1. Let Θ be the set of all continuous distribution functions on $\mathbb{R}$ and $\mathcal{A} = \mathbb{R}$ be the action space. Furthermore, suppose that the loss function is of the form $L(F, a) = W(F(a))$ for $F \in \Theta$ and $a \in \mathcal{A}$ for some function $W : [0, 1] \to \mathbb{R}$. Suppose $\mathcal{X} = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^n)$, and $\mathcal{P} = \{P_F : F \in \Theta\}$ where $P_F((-\infty, x]) = \prod_{i=1}^n F(x_i)$ for $x \in \mathbb{R}^n$. Also, let $T(X) = (X_{(1)}, \ldots, X_{(n)}))^\top$, which is sufficient for $\mathcal{P}$. Finally, let $\mathcal{Y} = T(\mathcal{X})$, and G be the group of transformations on $\mathcal{Y}$ as $G = \{g_\phi : \phi \in \mathcal{S}\}$ where $g_\phi(y) = (\phi(y_1), \ldots, \phi(y_n))^\top$ and $\mathcal{S}$ is the set of all functions which are continuous, strictly increasing, and surjective.

   (a) Show that this decision problem is invariant under G and find the group $\bar{G}$ and $\tilde{G}$.

Solution.

i. Note that there exists $\tau \in S_n$ such that $T(X) = (X_{\tau(1)}, \ldots, X_{\tau(n)})^\top$. Then,

$$\begin{aligned}
P_F(g_\phi T \le t) &= P_F(T \le g_\phi^{-1} t) = P_F(X \le T^{-1}(g_\phi^{-1} t)) \\
&= P_F(X \le (\phi^{-1}(t_{\tau^{-1}(1)}), \ldots, \phi^{-1}(t_{\tau^{-1}(1)}))^\top) \\
&= \prod_{i=1}^n (F \circ \phi^{-1})(t_{\tau^{-1}(i)}) = P_{F\circ\phi^{-1}}(X \le T^{-1}(t)) \\
&= P_{F\circ\phi^{-1}}(T \le t).
\end{aligned}$$

This implies that $\mathcal{P}$ is G-invariant with $\bar{G} = \{\bar{g}_\phi\}$ where $\bar{g}_\phi(F) = F \circ \phi^{-1}$.

ii. For all $g_\phi \in G$ and $a \in \mathcal{A}$, if $a^* = \phi(a)$, then

$$\begin{aligned}
L(F, a) &= W(F(a)) = W((F \circ \phi^{-1})(a^*)) \\
&= L(\bar{g}_\phi F, a^*).
\end{aligned}$$

Thus, L is also G-invariant with $\tilde{G} = \{\tilde{g}_\phi\}$ where $\tilde{g}_\phi = \phi$.

■

(b) For each $y \in \mathbb{R}^n$, let $\Phi_y = \{\phi \in \mathcal{S} : g_\phi(y) = y\}$. Suppose that δ is a behavioral decision rule $\delta : \mathcal{Y} \to \mathcal{A}^*$. For a fixed $y \in \mathcal{Y}$, let $Z \sim \delta(y)$. Show that if $\phi(Z) \stackrel{d}{=} Z$ for all $\phi \in \Phi_y$, then $Z = \phi(Z)$ a.e. for all $\phi \in \Phi_y$.

Solution. Let P be a generic notation of probability measures. Let $x, x' \in \mathbb{R}$ with $y_n < x < x' < y_{n+1}$. For all $n \in \mathbb{N}$, consider $\phi_n \in \mathcal{S}$ defined as $\phi_n^{-1}(t) = n^{-1}(t - x)$ for $t \in \mathbb{R}$. Then, since $Z \stackrel{d}{=} \phi_n(Z)$ for all $n \in \mathbb{N}$.

$$\begin{aligned} P(Z \in (x, x']) &= P(\phi_n(Z) \in (x, x']) = P(Z \in (\phi_n^{-1}(x), \phi_n^{-1}(x')]) \\ &= P(Z \in (0, n^{-1}(x' - x)]) \xrightarrow[n\to\infty]{} 0 \end{aligned}$$

by continuity of measures. Since x, x' are arbitrary, we have that $Z \in \{y_1, \ldots, z_n\}$ almost everywhere. Therefore, for all $\phi \in \Phi_y$, by definition of Φ_y, we conclude that

$$\begin{aligned} P(Z \neq \phi(Z)) &= \sum_{i=1}^{n} P(Z \neq \phi(Z) | Z = y_i) P(Z = y_i) \\ &= \sum_{i=1}^{n} P(y_i \neq \phi(y_i) | Z = y_i) P(Z = y_i) = 0, \end{aligned}$$

i.e., $Z = \phi(Z)$ a.e. ∎

(c) Show that there are only n non-randomized invariant decision rules, namely $d_i(y) = y_i$ for $i = 1, \ldots, n$.

Solution. By (b), for all $\delta \in \Delta$, the support of $\delta(y)$ is equal to $\{y_1, \ldots, y_n\}$. Thus, if δ is non-randomized, then for all $y \in \mathcal{Y}$, $\delta(y) \in \{y_1, \ldots, y_n\}$. It is obvious that for all $i = 1, \ldots, n$, $d_i(y) = y_i \in \{y_1, \ldots, y_n\}$ and d_i is invariant since $d_i(g_\phi(y)) = \phi(y_i) = \tilde{g}_\phi(d_i(y))$. Hence, there are only n non-randomized invariant decision rules $d_1, \ldots, d_n$, which are defined as $d_i(y) = y_i$ for $y \in \mathcal{Y}$ and $i = 1, \ldots, n$. ∎

(d) (The problem of estimating mdeian) When $W(p) = (p - 1/2)^2$, find the BEE.

Solution. Let $U_i = F(X_i)$, then $U_i \overset{iid}{\sim} U(0,1)$ for $i = 1, \dots, n$, and $U_{(i)} \sim Beta(i, n+1-i)$ for all $i = 1, \dots, n$. This implies that the risk of d_i is given as

$$\begin{aligned} R(F, d_i) &= E_F[\{F(d_i(T(X))) - 1/2\}^2] = E[(U_{(i)} - 1/2)^2] \\ &= var[U_{(i)}] + (E[U_{(i)}] - 1/2)^2 \\ &= \frac{i(n+1-i)}{(n+1)^2(n+2)} + \left(\frac{i}{n+1} - \frac{1}{2}\right)^2 \\ &= \frac{1}{(n+1)(n+2)}\left(i - \frac{n+1}{2}\right)^2 + \frac{1}{4(n+2)}. \end{aligned}$$

Thus, $R(F, d_i)$ is minimized at $i = (n+1)/2$ if n is odd and at $i = n/2$ and $i = (n+2)/2$ if n is even, i.e.,

$$i^* := \underset{i \in \{1,\dots,n\}}{\text{argmin}}\ R(F, d_i) = \begin{cases} (n+1)/2 & \text{if } n \text{ if even} \\ \{n/2, (n+2)/2\} & \text{if } n \text{ if odd} \end{cases}.$$

This means that the BEE is the sample median $X_{i^*} = \text{med} X_i$. ■

(e) (The prolem of estimating the lower bound) When $W(p) = p\mathrm{I}(p > 0) + \mathrm{I}(p = 0)$, find the BEE.

Solution. In the same way as before, the risk is

$$\begin{aligned} R(F, d_i) &= E_F[F(d_i(T(X)))\mathrm{I}(F(d_i(T(X))) > 0) + \mathrm{I}(F(d_i(T(X))) = 0)] \\ &= E[U_{(i)}\mathrm{I}(I_{(i)} > 0) + \mathrm{I}(I_{(i)} = 0)] = \frac{i}{n+1}. \end{aligned}$$

Thus, $\text{argmin}_{i \in \{1,\dots,n\}} R(F, d_i) = 1$. This means that the BEE is $X_{(1)}$. ■

2. In a hypothesis testing problem, suppose T is a boundedly complete statistic with distribution P_θ where $\theta \in \Theta$. Let $\mathcal{Y}$ be the range of T, i.e., $\mathcal{Y} = T(\mathcal{X})$, and suppose that the problem of testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ is invariant under a group of transformations G on $\mathcal{Y}$. Suppose $\psi : \mathcal{Y} \to [0, 1]$ is a test function and $\beta(\theta) = E_\theta[\psi(T)]$ is its power function. Show that $\beta(\theta)$ is invariant under G if and only if ψ is almost invariant under G.

Solution. Note that $E_\theta[\psi(gT)] = E_{\bar{g}\theta}[\psi(T)] = \beta(\bar{g}\theta)$.

$\Leftarrow$: Assume that ψ is almost invariant under G. Let $g \in G$ and $\theta \in \Theta$. Then, $\psi(gT) = \psi(T)$ a.e. with respect to P_θ. This implies that

$$\beta(\bar{g}\theta) = E_\theta[\psi(gT)] = E_\theta[\psi(T)] = \beta(\theta).$$

In brief, for all $g \in G$ and $\theta \in \Theta$, $\beta(\bar{g}\theta) = \beta(\theta)$, i.e., β is $\bar{G}$-invariant.

$\Rightarrow$: Assume that β is $\bar{G}$-invariant. Let $g \in G$ and $\theta \in \Theta$. Then,

$$E_\theta[\psi(gT)] = \beta(\bar{g}\theta) = \beta(\theta) = E_\theta[\psi(T)].$$

Since T is boundedly complete, $\psi(gT) = \psi(T)$ a.e. with respect to P_θ. In brief, for all $g \in G$ and $\theta \in \Theta$, $\psi(gT) = \psi(T)$ a.e. with respect to P_θ, i.e., ψ is almost invariant under G.

■

3. Generalize the proof of the following theorem to the case where G is a compact topological group with the Borel σ-field $\mathcal{B}(G)$ on G and the uniform density p on G, i.e., $p(dg) = \mathrm{vol}(G)^{-1}$.

Theorem. Suppose a decision problem is invariant under a finite group G. Then, if there exists a minimax rule, then there exists a minimax rule that is (behavioral) invariant. If a rule is minimax within the class of behavioral invariant rules, then it is minimax.

Solution. It suffices to prove that for all $\delta \in \Delta$, there exists $\delta^I \in \Delta$ such that δ^I is invariant and

$$\sup_{\theta\in\Theta} R_*(\theta, \delta^I) \le \sup_{\theta\in\Theta} R_*(\theta, \delta).$$

Let $\delta \in \Delta$. Let $M = \text{vol}(G)$ and define $\delta^I \in \Delta$ as

$$\{\delta^I(x)\}(A) = M^{-1} \int_G \{\delta(hx)\}(\tilde{h}A) dp(h), \quad A \in \mathcal{L}_{\mathcal{A}}, x \in \mathcal{X}.$$

The above integral is defined with *Haar* integral. Recall that the Haar integral satisfies

$$\int_G f(gx) d\mu(x) = \int_G f(x) d\mu(x)$$

where μ is a left Haar measure and f is a Borel measurable function on a locally compact topological group G.

(a) Let $g \in G$, $x \in \mathcal{X}$, and $A \in \mathcal{L}_{\mathcal{A}}$. Then, by using the property of Haar integral, we have

$$\begin{aligned} \{(\delta^I)^g(x)\}(A) = \{\delta^I(gx)\}(\tilde{g}A) &= M^{-1} \int \{\delta(ghx)\}(\tilde{g}\tilde{h}A) dp(h) \\ &= M^{-1} \int \{\delta(hx)\}(\tilde{h}A) dp(h) = \{\delta^I(x)\}(A). \end{aligned}$$

Thus, δ^I is G-invariant.

(b) By Fubini-Tonelli theorem, for all $\theta \in \Theta$,

$$\begin{aligned} L^*(\theta, \delta^I(x)) &= \int_{\mathcal{A}} L(\theta, \cdot) d(\delta^I(x)) \\ &= M^{-1} \int_G \int_{\mathcal{A}} L(\theta, \cdot) d\{\delta^g(x)\} dp(g) \\ &= M^{-1} \int_G L^*(\theta, \delta^g(x)) dp(g). \end{aligned}$$

Again by Funini-Tonelli theorem, this implies that for all $\theta \in \Theta$,

$$\begin{aligned} R_*(\theta, \delta^I) = E_\theta[L^*(\theta, \delta^I(X))] = M^{-1} \int_G R_*(\theta, \delta^g) dp(g) = M^{-1} \int_G R_*(\bar{g}\theta, \delta) dp(g) \\ \leq M^{-1} \int_G \sup_{\theta \in \Theta} R_*(\bar{g}\theta, \delta) dp(g) = \sup_{\theta \in \Theta} R_*(\theta, \delta). \end{aligned}$$

In brief, for all $\delta \in \Delta$, there exists $\delta^I \in \Delta$ such that δ^I is invariant and

$$\sup_{\theta \in \Theta} R_*(\theta, \delta^I) \leq \sup_{\theta \in \Theta} R_*(\theta, \delta),$$

and thus, the proof is complete. ■

4. Suppose in an experiment where the outcomes are either success and failure, 17 successes and 115 failures were observed. Let θ denote the success probability.

   (a) Assume the experiement was a Bernoulli trial so that $X \sim Bin(132, \theta)$ with observed $X = 17$. Find the MLE of θ and the 95% UMA interval for θ.

   (b) Assume the experiement was a negative binomial trial so that $X \sim NegBin(17, \theta)$ with observed $N = 132$. Find the MLE of θ and the 95% UMA interval for θ.

Solution. Omitted. ■

5. Consider the Cox's example in support of conditionality principle. Suppose that the more precise machine is available with a probability $p \in (0, 0.5]$.

   (a) Suppose $X = 29$ was observed. Compute a 95% frequentist confidence interval based on LRT for θ assuming that it was not told which machine was used.

Solution. Let $f_i(x|\theta) = (2\pi\sigma_i^2)^{-1/2} \exp\left[-(x-\theta)^2/(2\sigma_i^2)\right]$ for $i = 1, 2$ where $\sigma_1^2 = 0.1$ and $\sigma_2^2 = 10$. Then, the distribution of X is $f(x|\theta) = pf_1(x|\theta) + (1-p)f_2(x|\theta)$. Note that

$$\begin{aligned}
\partial_\theta f_i(x|\theta) &= (2\pi\sigma_i^2)^{-1/2} \exp\left[-(x-\theta)^2/(2\sigma_i^2)\right] \frac{2}{2\sigma^2}(x-\theta) \\
&= f_i(x|\theta)\frac{x-\theta}{\sigma_i^2}.
\end{aligned}$$

Thus, where $l(\theta|x) = \log f(x|\theta)$ is the log likliehood,

$$\begin{aligned}
l'(\theta|x) &= f(x|\theta)^{-1}\{p\partial_\theta f_1(x|\theta) + (1-p)\partial_\theta f_2(x|\theta)\} \\
&= (x-\theta)\frac{pf_1(x|\theta)\sigma_1^{-2} + (1-p)f_2(x|\theta)\sigma_2^{-2}}{f(x|\theta)},
\end{aligned}$$

which implies that $\hat{\theta} = x$ is the MLE. Then, a 95% confidence interval based on LRT for θ is

$$\begin{aligned}
C_\theta(x) &= \left\{\theta \in \mathbb{R} : \frac{L(\theta|x)}{L(\hat{\theta}|x)} \geq C\right\} \\
&= \{\theta \in \mathbb{R} : f(x|\theta) \geq K\}
\end{aligned}$$

where $K = f(x|\hat{\theta})C = \{(2\pi\sigma_1^2)^{-1/2} + (2\pi\sigma_2^2)^{-1/2}\}C$ is constant. Note that

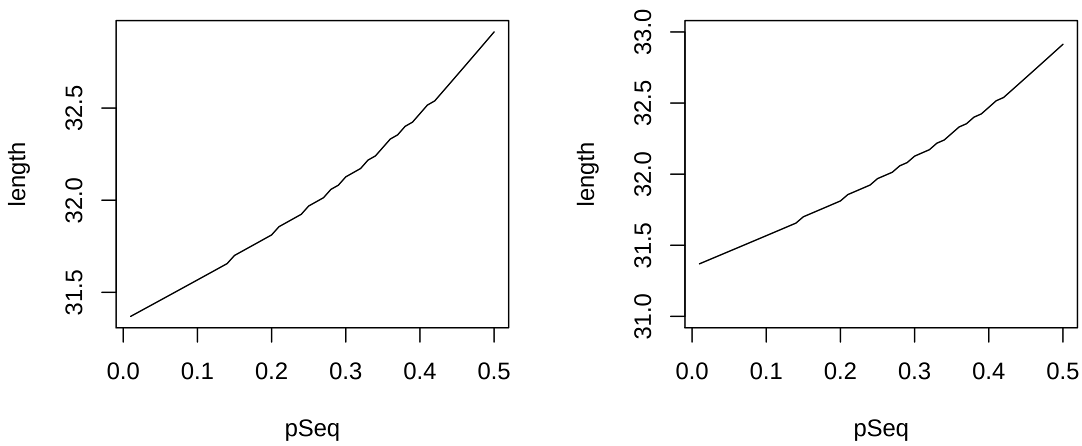
$$1-\alpha \le P_\theta(\theta \in C_\theta(X)) = \int f(x|\theta)\mathrm{I}(\theta \in C_\theta(x))dx = \int f(x|\theta)\mathrm{I}(f(x|\theta) \ge K)dx$$
$$= \int f(x|0)\mathrm{I}(f(x|0) \ge K)dx.$$

From this, we can find K and then C for given $\alpha \in (0,1)$, and finally we will obtain $C_\theta(x)$. We can numerically obtain this interval as in (b) and (c). ■

(b) Plot the width of the CI as a function of p.

Solution.
```
s1<-0.1; s2<-10; x<-29; thHat<-x; a<-0.05
pSeq<-seq(0.01, 0.5, by=0.01)
CI<-matrix(0, nrow = length(pSeq), ncol = 2)
LL<-rep(0, length(pSeq))
lthSeq<-rep(0, length(pSeq))
```

```
for(j in 1:length(pSeq)){
  p<-pSeq[j]
  cc<-(p*(2*pi*s1^2)^(-1/2) + (1-p)*(2*pi*s1^2)^(-1/2))
  lrt<-function(th, p=1/2){
    (p*dnorm(x, th, s1) + (1-p)*dnorm(x, th, s2)) /
      (p*(2*pi*s1^2)^(-1/2) + (1-p)*(2*pi*s1^2)^(-1/2)) }
  f<-function(x, th=0){ p*dnorm(x, th, s1) + (1-p)*dnorm(x, th, s2) }
  ff<-function(x, th=0){ p*pnorm(x, th, s1) + (1-p)*pnorm(x, th, s2) }
  k<-100; c<-0.01; d<-0.00001
  repeat{
    temp<-c(uniroot(function(x){f(x) - c}, c(-k, 0))$root,
            uniroot(function(x){f(x) - c}, c(0, k))$root)
    lth<-ff(temp[2]) - ff(temp[1])
    if(lth>1-a){ break }
    c<-c-d }
  lthSeq[j]<-lth
  CI[j,]<-c(uniroot(function(th){lrt(th) - c/cc}, c(29-k, 29))$root,
        uniroot(function(th){lrt(th) - c/cc}, c(29, 29+k))$root)
  LL[j]<-diff(CI[j,]) }
```



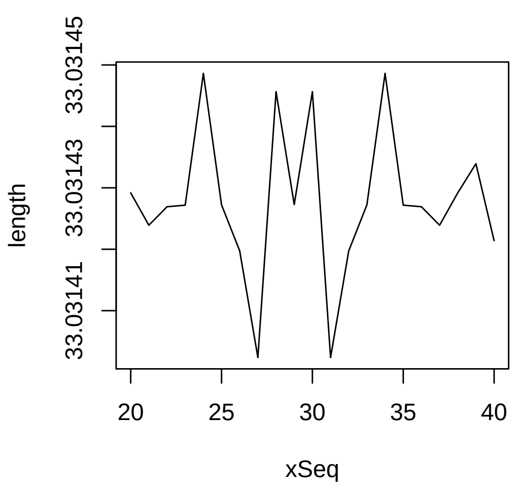
∎

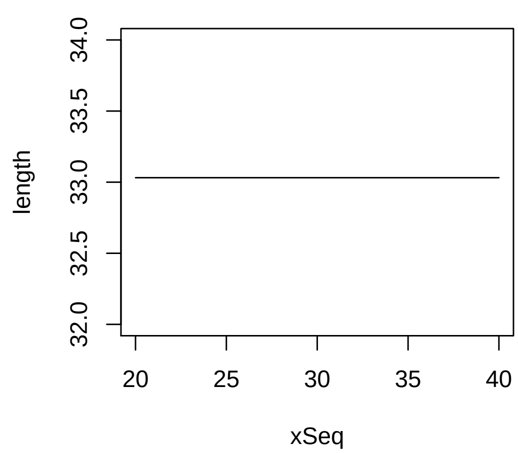
(c) If p is fixed, does the width depend on the value of X observed? Justifying intuitively using invariance and demonstrate numerically.

Solution. We can see that $C_\theta(x)$ does not depend on $\hat{\theta}$ as above. Also, $f(x|\theta)$ is symmetric about $\theta = x$ and have the maximum at $\theta = x$. Therefore, the width would not depend on the value of X. The example corresponding to $p = 0.5$ is as follows.

```
p<-1/2
xSeq<-seq(20, 40, by=1)
CI.x<-matrix(0, nrow = length(xSeq), ncol = 2)
LL.x<-rep(0, length(xSeq))
```

```
for(j in 1:length(xSeq)){
  x<-xSeq[j]
  cc<-(p*(2*pi*s1^2)^(-1/2) + (1-p)*(2*pi*s1^2)^(-1/2))
  lrt<-function(th, p=1/2){
    (p*dnorm(x, th, s1) + (1-p)*dnorm(x, th, s2)) /
      (p*(2*pi*s1^2)^(-1/2) + (1-p)*(2*pi*s1^2)^(-1/2)) }
  f<-function(x, th=0){ p*dnorm(x, th, s1) + (1-p)*dnorm(x, th, s2) }
  ff<-function(x, th=0){ p*pnorm(x, th, s1) + (1-p)*pnorm(x, th, s2) }
  k<-100; c<-0.01; d<-0.0001
  repeat{
    temp<-c(uniroot(function(x){f(x) - c}, c(-k, 0))$root,
            uniroot(function(x){f(x) - c}, c(0, k))$root)
    lth<-ff(temp[2]) - ff(temp[1])
    if(lth>1-a){ break }
    c<-c-d }
  CI.x[j,]<-c(uniroot(function(th){lrt(th) - c/cc}, c(29-k, 29))$root,
              uniroot(function(th){lrt(th) - c/cc}, c(29, 29+k))$root)
  LL.x[j]<-diff(CI.x[j,]) }
```





■

6. Suppose $p > 1$, and $X \sim N_p(\mu, \kappa^{-1}I)$ such that $\|\mu\| = 1$ and $\kappa > 0$, and $\varepsilon \sim N_p(0, \tau^{-1}I)$. Define the family of distributions $\mathcal{P} = \{P_\theta : \theta = (\mu^\top, \kappa, \tau)^\top \in S^{p-1} \times \mathbb{R}_+ \times \mathbb{R}_+\}$ where $S^{p-1} = \{x \in \mathbb{R}^p : \|x\| = 1\}$ and P_θ is the same density as that of $Y := X/\|X\| + \varepsilon$. Suppose $G = \{g_A : A \in O(p)\}$ where $g_A(y) = Ay$ and $O(p)$ is the set of all $p \times p$ orthogonal matrices.

(a) Show that the density (with respect to the Lebsegue measure) p_θ of P_θ is

$$p_\theta(y) = \frac{\tau^{p/2}\kappa^{p/2-1}\exp\left[-\frac{1}{2}\tau(\|y\|^2+1)\right] I_{p/2-1}(\|\tau y + \kappa\mu\|)}{(2\pi)^{p/2} I_{p/2-1}(\kappa)\|\tau y + \kappa\mu\|^{p/2-1}}$$

where I_α is the *modified Bessel function of the first order* α defined as

$$I_\alpha(x) = \sum_{m=0}^{\infty} m!\Gamma(m+\alpha+1)\left(\frac{x}{2}\right)^{2m+\alpha}$$

Solution. Note that $P(X = 0) = 0$, and let $Z = X/\|X\|$. Let p_θ be a generic notation of density. Then, the pdf of Z is $p_\theta(z) = C_p(\kappa)\exp[\kappa\mu^\top z]\mathrm{I}(\|z\| = 1)$ for $z \in \mathbb{R}^p$ where $C_p(\kappa) = \kappa^{p/2-1}(2\pi)^{-p/2}\mathrm{I}_{p/2-1}(\kappa)^{-1}$. Recall that $\int \exp[\alpha^\top z]\mathrm{I}(\|z\| = 1)dz = C_p(\|\alpha\|)^{-1}$. This implies that

$$\begin{aligned}
p_\theta(y) &= \int p_\theta(y|z)p_\theta(z)dz \\
&= \int |2\pi\tau^{-1}I|^{-1/2}\exp\left[-\frac{1}{2}(y-z)^\top(\tau^{-1}I)^{-1}(y-z)\right] C_p(\kappa)\exp[\kappa\mu^\top z]\mathrm{I}(\|z\|=1)dz \\
&= (2\pi)^{-p/2}\tau^{p/2}\exp\left[-\frac{\tau}{2}(\|y\|^2+1)\right] C_p(\kappa)\int \exp\left[(\tau y+\kappa\mu)^\top z\right]\mathrm{I}(\|z\|=1)dz \\
&= (2\pi)^{-p/2}\tau^{p/2}\exp\left[-\frac{\tau}{2}(\|y\|^2+1)\right]\frac{C_p(\kappa)}{C_p(\|\tau y+\kappa\mu\|)} \\
&= \frac{\tau^{p/2}\kappa^{p/2-1}\exp\left[-\frac{1}{2}\tau(\|y\|^2+1)\right] I_{p/2-1}(\|\tau y+\kappa\mu\|)}{(2\pi)^{p/2}I_{p/2-1}(\kappa)\|\tau y+\kappa\mu\|^{p/2-1}}.
\end{aligned}$$

■

(b) Show that $\mathcal{P}$ is G-invariant, and identify $\bar{G}$.

Solution. Let $A \in O(p)$. Note that $AY = AX/\|X\| + A\varepsilon$. Here, $A\varepsilon \overset{d}{=} \varepsilon$ since $A \in O(p)$. Also, $AX \sim N_p(A\mu, \kappa^{-1}I)$ and $\|AX\|^2 = X^\top A^\top AX = \|X\|^2$ since $A \in O(p)$. This implies that $g_AY = AX/\|AX\| + \varepsilon$, which means that the density of AY is also contained in $\mathcal{P}$. Finally, we have that $\bar{G} = \{\bar{g}_A : A \in O(p)\}$ where $\bar{g}_A(\theta) = ((A\mu)^\top, \kappa, \tau)^\top$. ■

(c) Suppose that κ and τ are known, and we want to estimate μ. The action space is also the unit surface S^{p-1} in $\mathbb{R}^p$. Show that the loss function $L(\mu, a) = 1 - \mu^\top a$ is G-invariant, and identify $\tilde{G}$.

Solution. Let $A \in O(p)$. Then,

$$L(\bar{g}_A \mu, a^*) = 1 - \mu^\top A^\top a^* = 1 - \mu^\top a$$

where $a^* = Aa$. Thus, L is also G-invariant, and we have that $\tilde{G} = G$. ■