Statistics - Homework 7 (Due June 14, 2019)

1. The Pareto Distribution is defined by the following density function:

   $$f(x \mid \alpha, k) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}}, \quad k \le x < +\infty, \quad \alpha > 0, \quad k > 0 .$$

   Given a random sample of size *n* and *k*= min(x_1, …, x_n), find the maximum likelihood estimator of α.
2. A random sample of size n_1 from a normal distribution with mean μ_1 and standard deviation σ_1 is available. Another random sample of size n_2 from a normal distribution with mean μ_2 and standard deviation σ_2 is also available. The two random samples are independent.
   (1) Find a maximum likelihood estimator of $\theta = \mu_1 - \mu_2$.
   (2) If the sum of n_1 and n_1 ($n = n_1 + n_2$) is fixed, how should the sizes of the two random samples be chosen in order to minimize the variance of the maximum likelihood estimator in (1).
3. Let $X_1, X_2, \cdots, X_n$ be independently and identically distributed (IID) from a continuous uniform distribution $U[0, \theta]$.
   (1) Find a maximum likelihood estimator of θ, and then find the mean and mean-squared error of your estimator.
   (2) Find a method of moments estimator of θ, and then find the mean and mean-squared error of your estimator.
4. The radius (*r*) of a circle is measured with an error of measurement which is normally distributed with zero mean and unknown variance σ^2. Given the following independent measurements of the radius:

51.07	52.49	48.07	48.94	49.33
50.34	51.30	40.33	41.38	44.05
51.45	48.89	56.42	49.21	49.41
46.03	53.64	41.13	46.17	57.07

   Find an unbiased estimator of the area of the circle and use the above data to calculate an unbiased estimate of the circle area.
5. Let *X* be a normally distributed random variable with expected value and standard deviation being 60 and 20, respectively. Let $\overline{X}_n$ be the sample mean of a random sample of size *n* from *X*. A random sample of size 25 from *X* is given in the following table:

37.3332	84.75534	56.2749	27.09361	63.11717
44.61746	46.38288	73.65585	91.7605	50.46811
86.2026	51.86157	78.05359	33.82873	75.01817
52.57203	19.59978	80.21883	72.44076	42.92938
61.5187	68.02203	68.10625	81.53383	60.46798

   (i) Determine a 95% confidence interval of the population mean of *X* using the above data. (Assuming the standard deviation of *X* is unknown.)
   (ii) Find the 95% acceptance interval of $\overline{X}_n$ with sample size *n*=25.
6. A random variable *X* has a gamma density with expected value 180 and standard deviation 104.
   (1) Generate 10,000 random samples (size *n*=100) from the above density. From the 10,000

samples, plot the corresponding 95% confidence intervals of the expected value of *X*. [Assuming the variance of *X* is unknown.]

(2) Among the 10,000 confidence intervals, how many of them actually contain the expected value of *X* ?

(3) What is the average length of the 10,000 confidence intervals?

An R code at the end of this file is provided for your reference.

7. The data in the following table is a random sample from an exponential density $f_X(x;\ \lambda) = \lambda e^{-\lambda x}$.

37.01	52.20	7.55	2.31	29.48	111.70	39.63	3.89	201.60	20.60
41.23	18.69	241.49	4.12	15.67	6.74	16.72	24.86	9.91	115.26
16.00	27.65	66.90	21.81	4.92	46.05	58.79	65.23	7.79	51.24
163.89	9.49	61.21	181.03	71.65	74.64	43.06	13.57	35.89	136.47

(1) Find a two-sided 90% confidence interval of λ.

(2) Find the lower bound (*a*) of the one-sided 95% confidence interval of λ, i.e. $P[a \le \lambda] = 0.95$.

8. A random sample of size 20 ($(x_1, x_2, \cdots, x_n)$, $n = 20$) drawn from a uniform density U[0, θ] is shown in the following table.

11.90	6.94	14.90	4.36	14.00
15.21	18.87	4.75	7.67	6.57
5.15	14.83	7.79	13.88	18.55
2.49	9.90	9.68	11.10	9.92

(1) Show that Y_n/θ ($Y_n = \max(X_1,\ldots,X_n)$) is a pivotal quantity.

(2) Find a 90% confidence interval of θ.

Reference R code for problem 6

```
# Interval estimation -- Normal Density
# Demonstration of confidence interval estimation for mean of a normal density
# ----------------------------------------------------------
ns=10
nr=100
siglevel=0.05
mx=c();stdx=c();lb=c();ub=c()
a=16 #location parameter
s=5  #scale parameter
exp.x=a
d=c() # 2d = length of the confidence interval
count=0
windows()
plot(0,0,type="n",xlim=c(a-4*s/sqrt(ns),a+4*s/sqrt(ns)),ylim=c(0,2*nr))
for (i in 1:nr)
{
      x=rnorm(ns,a,s)
      mx[i]=mean(x);stdx[i]=sqrt(var(x))
      d[i]=qt((1-siglevel/2),ns-1)*stdx[i]/sqrt(ns)
      lb[i]=mx[i]-d
      ub[i]=mx[i]+d
      if (lb[i]>exp.x |ub[i]<exp.x) count=count+1
      u=c(lb[i],ub[i])
      v=c(2*i,2*i)
      lines(u,v,type="l",lwd=2)
}
lines(c(a,a),c(0,2*nr),type="l",col="red",lwd=2)
print(c(count,count/nr,mean(2*d)))
```