Please provide complete and well-written solutions to the following exercises.

Due October 20, 9AM, to be submitted in blackboard, under the Assignments tab.

Homework 4

Exercise 1. Write down the generalized likelihood ratio estimate for the following alpha particle data, as we did in class for a slightly different data set. The corresponding test treats individual counts of alpha particles as independent Poisson random variables, versus the alternative that the probability of a count appearing in each box of data is a sequence of nonnegative numbers that sum to one. (In doing so, you should need to compute a maximum likelihood estimate using a computer.)

m	$0,1$ or 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	≥ 17
# of Intervals	16	26	58	102	125	146	163	164	120	100	72	54	20	12	10	4

Plot the MLE for the Poisson statistic (i.e. plot the denominator of the generalized likelihood ratio test statistic $\frac{\sup_{\theta\in\Theta} f_\theta(X)}{\sup_{\theta\in\Theta_0} f_\theta(X)}$) as a function of λ.

Finally, compute the value s of Pearson's chi-squared statistic S, and compute the probability that $S \geq s$. Does the probability $\mathbf{P}(S \geq s)$ give you confidence that the null hypothesis is true?

Exercise 2. Let $X_1, \ldots, X_n$ be i.i.d. random variables. Let $0 < \alpha < 1/2$. Define the α-trimmed sample mean to be

$$\overline{X}_n^{(\alpha)} := \frac{1}{n - 2\lfloor n\alpha \rfloor} \sum_{i=\lfloor n\alpha\rfloor+1}^{n-\lfloor n\alpha\rfloor} X_{(i)}.$$

For any $w = (w_1, \ldots, w_n) \in \{1, \ldots, n\}^n$, define the Winsorized sample mean to be

$$\overline{X}_n^{(w)} := \frac{1}{n} \sum_{i=1}^{n} X_{(w_i)}.$$

- Show that the jackknife estimator of $\overline{X}_n^{(\alpha)}$ is
$$\frac{1}{1-2\alpha}(\overline{X}_n^{(w)} - 2\alpha\overline{X}_n^{(\alpha)}),$$
for some vector w.
- Show that the jackknife variance estimator of $\overline{X}_n^{(\alpha)}$ is
$$\frac{1}{n(n-1)(1-2\alpha)^2} \sum_{i=1}^{n} (X_{(w_i)} - \overline{X}_n^{(w)})^2,$$
for some vector w.

Exercise 3. Let X_1, X_2, X_3 be i.i.d. continuous random variables such that X_1 has PDF $\{f_\theta \colon \theta \in \Theta\}$. Let W_1, W_2, W_3 be a bootstrap sample from X_1, X_2, X_3. Let Y denote the sample median of X_1, X_2, X_3. (That is, Y is the middle value among X_1, X_2, X_3, which is unique with probability one since the random variables are continuous.)

- Describe the distribution of $(W_{(1)}, W_{(2)}, W_{(3)})$.
- Describe the bootstrap estimator of Y.
- Describe the bootstrap estimator of the variance of Y.

Exercise 4. Let $\mu \in \mathbf{R}$ and let $0 < \sigma < \infty$. Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables each with mean μ and variance σ^2. Let $h \colon \mathbf{R} \to \mathbf{R}$ be a function such that h' exists and is continuous. Let $\overline{X}_n := (X_1 + \cdots + X_n)/n$. Let $Y_n := h(\overline{X}_n)$.

Show that the jackknife estimator of the variance of Y_n converges almost surely to the same estimate of the variance you get by applying the Delta Method to Y_n.

Exercise 5. Suppose $X_1, \ldots, X_n$ is a random sample from a Gaussian random variable X with unknown mean $\mu_X \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$. Suppose $Y_1, \ldots, Y_m$ is a random sample from a Gaussian random variable Y with unknown mean $\mu_Y \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$.

Assume that $X_1, \ldots, X_n$ is independent of $Y_1, \ldots, Y_m$, i.e. assume that X, Y are independent.

Assume that $n + m > 2$. Define

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \overline{Y} := \frac{1}{m}\sum_{i=1}^{m} Y_i,$$

$$S_X^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \qquad S_Y^2 := \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \overline{Y})^2,$$

$$S^2 := \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}.$$

Show that

$$\frac{\overline{X} - \overline{Y} - \mu_X + \mu_Y}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

has Student's t-distribution with $n + m - 2$ degrees of freedom. Deduce the following confidence intervals for the difference of the means

$$\mathbf{P}\Big(\overline{X} - \overline{Y} - tS\sqrt{\frac{1}{n} + \frac{1}{m}} < \mu_X - \mu_Y < \overline{X} - \overline{Y} + tS\sqrt{\frac{1}{n} + \frac{1}{m}}\Big) = \frac{\Gamma(\frac{p+1}{2})}{\sqrt{p}\sqrt{\pi}\,\Gamma(p/2)} \int_{-t}^{t} \Big(1 + \frac{s^2}{p}\Big)^{-(p+1)/2} ds,$$

where $p = n + m - 2$.