Answer sheet 8

Assignment 1. (i). Let $\overline{X}_I$ be the proportion of the sample points $(X_1, \ldots, X_n)$ that are in I. This is an average of a sample of Bernoulli random variables with success probability $p_I = \mathbb{P}(X \in I)$. A confidence interval for p_I is

$$\{p : n(p - \overline{X}_I)^2 \leq \overline{X}_I(1 - \overline{X}_I)\chi^2_{1,1-\alpha}\}.$$

(ii). Let F be the distribution function. Since h is small, we have

$$f(x) \approx \frac{F(x+h) - F(x)}{h} = \frac{p_I}{h}.$$

(iii). The approximate confidence interval for $f(x)$ is the rescaling of that of p_I, namely

$$\{p/h : n(p - \overline{X}_I)^2 \leq \overline{X}_I(1 - \overline{X}_I)\chi^2_{1,1-\alpha}\}.$$

The density estimator is constant at each bin, so the confidence interval of $f(y)$, $y \in I$ is the same as that of $f(x)$. Now, since f is assumed continuous, its values do not vary much in I, so this is sensible.

(iv). There are $(B - A)/h$ bins. More precisely, the number of bins is the smallest integer $\geq (B - A)/h$.

(v). The Bonferroni correction entails dividing α by the number of bins $m \approx (B - A)/h$. The confidence region is therefore the product set

$$\{p/h : n(p - \overline{X}_{I_j})^2 \leq \overline{X}_{I_j}(1 - \overline{X}_{I_j})\chi^2_{1,1-\alpha/m}\}, \qquad j = 1, \ldots, m.$$

Assignment 2. (i).

```
data("faithful", package = "datasets")
x <- faithful$waiting
```

(ii).

```
plot(density(x))
```

The default kernel used by `density` is Gaussian.

(iii).

```
hist(x, xlab = "Waiting times", ylab = "Frequency",
probability = TRUE, main = "Gaussian kernel",border = "gray")
lines(density(x, width = 12), lwd = 2)
```

(iv).

```
hist(x, xlab = "Waiting times)", ylab = "Frequency",
probability = TRUE, main = "Rect. kernel",border = "gray")
lines(density(x, width = 12,window = "rectangular"), lwd = 2)
rug(x)
hist(x, xlab = "Waiting times", ylab = "Frequency",
probability = TRUE, main = "Triang. kernel",border = "gray")
lines(density(x, width = 12, window ="triangular"), lwd = 2)
```

(v). Different kernels, same bandwidth.

(vi).
```
hist(x, xlab = "Waiting times", ylab = "Frequency",
probability = TRUE, main = "Manual bw selection, Gaussian kernel"
,border = "gray")
bandwidth <- 1:10
for(i in bandwidth)
lines(density(x, width = 12, bw=i), lwd = 2, col=i)
legend("topright",legend=bandwidth,
col=seq(bandwidth),lty=1)
```

We could chose 3 or 4 ?

(vii). The normal reference rule chooses a bandwidth of 4.7, CV a bandwidth of 2.66, manual selection here is 3. Here the comparison plot.

```
hist(x, xlab = "Waiting times", ylab = "Frequency",
probability = TRUE, main = "Manual bw selection,
Gaussian kernel",  border = "gray")
bandwidth <- c('manual','nrd0' , 'ucv')
lines(density(x,bw=3),col=1)
for(i in 2:length(bandwidth))
lines(density(x,bw=bandwidth[i]),col=i)
legend("topright",legend=bandwidth,
col=seq(bandwidth),lty=1)
```

Assignment 3. (a) $(AB)_{ik} = \sum_{j=1}^{m} a_{ij} b_{jk}$ thus

$$\mathrm{tr}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ji} = \sum_{j=1}^{n} \sum_{i=1}^{n} b_{ji} a_{ij} = \mathrm{tr}(BA).$$

(b) This follows from (a) with $A' = A$ and $B' = BC$.
(c) By linearity of the expected value, $\mathbb{E}(\mathrm{tr}(A)) = \mathbb{E} \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \mathbb{E}(a_{ii}) = \mathrm{tr}(\mathbb{E}(A))$.

Assignment 4. (a) Let $v \in \mathbb{R}^p \setminus \{0\}$ such that $Pv = \lambda v$. Then

$$\lambda v = Pv = PPv = P\lambda v = \lambda P v = \lambda^2 v.$$

As $v \neq 0$ this implies $\lambda = \lambda^2$; equivalently $\lambda \in \{0, 1\}$.
(b) There exists $u \in \mathbb{R}^p$ such that $v = Pu = PPu = Pv$.
(c) We have $(Pw)^T x = w^T P^T x = w^T (Px) = 0$ because $w \in W$ must be orthogonal to $Px \in V$. This means that Pw is orthogonal to everything and hence equals 0.
(d) Each $x \in \mathbb{R}^p$ can be written (uniquely) as $v + w$, $v \in V$, $w \in V^{\perp}$. Since P and Q agree on V and $V^{\perp}$, they must agree throughout $\mathbb{R}^p$.

Assignment 5. (a) For each $u = (u_1, \dots, u_p) \in \mathbb{R}^p$ we have $Xu = u_1 x_1 + \cdots + u_p x_p$, and these constitute precisely the elements of V.
(b) If $X^T X v = 0$, then

$$\|Xv\|^2 = v^T X^T X v = 0,$$

which means that $Xv = 0$. By part (a), Xv is a linear combination of the columns of X. Since these are independent, it must be that $v = 0$. As the $p \times p$ matrix $X^T X$ is injective, it must be invertible.
(c) To see that H is a projection simply note that

$$H^2 = X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T = X(X^T X)^{-1} X^T = H,$$

and

$$H^T = (X(X^T X)^{-1} X^T)^T = (X^T)^T [(X^T X)^{-1}]^T X^T = X([X^T X]^T)^{-1} X^T = X(X^T X)^{-1} X^T = H.$$

Clearly $Hy = X[(X^T X)^{-1} X^T y] \in V$, and so $M(H) \subseteq V$. Conversely, if $y \in V$ then $y = Xu$ for some $u \in \mathbb{R}^p$ and then $Hy = HXu = Xu = y$, so $y \in M(H)$. This completes the proof.

Assignment 6. (a) Otherwise, we can remove a subset of them without changing the span, and do so repeatedly until we have an independent set.
(b) This is so because Hy must belong to the column space of X, hence equal Xv for some v. Since everything is linear v should be a linear function X, $v = My$, and then $H = XM$.
(c) For any $y \in V^\perp$, $Hy = 0$, which means that $X_i^T y$ has to be zero. These are precisely the coordinates of the p-dimensional vector $X^T y$, which then should be zero. Conversely, if $y \notin V^\perp$, then $X_i^T y$ will be nonzero for some i, and so $X^T y$ will not be zero. Thus X^T is the "minimal" matrix with kernel $V^\perp$.
(d) We know that $Hx_i = x_i$ for all i, and using the hint

$$Xe_i = x_i = Hx_i = XBX^T x_i = XBX^T Xe_i.$$

Since X is injective, this means that $BX^T Xe_i = e_i$. This holds for all i, which means that $BX^T X$ is the identity and then $B = (X^T X)^{-1}$.

Assignment 7. Let $\Omega = U\Lambda U^T$ be the spectral decomposition of Ω, and let $\lambda_i = \Lambda_{ii}$ be the eigenvalues of Ω (in an arbitrary order). Then for any $v \in \mathbb{R}^p$ we have

$$v^T \Omega v = \sum_{i=1}^{p} [Uv]_i^2 \lambda_i.$$

If all the λ_i's are (strictly) positive, then this is (strictly) positive for all $v \neq 0$ (because U is injective, so $Uv \neq 0$). If one $\lambda_i < 0$ then choosing $[Uv]_j$ to be 0 for $j \neq i$ and 1 for $j = i$ gives $v^T \Omega v < 0$. Such a choice is possible since U is surjective.

Assignment 8. Clearly such Q is symmetric, and by orthonormality

$$Q^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} v_i v_i^T v_j v_j^T = \sum_{i=1}^{k} v_i v_i^T v_i v_i^T = \sum_{i=1}^{k} v_i v_i^T = Q.$$

Since $Qv_i = v_i$ for all i and $Qv = 0$ for all $v \in [span(v_1, \ldots, v_k)]^\perp$, Q is the projection on this span and hence of rank k.
Conversely, if Q is a projection, we can let $v_1, \ldots, v_k$ be an orthonormal basis of $M(Q)$. Let V be a matrix with columns $v_1, \ldots, v_k$. Then we know that $Q = V(V^T V)^{-1} V^T = VV^T$, and

it remains to show that this is the same matrix as $Q' = \sum_{i=1}^{k} v_i v_i^T$. Since $v_j = Ve_j$ for the unit vector e_j and the v_i's are orthogonal,

$$Qv_j = VV^TVe_j = Ve_j = \sum_{i=1}^{k} v_i v_i^T v_j = Q'v_j.$$

Hence Q and Q' agree on the basis of $M(Q)$ and thus on the whole $M(Q)$. On the complement, we have $v^T v_j = 0$ for all j, then clearly $Qv = 0 = Q'v$. Thus $Q = Q'$.

Assignment 9. (a) If U is orthogonal, then $W = UZ \sim N(0, UIU^T) = N(0, I)$. Let $H = U\Lambda U^T$ be a spectral decomposition of H with the first r elements of Λ equal to one and the rest equal to zero (in view of a previous assignment). Then

$$Z^T HZ = W^T \Lambda W = \sum_{i=1}^{r} W_i^2 \sim \chi_r^2.$$

(We used the fact that the marginal law of $(W_1, \ldots, W_r)$ is $N(0_r, I_{r\times r})$.
(b) Define $Z = \Omega^{-1/2}(Y - \mu) \sim N(0, \Omega^{-1/2}\Omega\Omega^{-1/2}) = N(0_p, I_{p\times p})$. Then

$$(Y - \mu)^T \Omega^{-1}(Y - \mu) = Z^T Z \sim \chi_p^2.$$