University of California, Santa Cruz
Department of Applied Mathematics and Statistics
Baskin School of Engineering
Classical and Bayesian Inference - AMS 132

Homework 4

Instructions: You have until Friday, March 2, to complete the assignment. It has to be returned during 10 first minutes of class (4:55 pm to 5:05 pm) or between 1:00 pm and 3:00 pm in office BE 357B.

1. Suppose that $X_1, \dots, X_n$ form a random sample from the normal distribution with mean μ and variance σ^2, and let $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Determine the smallest values of n for which the following relations are satisfied:

   (a) $P\left(\frac{\widehat{\sigma}^2}{\sigma^2} \leq 1.9\right) \geq 0.95$.
   Since $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, by the theorem of joint distribution of the sample mean and sample variance, it follows that $Y = \sum_{i=1}^n \left(\frac{X_i - \overline{X}_n}{\sigma}\right)^2 \sim \chi^2_{(n-1)}$. Now

   $$P\left(\frac{\widehat{\sigma}^2}{\sigma^2} \leq 1.9\right) = P(\frac{Y}{n} \leq 1.9) = P(Y \leq 1.9n).$$

   In order to $P(Y \leq 1.9n) \geq 0.95$, wee need to find n such that the 0.95 quantile of a $\chi^2_{(n-1)}$ is $1.9n$. After trying different values of n, it follows that $n = 5$ is such that $P(Y \leq 1.9 * 5) = 0.9502 \geq 0.95$, where $Y \sim \chi^2_{(4)}$.

   (b) $P\left(\mid \widehat{\sigma}^2 - \sigma^2 \mid \leq \frac{1}{2}\sigma^2\right) \geq 0.8$.
   Since $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, by the theorem of joint distribution of the sample mean and sample variance, it follows that $Y = \sum_{i=1}^n \left(\frac{X_i - \overline{X}_n}{\sigma}\right)^2 \sim \chi^2_{(n-1)}$. Now

   $$P\left(\mid \widehat{\sigma_0}^2 - \sigma^2 \mid \leq \frac{1}{2}\sigma^2\right) = P(-1/2 \leq \sigma_0^2/\sigma^2 - 1 \leq 1/2) = P(1/2 \leq \sigma_0^2/\sigma^2 \leq 3/2)$$

   $$= P(n/2 \leq Y \leq 3n/2),$$

   where $Y \sim \chi^2_{(n-1)}$. After trying different values for n, it follows that $n = 13$ is such that $P(n/2 \leq Y \leq 3n/2) = 0.8116 \geq 0.8$, where $Y \sim \chi^2_{(12)}$.

2. Suppose that the five random variables $X_1, \dots, X_5$ are i.i.d. and that each has the standard normal distribution. Determine a constant c such that the random variable

   $$\frac{c(X_1 + X_2)}{(X_3^2 + X_4^2 + X_5^2)^{1/2}},$$

will have the t distribution.

Since $X_1, \ldots, X_5$ are i.i.d. random variables having the standard normal distribution, it follows that $X_1 + X_2 \sim N(0, 2)$, and $X_i^2 \sim \chi^2_{(1)}$, for $i = 3, 4, 5$, and they are independent. This implies that $\frac{X_1+X_2}{\sqrt{2}} \sim N(0, 1)$ and $X_3^2 + X_4^2 + X_5^2 \sim \chi^2_{(3)}$, and they are independent. Finally, we get the t distribution

$$\frac{\frac{X_1+X_2}{\sqrt{2}}}{\sqrt{\frac{X_3^2+X_4^2+X_5^2}{3}}} = \sqrt{\frac{3}{2}}\frac{c(X_1 + X_2)}{(X_3^2 + X_4^2 + X_5^2)^{1/2}} \sim t_{(3)}.$$

Therefore, the constant c is equal to $\sqrt{\frac{3}{2}}$.

3. Suppose that a random sample of eight observations is taken from the normal distribution with unknown mean μ and unknown variance σ^2, and that the observed values are $3.1, 3.5, 2.6, 3.4, 3.8, 3.0, 2.9$, and 2.2. Find a confidence interval for μ with each of the following three confidence coefficients: (a) 0.90, (b) 0.95, and (c) 0.99. What effect has the confidence coefficient in the size of the interval?

   Since the variance of the random variables is unknown, a coeffcient gamma confidence interval for μ is given by

$$\left(\overline{X}_n - T^{-1}_{n-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma'}{\sqrt{n}}, \overline{X}_n + T^{-1}_{n-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma'}{\sqrt{n}}\right).$$

   From the data we have that $\overline{x}_n = 3.06$, $n = 8$, and $\sigma' = 0.5125$, also, $T^{-1}_{n-1}\left(\frac{1+0.9}{2}\right) = 1.8945$, $T^{-1}_{n-1}\left(\frac{1+0.95}{2}\right) = 2.36462$, $T^{-1}_{n-1}\left(\frac{1+0.99}{2}\right) = 3.4994$. Therefore, for $\gamma = 0.9$ the confidence interval is $(2.80, 3.31)$, for $\gamma = 0.95$ the confidence interval is $(2.75, 3.37)$, and for $\gamma = 0.99$ the confidence interval is $(2.62, 3.49)$. Therefore, the larger the confidence coefficient, the longer the interval, this is, less precise it is.

4. Suppose that $X_1, ..., X_n$ form a random sample from the normal distribution with unknown mean μ and unknown variance σ^2, and let the random variable L denote the length of the confidence interval for μ that can be constructed from the observed values in the sample. Find the value of $E(L^2)$ for the following values of the sample size n and the confidence coefficient γ:

   First we will find $E(L^2)$. Since the variance of the random variables is unknown, the length of the interval is $L = 2T^{-1}_{n-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma'}{\sqrt{n}}$, so $L^2 = 4\left[T^{-1}_{n-1}\left(\frac{1+\gamma}{2}\right)\right]^2\frac{\sigma'^2}{n}$. To find $E(L^2)$ we need $E(\sigma'^2)$. For this, notice that $E\left(\sum(X_i - \overline{X}_n)^2/\sigma^2\right) = n - 1$. This implies that

$$E(\sigma'^2) = E\left(\frac{\sum_{i=1}^n(X_i - \overline{X}_n)^2}{n-1}\right) = \frac{\sigma^2}{n-1}E\left(\sum(X_i - \overline{X}_n)^2/\sigma^2\right) = \sigma^2.$$

   Therefore,

$$E(L^2) = 4\left[T^{-1}_{n-1}\left(\frac{1+\gamma}{2}\right)\right]^2\frac{E(\sigma'^2)}{n} = 4\left[T^{-1}_{n-1}\left(\frac{1+\gamma}{2}\right)\right]^2\frac{\sigma^2}{n}$$

- for $\gamma = 0.95$: use $n = 5$, $n = 10$, and $n = 30$. For fixed γ, what effect has n in the size of the interval?

  For $\gamma = 0.95$, $E(L^2) = 6.16\sigma^2$ if $n = 5$, $E(L^2) = 2.04\sigma^2$ if $n = 10$, and $E(L^2) = 0.55\sigma^2$ if $n = 30$. So, for fixed γ coefficient, the confidence interval is more precise as the sample size increases.

- For $n = 8$: , use $\gamma = 0.90$,$\gamma = 0.95$, and $\gamma = 0.99$. For fixed n, what effect has the confidence coefficient γ?
  For $n = 8$, $E(L^2) = 1.79\sigma^2$ if $\gamma = 0.90$, $E(L^2) = 2.79\sigma^2$ if $\gamma = 0.95$, and $E(L^2) = 6.12\sigma^2$ if $\gamma = 0.99$. So, for fixed sample size n , the confidence interval is less precise as the coefficient γ increases.

Note: $E(L^2)$ will be a function of σ^2. For solving the exercise: first find L and show that $L^2 = 4c^2\sigma'^2/n$, where $\sigma'^2 = \sum(X_i - \overline{X}_n)^2/(n-1)$. Second, note that $W = \sum(X_i - \overline{X}_n)^2/\sigma^2$ has a χ square distribution with $n-1$ degrees of freedom, whose mean is $n-1$.

5. Suppose that $X_1, ..., X_n$ form a random sample from the exponential distribution with unknown mean μ. Find a 90 percent confidence interval for μ. Use $\gamma_1 = 0.05$ and $\gamma_2 = 0.95$
   Note: use the facts that: if $X \sim exp(\lambda)$, then $\sum_{i=1}^n X_i \sim Gamma(n, \lambda)$. If $X \sim Gamma(a, b)$, then $2Xb \sim \chi^2_{(2a)}$. Then, show that $\frac{2\sum_{i=1}^n X_i}{\mu}$ is a pivot quantity.

   Since $X \sim exp(1/\mu)$, then $\sum_{i=1}^n X_i \sim Gamma(n, 1/\mu)$. Now, since $\sum_{i=1}^n X_i \sim Gamma(n, 1/\mu)$, then $\frac{2\sum_{i=1}^n X_i}{\mu} \sim \chi^2_{2n}$. Therefore, $\frac{2\sum_{i=1}^n X_i}{\mu}$ is a function of the random variables and the parameter μ that has a distribution that is the same for every μ, so it is a pivot. Now, for finding the confidence interval we start from $P\left(G^{-1}(\gamma_1) < \frac{2\sum_{i=1}^n X_i}{\mu} < G^{-1}(\gamma_2)\right) = \gamma$, where $G^{-1}(\gamma)$ is the γ quantile of a χ^2 distribution with $2n$ degrees of freedom, and find random variables A and B such that $P(A < \mu < B) = \gamma$. Note that

   $$P\left(G^{-1}(\gamma_1) < \frac{2\sum_{i=1}^n X_i}{\mu} < G^{-1}(\gamma_2)\right) = \gamma$$

   is equivalent to

   $$P\left(\frac{1}{G^{-1}(\gamma_1)} > \frac{\mu}{2\sum_{i=1}^n X_i} > \frac{1}{G^{-1}(\gamma_2)}\right) = \gamma$$

   , which is equivalent to $P\left(\frac{2\sum_{i=1}^n X_i}{G^{-1}(\gamma_2)} < \mu < \frac{2\sum_{i=1}^n X_i}{G^{-1}(\gamma_1)}\right) = \gamma$. Therefore, the coefficient γ confidence interval for μ is given by (A, B), where $A = \frac{2\sum_{i=1}^n X_i}{G^{-1}(\gamma_2)}$ and $B = \frac{2\sum_{i=1}^n X_i}{G^{-1}(\gamma_1)}$, where $\gamma_1 = 0.05$ and $\gamma_2 = 0.95$.

6. In the June 1986 issue of *Consumer Reports*, some data on the calorie content of beef hot dogs is given. Here are the numbers of calories in 20 different hot dog brands:
   $186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111, 141, 153, 190, 157, 131, 149, 135, 132.$

Assume that these numbers are the observed values from a random sample of twenty independent normal random variables with mean μ and variance σ^2, both unknown. Find a 90% confidence interval for the mean number of calories μ.

Since the variance is unknown the confidence interval is

$$\left(\overline{X}_n - T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma'}{\sqrt{n}}, \overline{X}_n + T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma'}{\sqrt{n}}\right).$$

From the data we have that $\overline{x}_n = 156.85$, $n = 20$, $\sigma' = 22.6420$, and $T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) = 1.72913$, so the confidence interval is $(148.0956, 165.6044)$. So we can say that with a 90% confidence the mean calories in hot dogs is between 148.0956 and 165.6044.

7. Consider the problem and data from exercise 6, but now assume that the variance is known and equal to 510.

   (a) Find a 90% confidence interval for the mean number of calories μ. You can use the result of exercise 1 in section 8.5 from the textbook (a confidence interval in this case was also solved in discussion section).

   Since the variance of the random variables is known, a 90% confidence interval for the mean number of calories μ is given by

   $$\left(\overline{X}_n - \Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma'}{\sqrt{n}}, \overline{X}_n + \Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma'}{\sqrt{n}}\right).$$

   So, from the information in exercise 6, the observed value of the interval is $(148.5439, 165.1561)$. Note that this interval has a smaller length than the one from exercise 6. This is because now the variance is known, there is less uncertainty.

   (b) Now, assume that μ has a normal prior distribution with mean 0 and variance 10000. Find the posterior distribution of μ and compute the probability that the posterior distribution of μ lies between the confidence interval computed in a).

   Here we are assuming that $\mu \sim N(0, 10000)$ and $X_i \sim N(\mu, 510)$, so from the theorem of conjugate prior distributions, it follows that the posterior distribution of μ is normal with mean equal to $\mu_1 = \frac{10000n\overline{x}_n}{510+10000n}$ and variance $V_1^2 = \frac{510*10000}{510+10000n}$. The probability that the posterior distribution of μ lies between the confidence interval computed in a) is

   $$P(148.5439 < \mu < 165.1561 \mid \boldsymbol{x}) = P(\mu < 165.1561 \mid \boldsymbol{x}) - P(\mu < 148.5439 \mid \boldsymbol{x}) = 0.899372,$$

   where $\mu \mid \boldsymbol{x} \sim N(\mu_1, v_1^2)$.

```
mu1 <- (20*10000*mean(x))/(510+20*10000)
v12 <- (510*10000)/(510+20*10000)
pnorm(165.1561, mu1, sqrt(v12))-  pnorm(148.5439, mu1, sqrt(v12))
```

So under the prior $\mu \sim N(0, 10000)$ the posterior probability of μ being inside the confidence interval is very close to the confidence coefficient, 90%. This is because, this prior has a very large variance.

(c) What kind of prior distribution was considered in b)?

The prior distribution considered in b) is a conjugate prior distribution.