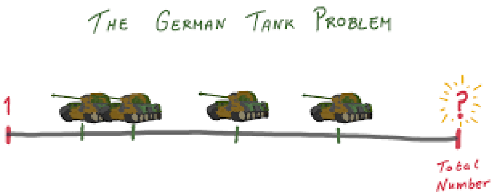
ECE 2 - Année 2021-2022
ENC Bessières
Mathématiques - F. Gaunard
http://frederic.gaunard.com



Devoir Maison n°7

German Tank Problem
Solution

On considère dans tout le problème un $n-$échantillon $(X_1, ..., X_n)$ de la loi $\mathcal{U}([\![1, N]\!])$ et on commence par poser

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(1) On commence par rappeler que l'espérance et la variance de chaque variable de l'échantillon vaut

$$E(X_i) = \frac{N+1}{2}, \qquad V(X_i) = \frac{N^2-1}{12}$$

ainsi, par linéarité de l'espérance

$$(\overline{X}_n) = \frac{E(X_1) + ... + E(X_n)}{n} = E(X_1) = \frac{N+1}{2}$$

ce qui n'est bien évidemment pas égal à N. Mais, en posant

$$T_n = 2\overline{X}_n - 1 = \frac{2}{n}(X_1 + ... + X_n) - 1,$$

la linéarité de l'espérance garantit que $E(T_n) = N$ et T_n est, lui, un estimateur sans biais de N.

(2) Le risqua quadratique est égal à la variance lorsque l'estimateur est sans biais. Ainsi,

$$\begin{aligned}
r(T_n) &= V(T_n) = V\left(\frac{2}{n}(X_1 + ... + X_n) - 1\right) \\
&= \frac{4}{n^2} V(X_1 + ... + X_n) \\
&= \frac{4}{n^2}(V(X_1) + .. + V(X_n)) && \text{par indépendance des } X_i \\
&= \frac{N^2-1}{3n}
\end{aligned}$$

En particulier, on a bien $r(T_n) \to 0$ lorsque $n \to +\infty$, ce qui permet d'affirmer que T_n est un estimateur convergent d'après un résultat du cours (qui lui même découle de l'inégalité de Bienayme-Tchebychev).

(3) **Application numérique &** `SciLab`.

(a) Il suffit d'utiliser la commande `mean( )` qui renvoie la moyenne des composantes d'une liste ou d'un vecteur.

```
function y=T(X)
    y=2*mean(X)-1
endfunction
```

(b) On observe

```
X=[8, 322, 15, 135, 69]
```

La fonction précédente renvoie, pour cet échantillon, une estimation $N \simeq T_5 = 219$. Or on sait qu'il y a au moins 322 tanks. Cet estimateur ne semble pas très pertinent avec un petit échantillon...

(c) Cette fois on demande à ce que l'échantillon X soit simulé à l'intérieur de la fonction au sein de laquelle on utilise notre fonction `T` précédente.

```
function y=T_moy(N,n)
    Y=zeros(1,n)
    for k=1:n
        X=grand(1, 1000, 'uin', 1, N)
        Y(k)=T(X)
    end
    y=mean(Y)
endfunction
```

Regardons ce que renvoie donc cette fonction pour $N = 1000$ avec $n = 20$ puis $n = 100$.

```
--> T_moy(1000, 20)
 ans  =

   997.81770

--> T_moy(1000, 100)
 ans  =

   998.20482
```

L'espérance théorique étant $N = 1000$, ce n'est pas si mal.

(4) C'est une question classique. On rappelle que, pour chaque $i \in [\![1;n]\!]$, on a, pour $k \in \mathbb{N}$,

$$P(X_i \leq k) = \begin{cases} 0, & \text{si } k = 0 \\ \dfrac{k}{N}, & \text{si } 1 \leq k \leq N \\ 1, & \text{si } k > 1 \end{cases}$$

Étant clair que $P(M_n \leq k) = 0$ si $k = 0$ et $P(M_n \leq k) = 1$ si $k > N$, on calcule pour $1 \leq k \leq N$.

$$\begin{aligned} P(M_n \leq k) &= P\left(\max(X_1, ..., X_n) \leq k\right) \\ &= P\left([X_1 \leq k] \cap [X_2 \leq k] \cap ... \cap [X_n \leq k]\right) \\ &= P(X_1 \leq k) \times P(X_2 \leq k) \times ... \times P(X_n \leq k) \qquad \text{par indépendance des } X_i \\ &= \left(\frac{k}{N}\right)^n \end{aligned}$$

(5) Soit Y un v.a à valeurs dans $[\![1;N]\!]$. La variable étant finie, elle admet une espérance qu'on va calculer en utilisant l'égalité ultra-classique suivante

$$[Y = k] \cup [Y > k] = [Y > k-1]$$

et donc, par incompatibilité,

$$P(Y = k) = P(Y > k-1) - P(Y > k).$$

Il suit que

$$\begin{aligned}
E(Y) &= \sum_{k=1}^{N} kP(Y=k) = \sum_{k=1}^{N} k\left(P(Y>k-1)-P(Y>k)\right)\\
&= \sum_{k=1}^{N} kP(Y>k-1) - \sum_{k=1}^{N} kP(Y>k)\\
&= \sum_{k=0}^{N-1} (k+1)P(Y>k) - \sum_{k=1}^{N} kP(Y>k)\\
&= \sum_{k=0}^{N-1} kP(Y>k) + \sum_{k=0}^{N-1} P(Y>k) - \sum_{k=1}^{N} kP(Y>k)\\
&= \sum_{k=0}^{N-1} P(Y>k) - NP(Y>N)\\
&= \sum_{k=0}^{N-1} P(Y>k)
\end{aligned}$$

car $(Y>N)=0$ vu que $Y(\Omega)=[\![1;N]\!]$. C'est bien la formule demandée.

(6) Vu que $M_n(\Omega)=[\![1;N]\!]$, on peut lui appliquer le résultat précédent pour le calcul de son espérance:

$$\begin{aligned}
E(M_n) &= \sum_{k=0}^{N-1} P(M_n>k) = \sum_{k=0}^{N-1} \left(1-P(M_n\leq k)\right)\\
&= \sum_{k=0}^{N-1} 1 - \sum_{k=0}^{N-1}\left(\frac{k}{N}\right)^n\\
&= N - \sum_{k=0}^{N-1}\left(\frac{k}{N}\right)^n,
\end{aligned}$$

comme attendu.

(7) Soit $k\in[\![0;N-1]\!]$. La fonction $t\mapsto t^n$ est croissante sur $[k/N;(k+1)/N]$ et donc, pour tout $t\in[k/N;(k+1)/N]$, on a

$$t^n \geq \left(\frac{k}{N}\right)^n \geq 0$$

puis, par croissance de l'intégrale

$$\int_{k/N}^{(k+1)/N} t^n \mathrm{d}t \geq \int_{k/N}^{(k+1)/N}\left(\frac{k}{N}\right)^n \mathrm{d}t = \frac{1}{N}\times\left(\frac{k}{N}\right)^n.$$

On a bien

$$0\leq \left(\frac{k}{N}\right)^n \leq N\int_{k/N}^{(k+1)/N} t^n\mathrm{d}t.$$

(8) On somme, pour k de 0 à $N-1$, et on a

$$0\leq \sum_{k=0}^{N-1}\left(\frac{k}{N}\right)^n \leq \sum_{k=0}^{N-1} N\int_{k/N}^{(k+1)/N} t^n\mathrm{d}t = N\int_0^1 t^n\mathrm{d}t = \frac{N}{n+1}.$$

En appliquant le résultat de la Question (6), on a bien
$$N - \frac{N}{n+1} \leq E(M_n) \leq N.$$
Le théorème des gendarmes donne alors
$$\lim_{n\to+\infty} E(M_n) = N$$
ce qui est bien la définition d'un estimateur de N asymptotiquement sans biais.

On introduit
$$\tilde{M}_n = M_n + (m_n - 1), \qquad \text{où} \qquad m_n = \min(X_1, ..., X_n).$$

(9) On commence par écrire, pour $k \in [\![0, N-1]\!]$, que
$$\begin{aligned} P(m_n > k) &= P\left(\min(X_1, ..., X_n) > k\right) \\ &= P\left([X_1 > k] \cap ... \cap [X_n > k]\right) \\ &= P(X_1 > k) \times ... \times P(X_n > k) \qquad \text{par indépendance des } X_i \\ &= \left(\frac{N-k}{N}\right)^n \end{aligned}$$
et donc
$$E(m_n) = \sum_{k=0}^{N-1} P(m_n > k) = \sum_{k=0}^{N-1} \left(\frac{N-k}{N}\right)^n = \sum_{j=1}^{N} \left(\frac{j}{N}\right)^n.$$

(10) Par linéarité de l'espérance,
$$E(\tilde{M}_n) = E(M_n) + E(m_n) - 1 = N - \sum_{k=0}^{N-1} \left(\frac{k}{N}\right)^n + \sum_{j=1}^{N} \left(\frac{j}{N}\right)^n - 1 = N,$$
et $\tilde{M}_n$ est bien un estimateur sans biais de N.

(11) Avec

```
X=[14,   44,   50,   101,   117,   127,   134,   139,   165,   188,
192,  201,   204,  215,   234,   243,   244,   253,   269,   269,
282,  287,   288,  322,   345]
```

La commande `max(X)+min(X)-1` renvoie une estimation $N \sim \tilde{M}_{25} = 358$, ce qui est déjà plus grand que la valeur maximale de l'échantillon et n'a donc pas le défaut précédent.

(12) **Comparaison**.

La fonction proposée renvoie les moyennes de chacun des trois estimateurs T_n, M_n et $\tilde{M}_n$, sur 1000 échantillons de taille n de la loi $\mathcal{U}[\![1; N]\!]$.

On prend ensuite $N = 1000$.
On regarde comment évolue chaque estimateur en fonction de n.
Pour cela on représente graphiquement cette évolution en fonction de n en prenant des valeurs de n entre 10 et 200 avec un pas de 10 (donc $n = 10, 20, 30, ..., 100, 110, ..., 200$).
Les losanges correspondent à l'estimateur T_n, les $+$ à l'estimateur M_n et les $\times$ à $\tilde{M}_n$. Il est clair que ce sont les $\times$ qui ont des valeurs au plus proche de $N = 1000$ et on peut penser raisonnablement que $\tilde{M}_n$ est le meilleur des trois estimateurs.