar{p} \pm z_{0.95} * 0.0491] \approx [\mathbf{73}\%, \mathbf{87}\%].$$