És ben conegut que, si $X_1, \ldots, X_n$ és una mostra aleatòria simple provinent d'una distribució normal de mitjana μ i variància σ^2 desconegudes, $N(\mu, \sigma^2)$, l'estadístic pivotal:

$$V = \frac{(n-1)\hat{S}^2}{\sigma^2}$$

segueix una distribució khi-quadrat amb $n - 1$ graus de llibertat, on

$$\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Al fitxer "exer1-2016.R" hi ha un conjunt de dades format per $n = 114$ valors independents, corresponents a temps d'espera fins a l'ocurrència d'un determinat esdeveniment.

És fa molt difícil creure que aquests valors corresponen a realitzacions d'un distribució normal.

1) Mitjançant bootstrap no paramètric, estima gràficament (un histograma, o una estimació suavitzada de la funció de densitat, és suficient) la distribució mostral de V. Compara gràficament aquesta estimació bootstrap amb una khi-quadrat amb $n - 1$ graus de llibertat, i amb la que seria la veritable distribució de V si les dades provinguessin d'un distribució exponencial (la informació per disposar d'aquesta darrera distribució mostral és al fitxer "trueDens.RDa", com accedir-hi i dibuixar-la s'indica a "exer1-2016.R"[1]).
2) Utilitzant l'estimació bootstrap anterior de la distribució mostral de V, crea una taula de quantils de V pels valors de probabilitat 0.01, 0.025, 0.05, 0.1, 0.9, 0.95, 0.975 i 0.99.
3) i 4) Repeteix els passos 1) i 2) però amb bootstrap paramètric, utilitzant el fet que, molt versemblantment, la distribució d'aquests temps d'espera és exponencial.
5) Calcula els intervals de confiança percentil i BCa mitjançant bootstrap no paramètric per a la variància σ^2. (Atenció, recorda que V NO és un estimador de σ^2.)
6) Fes el mateix que a la pregunta anterior, però mitjançant bootstrap paramètric.
7) Utilitzant el caràcter pivotal de V, calcula un interval de confiança bootstrap (versió no paramètrica i paramètrica exponencial) per σ^2.

Atenció: la pregunta anterior **NO** demana un interval de confiança bootstrap-t, que hauria d'estar basat en el possible caràcter pivotal de $t = \left(S^2 - \sigma^2\right) / \mathrm{SE}_{S^2}$ –encara que entendre com es construeixen els intervals de confiança bootstrap-t pot ser de gran ajuda per saber resoldre això. Pista: recordeu que, pel cas de dades normals, V segueix una distribució khi-quadrat amb $n - 1$ graus de llibertat. Aleshores, hi ha constants diguem-ne $\chi^2_{\frac{\alpha}{2}}$ i $\chi^2_{1-\frac{\alpha}{2}}$ (els quantils $\alpha/2$ i $1 - \alpha/2$ d'aquesta distribució) tals que:

$$\Pr\left\{\chi^2_{\frac{\alpha}{2}} \le \frac{(n-1)\hat{S}^2}{\sigma^2} \le \chi^2_{1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

[1] Aquesta distribució ha estat obtinguda per simulació. Com un exercici complementari, que milloraria la valoració d'aquest exercici de bootstrap, pots fer una simulació de Monte Carlo per demostrar que, efectivament, si les dades fossin exponencials, s'obtindria la distribució que es pot carregar a partir del fitxer trueDens.RDa, i que l'estadístic V també té caràcter pivotal sota la distribució exponencial.

i aïllant σ^2, tenim que:

$$\Pr\left\{\frac{(n-1)\hat{S}^2}{\chi^2_{1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)\hat{S}^2}{\chi^2_{\frac{\alpha}{2}}}\right\} = 1-\alpha.$$

Per tant, $\left[(n-1)\hat{S}^2 \Big/ \chi^2_{1-\frac{\alpha}{2}}, (n-1)\hat{S}^2 \Big/ \chi^2_{\frac{\alpha}{2}}\right]$ és un interval de confiança de nivell $1-\alpha$ per σ^2 (pel cas de dades normals). Del raonament anterior per obtenir aquest interval de confiança, què continua sent vàlid si no assumim normalitat?

8) Amb les mateixes dades utilitzades a les qüestions anteriors, simula B = 20000 remostres bootstrap SEMIPARAMÈTRIC a partir d'una estimació nucli gaussià de la funció de densitat i obté a partir d'elles l'interval de confiança bootstrap percentil.

Nota: amb la presentació teòrica i amb l'exemple pràctic a la carpeta "Estimació no paramètrica de la densitat", juntament amb les explicacions de classe, heu d'estar perfectament en condicions de resoldre aquesta darrera qüestió. Com a bibliografia complementària podeu també consultar el document: "No paramètrica - Mètodes de suavitzat (Prof. P. Delicado)" també al Campus virtual.

A lliurar com a resultat d'aquesta pràctica:

a. Script R comentat amb el codi que realitza tot allò demanat anteriorment.
b. Document Pdf o Word explicant els resultats obtinguts.

Si aquests documents s'han generat amb Sweave, knitr o algun sistema similar, juntament amb els fitxers font (per exemple, els fitxers amb extensió Rmd o Rnw) heu de lliurar els fitxers R i Pdf o Word que es generin.