Exercises October 23 **STA 4508S (Fall, 2018)**

1. The Kullback-Leibler divergence from the distribution G to the distribution F is given by

$$KL(F : G) = \int \log \frac{f(y)}{g(y)} f(y) dy, \tag{1}$$

where f and g are and density functions with respect to Lebesgue measure. Note that the divergence is not symmetric in its arguments. This is called the directed information distance in Barndorff-Nielsen and Cox (1994) where the more general definition $KL(F : G) = \int \log(dF/dG)dF$ is used, assuming F and G are mutually absolutely continuous.

   (a) In the canonical exponential family model with density $f(s; \varphi) = \exp\{\varphi^T s - k(\varphi)\}h(s)$, $s \in \mathbb{R}^p$, find an expression for the KL divergence between the model with parameter φ_1 and that with parameter φ_2. I should have said "from one to the other" to be clear. If we say "from $f(s; \varphi_1)$ to $f(s; \varphi_2)$ then it's $(\varphi_1 - \varphi_2)^T k'(\varphi_1) - k(\varphi_1) + k(\varphi_2)$. If $\varphi_2 = \hat{\varphi}$ and $\varphi_1 = \varphi_0$, then it's $k(\hat{\varphi}) - k(\varphi_0) - (\hat{\varphi} - \varphi_0)k'(\varphi_0)$, so close to quadratic in $\hat{\varphi} - \varphi_0$.

   (b) Show that for a sample of observations from a model with density $f(y; \theta)$ the maximum likelihood estimator minimizes the KL divergence from $F(\cdot; \theta)$ to $G_n(\cdot)$, where $G_n(\cdot)$ is the empirical distribution function putting mass $1/n$ at each observation y_i.

   The definition from BNC makes the wording confusing; "from F to G" means $\int \log(dG/dF)dG$. See handwritten notes for Nov 6.

2. Suppose $y_i \sim N(\mu_i, 1/n), i = 1, \dots, k$ and $\psi^2 = \Sigma_{i=1}^k \mu_i^2$ is the parameter of interest.[1]

   (a) Show that the marginal posterior density for $n\psi^2$, assuming a flat prior $\pi(\underline{\mu}) \propto 1$, is a non-central χ^2_k distribution, with non-centrality parameter $n\Sigma y_i^2$.

   (b) Show that the maximum likelihood estimate of ψ^2 is $\hat{\psi}^2 = \Sigma y_i^2$, and that $n\hat{\psi}^2$ has a non-central χ^2_k distribution with non-centrality parameter $n\psi^2$.

[1] It will be convenient to use $\lambda_i = \mu_i/\sqrt{\Sigma\mu_i^2}$ for the nuisance parameters.

(c) Compare the normal approximations to $r_u(\psi)$, $r_e(\psi)$ and $r(\psi)$ with the exact distribution of the maximum likelihood estimate.

(d) Compare the 95% Bayesian posterior probability interval for ψ^2, based on (a) to the 95% confidence interval for ψ^2, based on (b).

This question is awfully vague. First need to show that $\hat{\lambda}_i = y_i/||y||$, and further that $\hat{\lambda}_{i,\psi} = \hat{\lambda}_i$ (not y_i/ψ as was claimed in class). With this in hand we can show that $\ell_p(\psi) = n\psi||y|| - n\psi^2/2$, and $r_p = r_e = r$. We have 3 p-value functions: based on the normal approximation to any of the pivots, the exact frequentist distribution of $\hat{\psi}$ and the Bayes posterior for ψ:

$$\begin{aligned} p_{norm}(\psi) &= \Pr\{N(0,1) \geq \sqrt{n}(||y|| - \psi)\} \\ p_{exact}(\psi) &= \Pr\{\chi^2_k(n\psi^2) \geq n||y||^2\} \\ p_{Bayes}(\psi) &= \Pr\{\chi^2_k(n||y||^2) \leq n\psi^2\}. \end{aligned}$$

(The Bayes posterior is switched so that all the curves are descending.) The somewhat ugly plots on the next page illustrate convergence of both as $n \to \infty$, fixed k, but not (of course) as $k \to \infty$, fixed n.

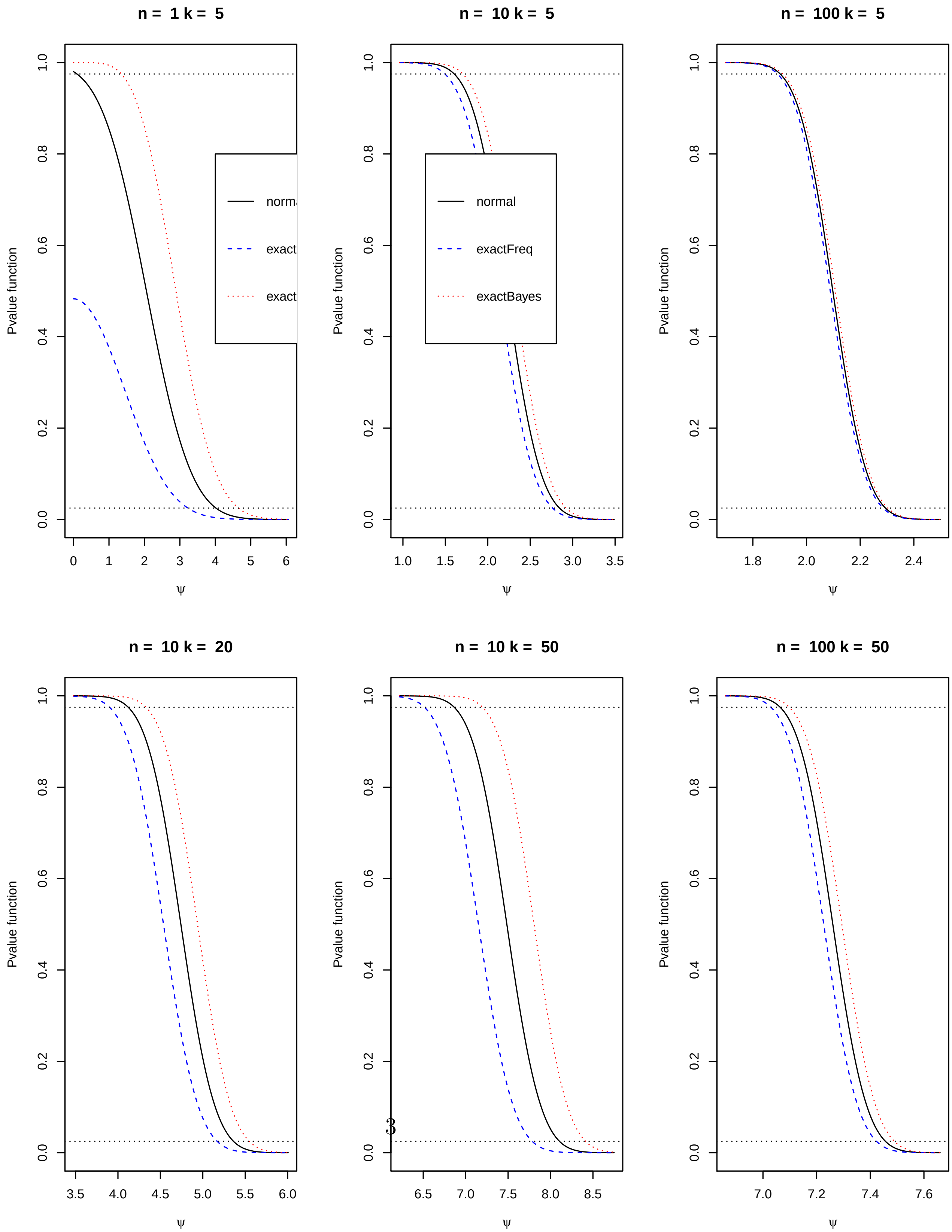
```
p1= function(psi){pnorm(sqrt(n)*(normy-psi))}
p2=function(psi){pchisq(n*normy^2,df=k,ncp=n*psi^2)}
p3=function(psi){1-pchisq(n*psi^2,df=k,ncp=n*normy^2)}

n=10
k=5

y = rnorm(k,1,sqrt(1/n)) #lazy choice of mu

normy = sqrt(sum(y^2))
lower = normy-4/sqrt(n)
upper = normy+4/sqrt(n)
# should be a good range for psi, but it shouldn't go negative
psivals=seq(max(0,lower),upper,length=100)

plot(psivals,p1(psivals),type="l",main = paste("n = ", n, "k = ", k),
     ylab = "Pvalue function", xlab = expression(psi),ylim=c(0,1))
lines(psivals,p2(psivals),col="blue",lty=2)
lines(psivals,p3(psivals),col="red",lty=3)
legend(upper-normy,.8, c("normal","exactFreq","exactBayes"),
       lty=c(1,2,3),col=c("black","blue","red"))
#legend placement is not always ideal
```



n = 1 k = 5
n = 10 k = 5
n = 100 k = 5
n = 10 k = 20
n = 10 k = 50
n = 100 k = 50
Pvalue function
ψ
normal
exactFreq
exactBayes