Estimateurs

Exercice 1

Les données se trouvant à la fin de l'exercice sont des tirages aléatoires de valeurs suivant une loi uniforme $\mathcal{U}_{[0;\vartheta]}$ pour un certain paramètre réel $\vartheta > 0$ que l'on souhaite estimer.

Les questions marquées d'un astérisque sont facultatives et on peut admettre ce qu'elles demandent.

Estimation ponctuelle. Soit $n \in \mathbb{N}_{>0}$ et $(X_1, \dots, X_n)$ un échantillon de $\mathcal{U}_{[0;\vartheta]}$

0. * Montrer que pour tout réel $\alpha \neq -1$, $\mathbb{E}\left(X_1^\alpha\right) = \dfrac{\vartheta^\alpha}{\alpha+1}$.
1. On considère $T_n^{(1)} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} 2X_i$.
   (a) Montrer que $T_n^{(1)}$ est un estimateur convergent et sans biais de ϑ.
   (b) Calculer $\mathbb{V}\left(T_n^{(1)}\right)$ la variance de $T_n^{(1)}$.
   (c) En déduire $EQ(T_n^{(1)}, \vartheta)$ l'erreur quadratique de l'estimateur.
2. On considère $T_n^{(2)} = e\displaystyle\prod_{i=1}^{n} X_i^{\frac{1}{n}}$
   (a) Montrons que $T_n^{(2)}$ est un estimateur convergent de ϑ.
      i. * Montrer que $\mathbb{E}\left(\ln(X_1)\right) = \ln(\vartheta) - 1$.
      ii. En déduire que $L_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \ln(X_i)$ est un estimateur convergent de $\ln(\vartheta) - 1$.
      iii. Conclure.
   (b) Déterminons $B(T_n^{(2)}, \vartheta)$ le biais de l'estimateur.
      i. Montrer que $\mathbb{E}\left(T_n^{(2)}\right) = e\left(\dfrac{n}{n+1}\right)^n \vartheta$. En déduire $B(T_n^{(2)}, \vartheta)$ le biais de $T_n^{(2)}$.
      ii. * Montrer $\left(\dfrac{n}{n+\alpha}\right)^n \sim e^{-\alpha + \frac{\alpha^2}{2n}}$
      iii. En déduire que $T_n^{(2)}$ est asymptotiquement sans biais.
      iv. * Montrer que $B(T_n^{(2)}, \vartheta) \sim \dfrac{\vartheta}{2n}$.
   (c) Calculer $\mathbb{V}\left(T_n^{(2)}\right)$ la variance de $T_n^{(2)}$.
   (d) * En déduire que $EQ(T_n^{(2)}, \vartheta)$ l'erreur quadratique de l'estimateur, est équivalent en l'infini à $\dfrac{\vartheta^2}{n}$.
3. **Application :** Avec le premier échantillon de données simulées, estimer la valeur de ϑ via $T_n^{(1)}$ et $T_n^{(2)}$ et donner l'erreur quadratique de ces estimations. Faire la même chose avec le second échantillon. Lequel de ces deux estimateurs est le meilleur ?
4. On considère $T_n^{(3)} = \max(X_i | i \in [1, n])$.
   (a) * Montrer que pour tout $n \in \mathbb{N}_{>0}$ et tout $k \in \mathbb{N}$, $\displaystyle\int_{[0;\vartheta]^n} \max(x_1, \dots, x_n)^k \, dx_1 \cdots dx_n = \dfrac{n}{n+k}\vartheta^{n+k}$.
   (b) * Montrer que la fonction de répartition de $T_n^{(3)}$ est

$$\forall t \in \mathbb{R},\ F_{T_n^{(3)}}(t) = \frac{t^n}{\vartheta^n}\mathbb{1}_{[0,\vartheta]}(t) + \mathbb{1}_{]\vartheta;+\infty[}(t)$$

   (c) Montrer que $T_n^{(3)}$ est un estimateur convergent de ϑ asymptotiquement sans biais.
   (d) Déterminer $EQ(T_n^{(3)}, \vartheta)$ et montrer que qu'il est équivalent en l'infini à $\dfrac{2\vartheta}{n^2}$.
5. On considère $T_n^{(4)} = \dfrac{n+1}{n}\max(X_i | i \in [1, n])$

(a) Montrer que T_n^4 est un estimateur convergent et sans biais de ϑ.

(b) Déterminer $\mathbb{V}\left(T_n^{(4)}\right)$ et $\mathrm{EQ}(T_n^4, \vartheta)$.

6. **Application :** Avec le premier échantillon de données simulées, estimer la valeur de ϑ via $T_n^{(3)}$ et $T_n^{(4)}$ et donner l'erreur quadratique de ces estimations. Faire la même chose avec le second échantillon. Classer les 4 estimateurs de ϑ du moins bon au meilleur.

Intervalle de confiance. On note $T_n = T_n^{(4)}$ le meilleur estimateur étudié dans la première partie de l'exercice.

1. Montrer que $Q_{T_n}(t)$ la fonction quantile de T_n est définie pour tout $t \in [0, 1]$ par

$$Q_{T_n}(t) = \vartheta t^{\frac{1}{n}}$$

2. On admettra que l'intervalle de dispersion unilatéral supérieur de niveau $1 - \alpha$ est optimal.

(a) **(Application).** Déterminer en fonction de ϑ les intervalles de dispersions optimaux de niveau 95%, de niveau 99% et de niveau 99, 9% pour les deux jeux de données de l'exercice.

(b) Montrer que $\left[T_n; T_n \alpha^{\frac{1}{n}}\right]$ est un intervalle de confiance de niveau $1 - \alpha$.

(c) **(Application).** Déterminer les intervalles de confiance de niveau 95%, 99% et 99, 9% des deux jeux de données de l'exercice.

Seul au monde. Voici des données correspondant au tirage aléatoire de valeur suivant une certaine loi de probabilité inconnue.

$\rightarrow$ Données :

−1.237	−1.444	+0.804	+0.843	+0.189	−2.705	−2.224	−1.904	−1.299	−1.426
−2.307	+0.526	+0.009	+0.537	+0.051	−2.679	−0.043	−1.356	−1.398	+0.038
+0.459	+0.449	−0.319	−2.020	+0.698	+0.464	−0.542	−1.867	−2.327	−0.469
−1.799	−0.443	−2.682	−1.736	−1.040	+0.092	−2.439	+0.315	+0.139	−0.419
−0.004	−1.862	−1.886	+0.024	−1.350	−1.276	−2.271	−1.889	−0.689	−0.964

$\rightarrow$ Taille de l'échantillon : 50

$\rightarrow$ $\overline{x} = \dfrac{1}{50}\displaystyle\sum_{i=1}^{50} x_i = -0.89356$

$\rightarrow$ $\overline{v} = \dfrac{1}{50}\displaystyle\sum_{i=1}^{50} (x_i - \overline{x})^2 = 1.1569419264$

Premier modèle : loi normale. On suppose que les valeurs tirées suivent une loi normale $\mathcal{N}(\mu, \sigma)$.

Déterminer, en vous servant du cours, les intervalles de confiances symétrique de niveau 95%, 99% et 99, 9% de μ et σ.

Second modèle : loi uniforme. On suppose que les valeurs suivent une loi uniforme $\mathcal{U}_{[\vartheta;1]}$ pour un certain paramètre réel $\vartheta < 1$.

Donner au moins deux estimateurs convergent de ϑ dont l'un sera sans biais et l'autre asymptotiquement sans biais. Vous comparerez la qualité de ces estimateurs et déterminerez pour le meilleur les intervalles de confiances de niveau 95%, 99% et 99, 9%.

Echantillon 1 .

→ Données :

0.178	0.573	1.725	2.930	1.197	1.251	3.000	0.262	2.608	0.926	0.968	1.587	2.383	1.851	0.417
3.051	1.656	0.639	0.995	2.764	1.975	2.812	1.602	2.118	1.132	3.016	2.659	0.549	0.446	0.636

→ Taille de l'échantillon : 30

→ $\frac{1}{30}\sum_{i=1}^{30} 2x_i = 3.193733333333333$

→ $e\prod_{i=1}^{30} x_i^{\frac{1}{30}} = 3.4197348436952573$

→ $max(x_i | i \in [\![1, 30]\!]) = 3.051$

Echantillon 2 .

→ Données :

0.596	2.089	0.117	0.783	2.814	3.085	0.556	1.938	2.079	2.071	2.552	1.640	0.656	1.096	0.267
1.165	0.680	0.775	0.091	0.887	0.210	0.832	2.444	2.325	1.471	0.605	0.009	2.388	0.912	2.942
2.851	0.170	1.025	2.403	0.755	2.402	0.330	3.026	2.254	1.845	0.147	0.326	2.040	3.016	2.982
1.967	0.813	0.603	2.425	0.782	1.961	0.993	0.802	1.504	1.906	0.275	2.167	3.065	0.131	0.135
1.261	0.830	2.127	2.763	2.753	0.178	0.218	0.293	2.254	1.770	1.445	1.540	1.554	2.865	0.119
0.848	1.914	0.964	0.958	0.241	0.659	1.882	2.681	2.645	2.334	1.399	0.024	2.236	0.912	1.870
2.541	1.623	2.403	0.912	1.849	1.608	0.144	1.181	1.667	0.658	1.159	1.277	1.824	0.462	0.172
2.098	2.435	3.003	0.049	0.362	2.565	1.052	0.153	2.668	3.098	1.819	2.515	1.385	2.475	2.495
2.425	1.413	0.917	2.630	1.050	2.262	0.168	1.093	0.587	0.461	2.425	3.072	2.575	2.595	1.036
0.478	2.979	0.876	0.344	1.383	0.141	3.023	1.399	1.407	0.894	2.579	2.414	2.870	0.014	2.827
1.258	2.776	3.047	0.504	0.650	2.485	1.677	1.059	2.263	1.698	1.273	0.625	1.954	2.019	2.187
0.154	2.541	2.821	0.385	0.322	2.764	1.628	2.893	0.598	2.587	2.162	2.553	2.205	2.489	0.573
0.849	1.842	2.993	0.371	0.143	0.378	3.031	1.975	1.815	1.874	0.068	2.845	0.731	1.928	0.962
1.907	1.962	0.535	0.478	3.079	1.024	0.185	0.389	1.093	0.133	2.646	0.555	1.752	2.907	0.649
2.049	1.324	2.462	1.946	2.670	0.368	2.801	1.928	2.087	0.811	0.832	1.871	0.254	0.387	0.623
0.072	2.303	0.762	1.046	1.696	2.638	1.901	0.106	2.429	2.439	0.011	1.470	0.081	1.049	1.212
1.643	0.670	2.779	1.957	2.687	3.122	1.699	1.272	2.468	0.208	1.452	0.708	0.477	1.920	2.821
0.892	2.615	2.268	0.163	2.436	0.790	0.073	0.808	1.483	0.097	1.834	1.836	1.259	2.246	0.641
2.303	2.846	2.589	2.856	0.011	2.240	0.887	1.572	2.254	0.561	2.467	0.881	1.167	3.130	0.243
0.323	1.010	1.584	2.904	2.996	2.497	2.010	0.679	2.281	2.289	2.189	0.548	1.396	1.846	2.798

→ Taille de l'échantillon : 300

→ $\frac{1}{300}\sum_{i=1}^{300} 2x_i = 3.0550666666666664$

→ $e\prod_{i=1}^{300} x_i^{\frac{1}{300}} = 2.8652098608233163$

→ $max(x_i | i \in [\![1, 300]\!]) = 3.13$