Inferencia Estadística
Maestría en Análisis Estadístico y Computación
Tarea 4

Fecha de entrega: Jueves 12 de marzo, 2020

Problemas a entregar: 1, 3, 4, 6, 7, 8, 9, 10

1. **Wasserman, Cap 7.** Let $X_1, \ldots, X_n \sim$ Bernoulli (p) and let $Y_1, \ldots, Y_m \sim$ Bernoulli (q). Find the plug-in estimator and estimated standard error for p. Find an approximate 90 percent confidence interval for p (Use el intervalo de Wald (Aproximación Normal, expresión (7.6) del libro)). Find the plug-in estimator and estimated standard error for $p - q$. Find an approximate 90 percent confidence interval for $p - q$ (También usando la aproximación Normal).

2. **Wasserman, Cap 7.** (Computer Experiment.) Generate 100 observations from a $\mathrm{N}(0, 1)$ distribution. Compute a 95 percent confidence band for the CDF F (as described in the appendix). Repeat this 1000 times and see how often the confidence band contains the true distribution function. Pueden usar las funciones rnorm de R.

3. **Wasserman, Cap 7.** Let $X_1, \ldots, X_n \sim F$ and let $\widehat{F}$ be the empirical distribution function. Let $a < b$ be fixed numbers and define $\theta = T(F) = F(b) - F(a)$. Let $\widehat{\theta} = T\left(\widehat{F}_n\right) = \widehat{F}_n(b) - \widehat{F}_n(a)$. Find the estimated standard error of $\widehat{\theta}$ Find an expression for an approximate (usando la aproximación Normal) $1 - \alpha$ confidence interval for θ.

4. **Wasserman, Cap 7.** Get the data on eruption times and waiting times between eruptions of the Old Faithful geyser from the website. Estimate the mean waiting time and give a standard error for the estimate. Also, give a 90 percent confidence interval for the mean waiting time. Now estimate the median waiting time. In the next chapter we will see how to get the standard error for the median.

5. **Wasserman, Cap 7.** 100 people are given a standard antibiotic to treat an infection and another 100 are given a new antibiotic. In the first group, 90 people recover; in the second group, 85 people recover. Let p_1 be the probability of recovery under the standard treatment and let p_2 be the probability of recovery under the new treatment. We are interested in estimating $\theta = p_1 - p_2$. Provide an estimate, standard error, an 80 percent confidence interval, and a 95 percent confidence interval for θ.

6. **Wasserman, Cap 8.** (Computer Experiment.) Conduct a simulation to compare the various bootstrap confidence interval methods. Let $n = 50$ and let $T(F) = \int (x - \mu)^3 dF(x)/\sigma^3$ be the skewness. Draw $Y_1, \ldots, Y_n \sim N(0, 1)$ and set $X_i = e^{Y_i}, i = 1, \ldots, n$. Construct the three types of bootstrap 95 percent intervals (Intervalo Normal, Intervalo Pivotal, Intervalo Percentil) for $T(F)$ from the data $X_1, \ldots, X_n$. Repeat this whole thing many times (mil veces cada intervalo)and estimate the true coverage of the three intervals. Esto es, de las mil simulaciones, calcule el porcentaje de veces que los intervalos producidos contiene el verdadero valor $T(F)$.

7. **Wasserman, Cap 8.** (Computer Experiment.) Let $X_1, \ldots, X_n$ Normal $(\mu, 1)$. Let $\theta = e^{\mu}$ and let $\widehat{\theta} = e^{\bar{X}}$. Create a data set (using $\mu = 5\Big)$ consisting of n = 100 observations.

   a. Use the bootstrap to get the se and 95 percent confidence interval for θ (Usando el intervalo Normal).

b. Plot a histogram of the bootstrap replications. This is an estimate of the distribution of $\widehat{\theta}$. Compare this to the true sampling distribution of $\widehat{\theta}$.

8. **Wasserman, Cap 8.** Let $X_1, \ldots, X_n \sim$ Uniform $(0, \theta)$. Let $\hat{\theta} = X_{\max} = \max\{X_1, \ldots, X_n\}$. Generate a data set of size 50 with $\theta = 1$. Find the distribution of $\widehat{\theta}$. Compare the true distribution of $\widehat{\theta}$ to the histograms from the bootstrap. This is a case where the bootstrap does very poorly. In fact, we can prove that this is the case. Show that $P(\widehat{\theta} = \widehat{\theta}) = 0$ and yet $P\left(\widehat{\theta}^* = \widehat{\theta}\right) \approx .632$. In fact, $P\left(\widehat{\theta}^* = \widehat{\theta}\right) = 1 - (1 - (1/n))^n$.

9. **Dekking, Cap 18** We generate a bootstrap dataset $x_1^*, x_2^*, x_3^*, x_4^*$ from the empirical distribution function of the dataset

$$1, 3, 4, 6.$$

   a. Compute the probability that the bootstrap sample mean is equal to 1.
   b. Compute the probability that the maximum of the bootstrap dataset is equal to 6.
   c. Compute the probability that exactly two elements in the bootstrap sample are less than 2 .

10. De los datos recolectados en clase tenemos que la proporción de alumnos unidos es 63.33%. Muestreamos 10 (de los 30 alumnos) con reemplazo y 7 de los 10 muestreados son unidos.

   a. Usando Jackknife, obtenga el estimador corregido por sesgo para la proporción de personas unidas
   b. Calcule el estimador de la varianza del estimador de la proporción de personas unidas
   c. Utilizando bootstrap, obtenga los intervalos normal, pivotal y de percentil al 95%.