

TP Principes et Méthodes Statistiques

Gabriel Sarrazin, Nejmeddine Douma, Simon Rabourg

Avril 2015

1 Analyse des défauts de cuves

2 Vérifications expérimentales à base de simulations

1. Il est possible de simuler n échantillons de la loi $Pa(a,b)$ car nous connaissons sa fonction de répartition.

$P_a(a,b)$ est une fonction continue, elle peut donc s'apparenter à une loi uniforme. Dans un premier temps, simuler n échantillons de cette loi va donc consister à tirer, au hasard, n valeurs aléatoires sur l'intervalle $[0,1]$. Connaissant la fonction de répartition de la loi $P_a(a,b)$, nous allons ensuite calculer l'image inverse $F^{-1}(U_i)$ pour obtenir un échantillon de loi $P_a(a,b)$ et nous ferons cela pour les n valeurs obtenues sur $[0,1]$

Nous pouvons représenter cette méthode sous forme d'un graphique : en mettant en ordonnée les n valeurs de la loi U_i et en abscisse la projection pour chacune de ses valeurs de son image inverse ($F^{-1}(U_i)$).

2. En suivant la méthode décrite précédemment, nous avons simulé m échantillon de taille n avec différentes valeurs pour m, n et a .

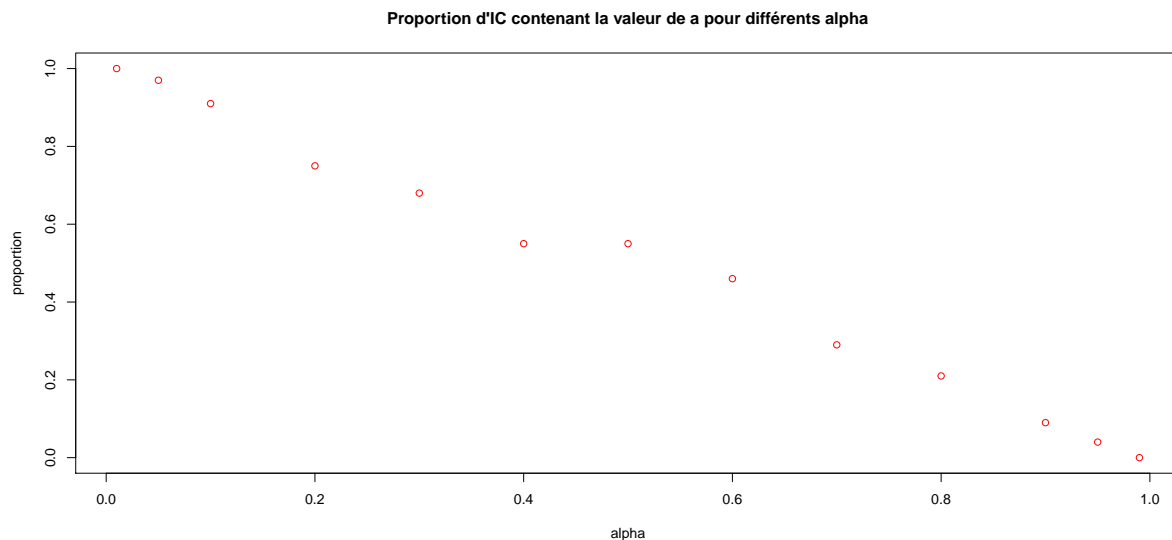


FIGURE 1 – Graphe des proportions d'IC(α) contenant la valeur exacte de a pour différentes valeurs de α

Nous avons ensuite, pour chaque échantillon de taille n , calculer l'intervalle de confiance bilatéral. Pour cela nous avons utilisé l'intervalle de confiance trouvé en première partie qui s'utilise avec $Y = \ln \frac{X}{2}$. A chaque fois que a est bien contenu dans cet intervalle, nous incrémentons une variable compteur. La proportion Pr d'IC contenant a est donc $Pr = \text{Compteur}/m$.

Afin que cela soit plus représentatif, nous avons tracé un graphique avec les différentes proportions obtenues en fonction des α . (Voir figure page 2)

Nous pouvons voir qu'il existe une linéarité entre la proportion et les α . Plus on augmente le nombre d'échantillon m et leur taille n et plus l'approximation est exacte. Ces points ont été obtenus pour les valeurs de $m = 100$, $n = 10, 30, 50, 100, 200, 500$, $a = 3, 5, 10, 20, 50$ et $\alpha = 0.01, 0.05, 0.10, 0.20, \dots, 0.80, 0.90, 0.95, 0.99$.

Quand on simule un grand nombre m d'échantillons de taille n de la loi $P_a(a, 2)$ alors la proportion d'IC(α) contenant a est approximativement égale à $1 - \alpha$.

3. Afin d'estimer le paramètre a , nous disposons de trois méthodes : la méthode des moments, la méthode du maximum de vraisemblance qui nous mènent au même résultat pour l'estimation du paramètre a et nous pouvons aussi déterminer l'estimateur sans biais de variance minimale si celui-ci existe.

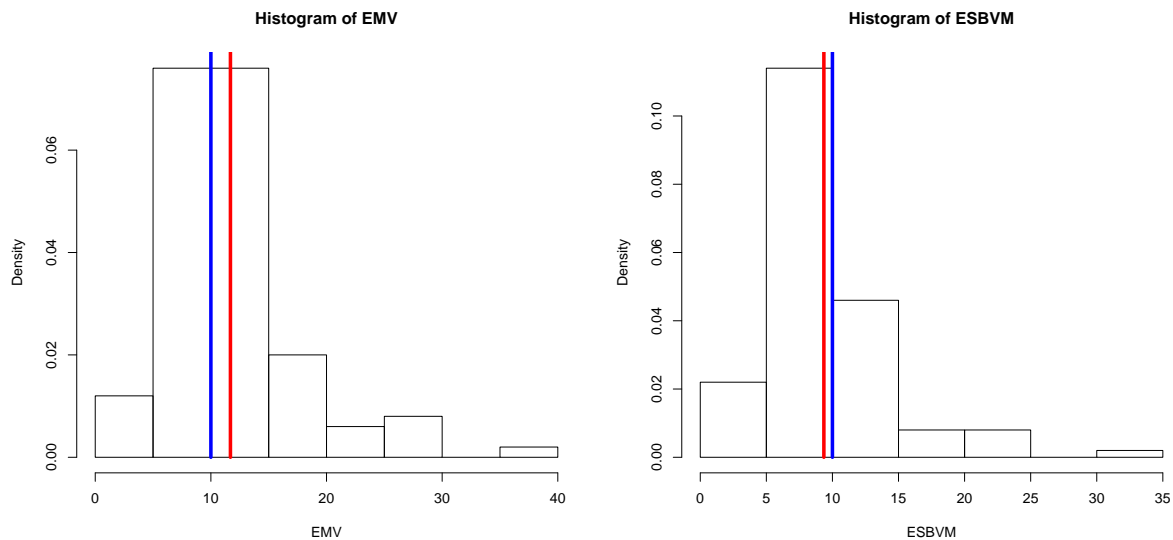


FIGURE 2 – Histogramme des EMV et ESBVM pour $m = 100, n = 5, a = 10$

Pour chaque échantillon (m en tout) nous mis au point sur R une fonction qui calcule les estimateurs proposés, qui estime le biais et l'erreur quadratique de chaque estimateur, et qui fait une moyenne pour chacun de ces résultats. Par ces résultats, nous pouvons tracer un histogramme des estimateurs obtenus avec en rouge : la moyenne de ces estimateurs et en bleu : la valeur exacte de a pour laquelle il faut être la plus proche.

Pour chaque graphique voici les moyennes des biais et des EQM obtenus pour chaque estimateur :

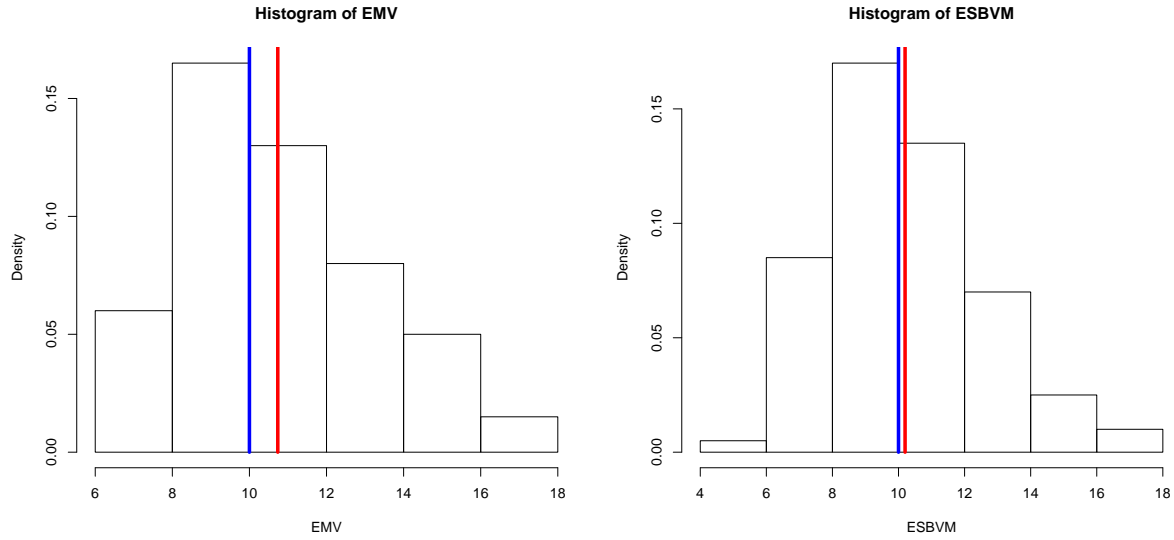


FIGURE 3 – Histogramme des EMV et ESBVM pour $m = 100, n = 20, a = 10$

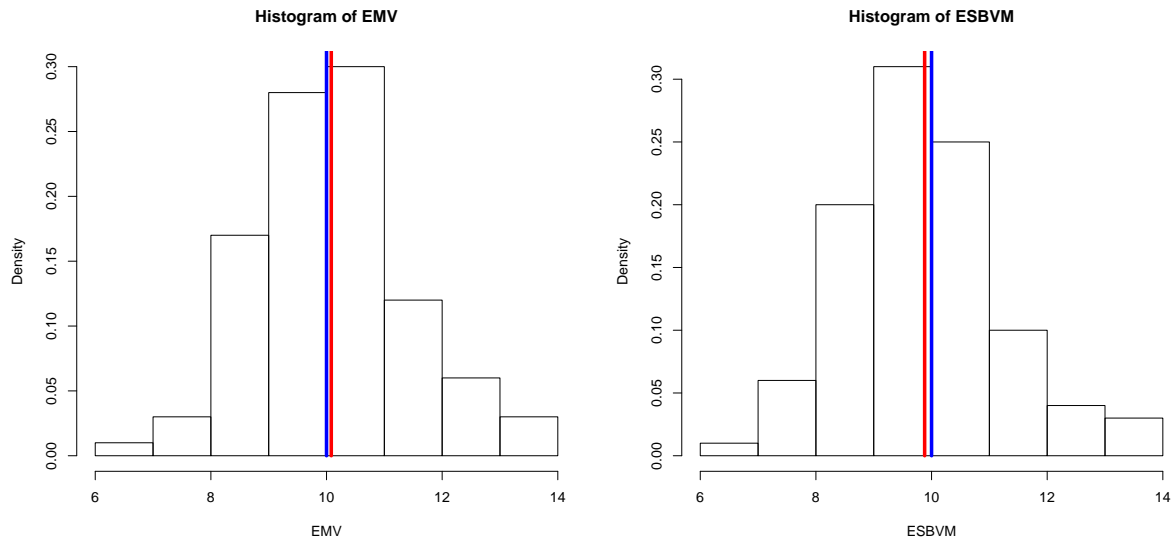


FIGURE 4 – Histogramme des EMV et ESBVM pour $m = 100, n = 50, a = 10$

- Moyenne des biais EMV 1.691318 Moyenne des biais ESBVM -0.6469459 Moyenne des EQM EMV 38.3953 Moyenne des EQM ESBVM 23.16078
- Moyenne des biais EMV 0.7355605 Moyenne des biais ESBVM 0.1987825 Moyenne des EQM EMV 6.691049 Moyenne des EQM ESBVM 5.589889
- Moyenne des biais EMV 0.08228447 Moyenne des biais ESBVM -0.1193612 Moyenne des EQM EMV 1.867297 Moyenne des EQM ESBVM 1.801097
- Moyenne des biais EMV 0.1657237 Moyenne des biais ESBVM 0.06406643 Moyenne des EQM EMV 0.8215394 Moyenne des EQM ESBVM 0.7823774
- Moyenne des biais EMV 0.02238977 Moyenne des biais ESBVM 0.002344989 Moyenne des EQM EMV 0.207069 Moyenne des EQM ESBVM 0.2057478

Après analyse des résultats et des graphiques, nous pouvons en déduire que le meilleur estimateur est l'*ESBVM* car il possède le biais et l'erreur quadratique

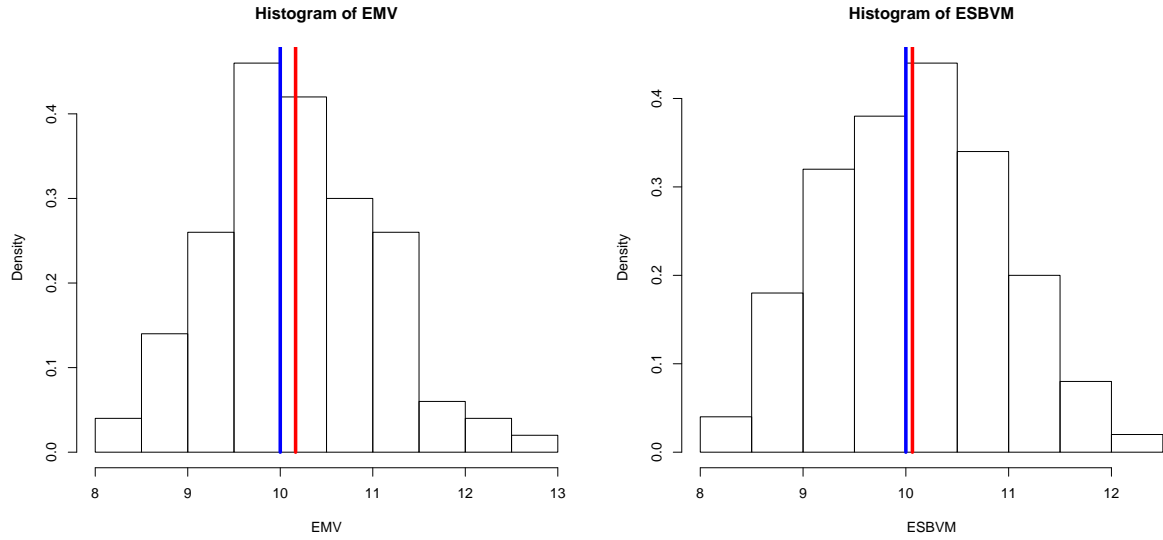


FIGURE 5 – Histogramme des EMV et ESBVM pour $m = 100, n = 100, a = 10$

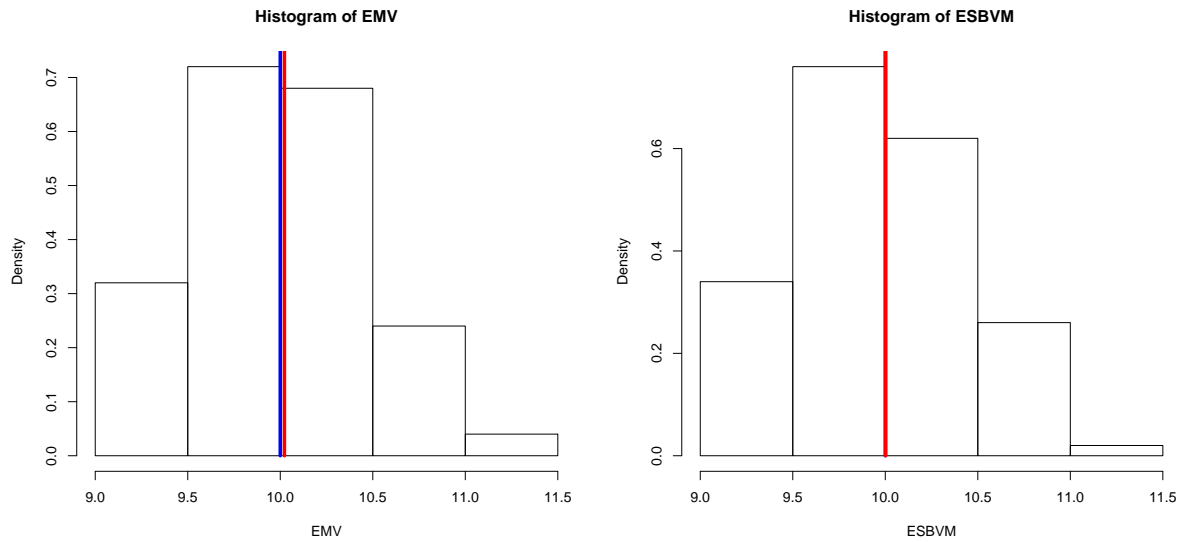


FIGURE 6 – Histogramme des EMV et ESBVM pour $m = 500, n = 5, a = 10$

moyenne le plus faible sur l'ensemble des échantillons. Il faut noter que plus la taille des échantillons est élevée, plus les estimateurs sont précis. Nous pouvons tout de même constater que quelque soit la taille de l'échantillon, l'*ESBVM* est le meilleur estimateur.

4. Dans cette partie nous simulons à nouveau m échantillons de taille n suivant la loi $P_a(a, 2)$. Nous calculons la moyenne empirique à l'aide de la fonction *mean* disponible sur R et nous calculons son Esperance. Pour chaque valeur n allant de 5 à 500 nous calculons la différence absolue de ces deux paramètres. Nous avons ensuite enregistré le nombre de fois où la valeur dépasse une valeur ϵ (*Ici*, $\epsilon = 1$). Sur la figure qui suit, nous pouvons remarquer que plus la taille de l'échantillon grandit, moins la différence entre la moyenne et l'esperance de cet échantillon est grande.

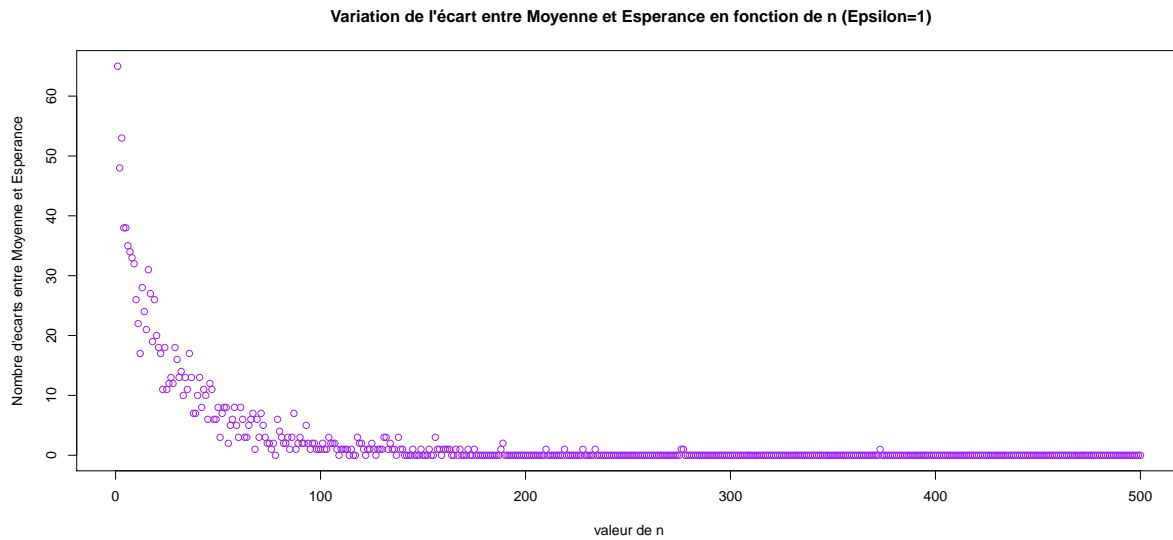


FIGURE 7 – Variation de la différence absolue entre Moyenne et Esperance en fonction de n

Par conséquent, plus n est grand, moins la moyenne empirique \bar{X}_n s'éloigne de l'Esperance $E(X)$ d'au moins ϵ . On a donc

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P} \left(\left| \frac{X_1 + X_2 + \dots + X_n}{n} - E(X) \right| \geq \epsilon \right) = 0$$

Autrement dit, (X_n) converge en probabilité vers $E(X)$. La moyenne empirique est bien un estimateur convergent de l'esperance.

5. Après avoir simulé m échantillons de taille n suivant la loi $P_a(a, 2)$, nous calculons leur moyenne. Nous obtenons donc un échantillon de moyennes empiriques. Pour différentes valeurs de n (Voir figures), nous avons tracé un histogramme et un graphe de la probabilités pour la loi normale à l'aide de la fonction `R qqnorm` qui permet de comparer graphiquement la distribution de l'échantillon des m moyennes empiriques avec une distribution normale.

Nous pouvons constater que pour $n = 5$ il est difficile de dire que la loi normale est un modèle approprié. Cependant, plus n augmente et plus les points sont alignés. Nous pouvons donc en déduire que pour n "suffisamment" grand, la loi de \bar{X}_n est approximativement une loi normale.

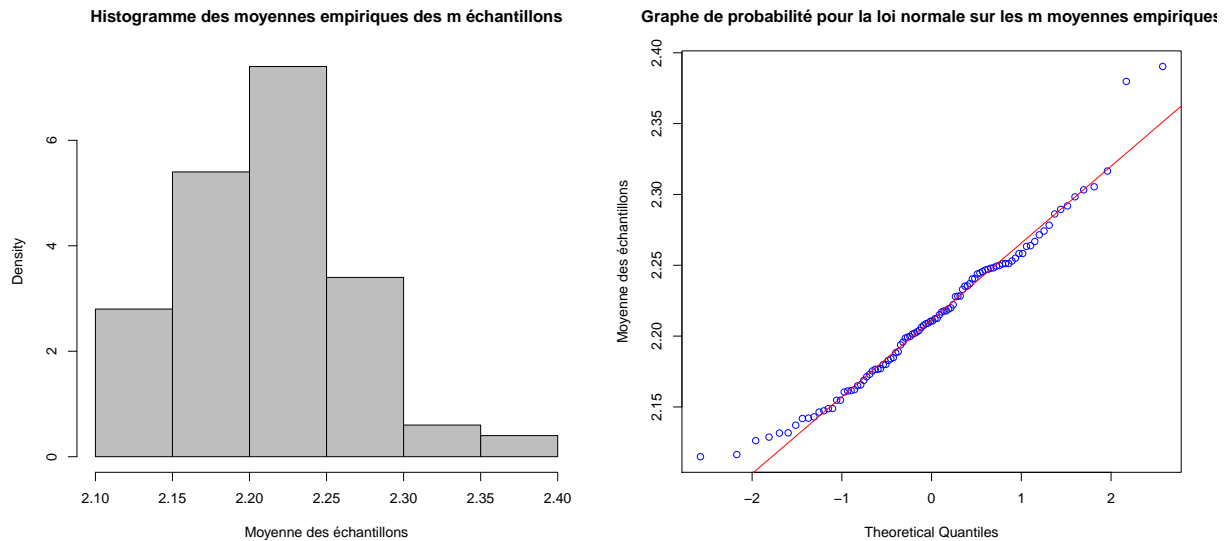


FIGURE 8 – Etude de la distribution \bar{X}_n par une distribution normale pour $n=10$

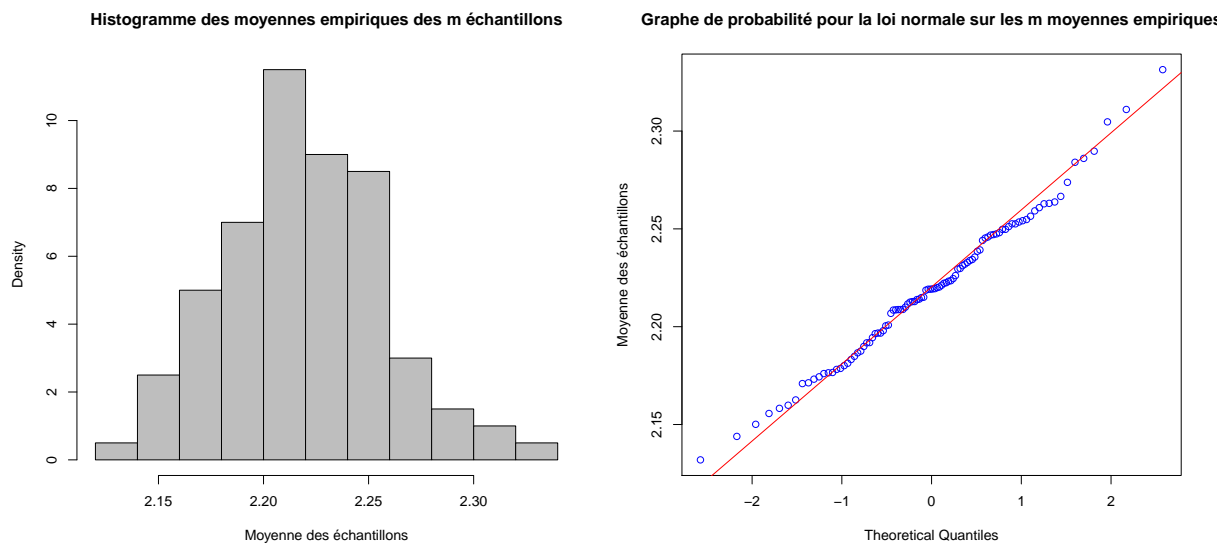


FIGURE 9 – Etude de la distribution \bar{X}_n par une distribution normale pour $n=50$

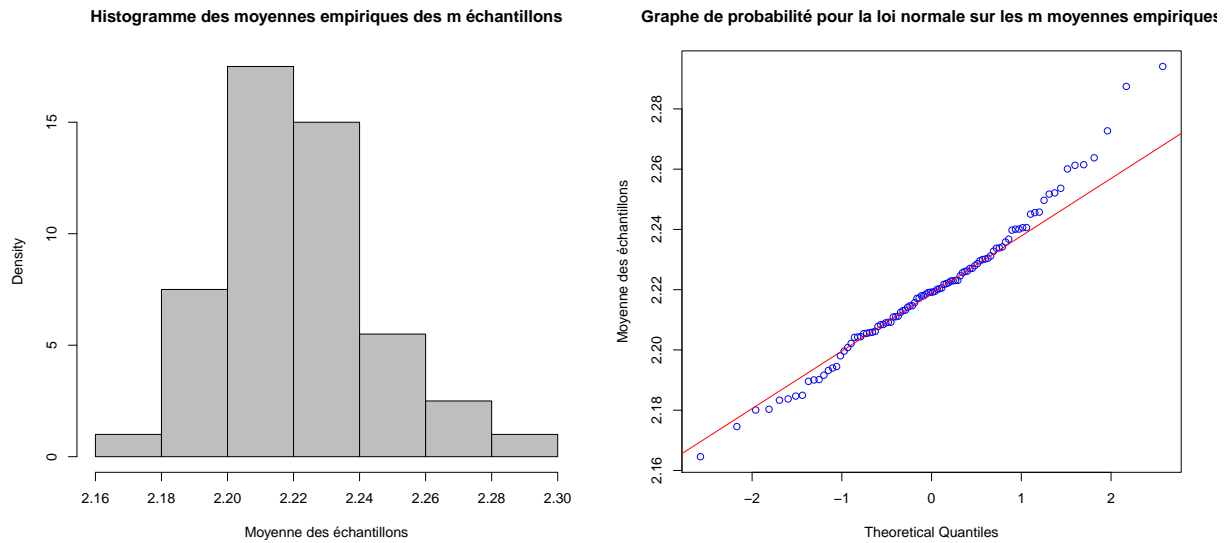


FIGURE 10 – Etude de la distribution \bar{X}_n par une distribution normale pour $n=100$

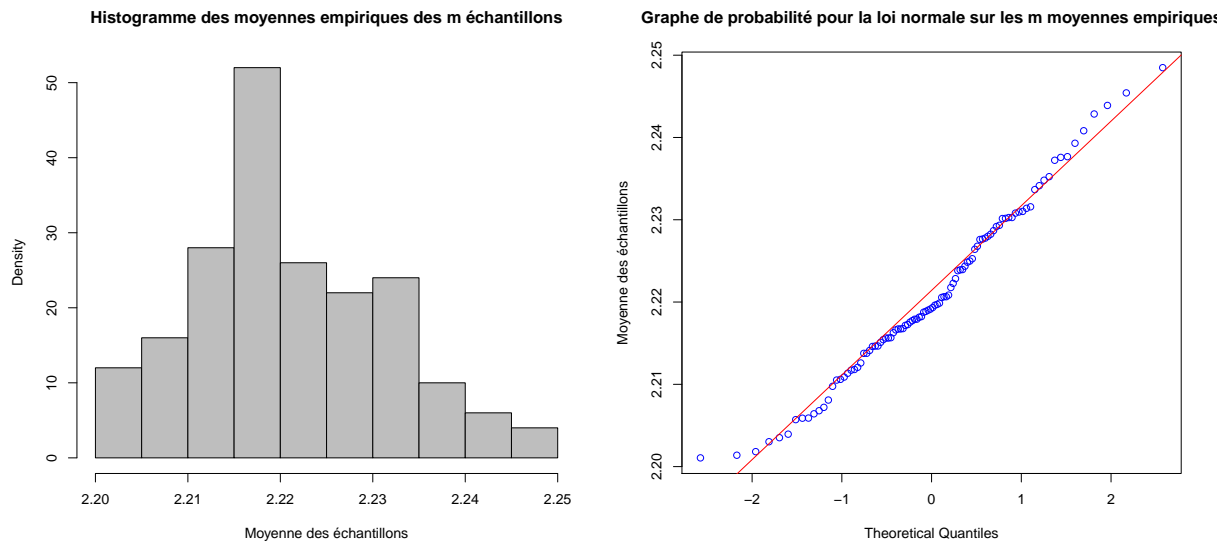


FIGURE 11 – Etude de la distribution \bar{X}_n par une distribution normale pour $n=500$

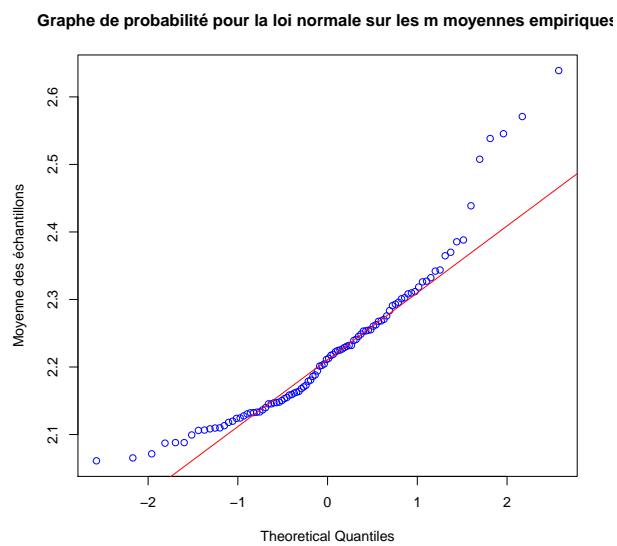
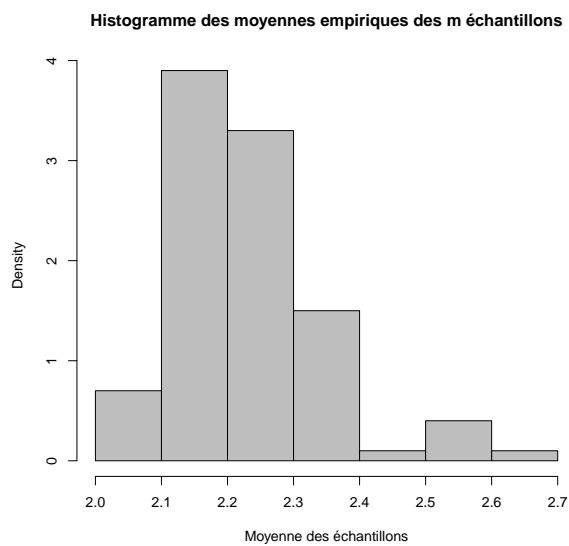


FIGURE 12 – Etude de la distribution \bar{X}_n par une distribution normale pour $n=5$