Assignment sheet 8

November 7, 2018

Assignment 1. This assignment pertains to confidence bands and histograms. Let $f : [A, B] \to [0, \infty)$ be a continuous density function, and $X_1, \dots, X_n$ a sample from f. Let I be a histogram bin of length h, and recall that the histogram estimates (the restriction to I of) f by $(nh)^{-1} \sum_{j=1}^{n} \mathbf{1}\{X_j \in I\}$; this is the average number of sample points in I, normalized by the length of the interval (h).

(i). Invert a Wald test to construct an approximate $(1 - \alpha)$-confidence interval for $p_I = \mathbb{P}(X \in I)$ at level $\alpha \in (0, 1)$. *Hint : What is the probability of falling into a particular bin ?*

(ii). Suppose that h is small. Let $x = \inf I$ be the leftmost point of I. What is the (approximate) relation between $f(x)$ and p_I ?

(iii). Using the previous two parts, construct an approximate confidence interval for $f(x)$.

(iv). We would now like to construct a simultaneous confidence band for the entire density f on $[A, B]$ using the histogram. We need to correct for the multiple testing, so the first step is to understand : how many bins does the histogram contain ?

(v). Let $I_1, I_2, \dots, I_m$ denote the intervals corresponding to the histogram bins, and let $x = \inf I_1$. Using a Bonferroni correction, construct an approximate confidence region (product of intervals) that contains simultaneously all the density values $f(x), f(x + h), f(x + 2h), \dots, f(x + mh)$.

Remark. Since the number of points in the bins are correlated, we cannot use the independence correction.

Assignment 2. (Optional) Choosing the bandwidth is a crucial step in KDE. In class you saw that it regulates the variance-bias trade-off. In this exercise we are going to see this through a practical example.

The dataset `faithful` collects the duration of the eruptions and the waiting time between eruptions for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA.

(i). Search for and download the dataset. Save the waiting times in a vector `x`.

(ii). Use the functions `plot` and `density` to plot an estimated density for `x`. Which is the default kernel used by `density` ?

(iii). Plot an histogram of `x` and overline the curve plotted by the `density` function.

(iv). Repeat the previous point for different kernels.

(v). Run the following command and comment

```
par(mfrow=c(1,1))
kernels <- eval(formals(density.default)$kernel)
plot(density(precip),main = "Different kernels, bw not selected")
for(i in 2:length(kernels))
lines(density(precip,kern=kernels[i]),col=i)
legend("topright",legend=kernels,
col=seq(kernels),lty=1).
```

(vi). By default, bandwidth selection is done with the normal reference rule, but can also be done manually. Select manually the bandwidth within density. For example, plot several estimated densities over the histograms with bandwidth varying from 1 to 10 and chose the most suitable one by eye.

(vii). Plot the chosen bandwidth against the default one and the one picked by cross validation.

Assignment 3. (a) Let $A_{n\times m}$ and $B_{m\times n}$ be matrices such that AB is square. Show that $tr[AB] = tr[BA]$.
(b) Suppose that ABC is well-defined as a square matrix. Show that $tr[ABC] = tr[BCA]$.
Remark. In general case, trace of a product of matrices is invariant under cyclic permutations. For example, $ABC \mapsto ACB$ is not a cyclic permutation, thus it may be $tr[ACB] \neq tr[ABC]$.
(c) Let A be a random matrix (each of its coordinates is a random variable). Show that $\mathbb{E}(tr[A]) = tr[\mathbb{E}(A)]$, where the last expectation is interpreted coordinate-wise.

Assignment 4. (288) (a) Let P be a projection on a subspace V. If λ is an eigenvalue of P, Show that $\lambda = 0$ or 1.
(289) (b) Show that $Pv = v$ for all $v \in V$.
(c) Show that $Pw = 0$ for all $w \in V^{\perp}$. *Hint : compute* $(Pw)^T x$ *for* $w \in V^{\perp}$ *and* $x \in \mathbb{R}^p$.
(d) Let Q be another projection on the same space V. Show that $P = Q$.

Assignment 5. (289) Let $x_1, \dots, x_p$ be linearly independent vectors in $\mathbb{R}^n$ and X be an $n \times p$ matrix with columns $x_1, \dots, x_p$. Show that :
(a) the subspace $V = span(x_1, \dots, x_p)$ equals $M(X)$.
(b) $X^T X$ is invertible. *Hint : take* v *in the kernel and compute* $\|Xv\|^2$.
(c) the projection onto V is

$$H = X(X^T X)^{-1} X^T. \tag{1}$$

Assignment 6. (Optional) Here we display the rationale behind the important formula (1). Let $x_1, \dots, x_p$ be vectors in $\mathbb{R}^n$, and suppose that we wish to find the projection H onto their span V. Let X be a matrix with columns $x_1, \dots, x_p$.
(a) Explain why we can assume without loss of generality that (x_j) are independent.
(b) Explain why it makes sense to guess that H should take the form XM for some matrix M.
(c) Explain why it makes sense to guess that H should take the form NX^T for some matrix N.
(d) In view of (b) and (c), write $H = XBX^T$ for some $p \times p$ matrix B. Find B. *Hint : let* $e_i \in \mathbb{R}^p$ *be the i-th unit vector, then* $x_i = Xe_i$.

Assignment 7. (293) show that the two definitions of a positive definite matrix are equivalent.

Assignment 8. (295) Show that Q is an orthogonal projection of rank k if and only if there exist orthonormal vectors $v_1, \dots, v_k$ such that $Q = \sum_{i=1}^k v_i v_i^T$.

Assignment 9. (a) Let $Z \sim N(0_p, I_{p\times p})$ and H be an orthogonal projection of rank r. Show that $Z^T HZ \sim \chi^2_r$. *Hint : use the spectral decomposition of* H.
(b) Let $Y \sim N(\mu_p, \Omega_{p\times p})$ with Ω nonsingular. Show that $(Y-\mu)^T \Omega^{-1}(Y-\mu) \sim \chi^2_p$.