STATISTICS: Problem Set 3-Solutions

October 3, 2007

1. Consider an even number of independent Bernoulli-distributed observations where the two subsamples $Y_1, Y_2, .., Y_{n/2}$ and $Y_{n/2+1}, .., Y_n$ have parameters p_1 and p_2 respectively denoting $P(Y_i = 1)$.

(a) Work out the log-likelihood function for the whole sample, and show how you might construct the maximum likelihood estimators of p_1 and p_2.

(b) Using your result in (a) above, show that $Z_1 = \frac{2}{n}\sum_{i=1}^{n/2} Y_i \longrightarrow p_1$ and $Z_2 = \sum_{i=n/2+1}^{n} Y_i \longrightarrow p_2$, where $\longrightarrow$ denotes "convergence in probability".

(c) Combine the results in (b) to show that $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \longrightarrow \frac{p_1+p_2}{2}$, and find the maximum likelihood estimate of the parameter p for the whole sample(without taking into account the change in p).

(d) Comment briefly on why the Law of Large Numbers cannot be used for analyzing $\bar{Y}$ directly.

(e) Describe (but do not solve in detail) a likelihood ratio test of the null hypothesis $H_0 : p_1 = p_2$ against the alternative hypothesis $H_A : p_1 \neq p_2$.

(a) Recalling Exercise 6, part (b), in the Problem set 1, the likelihood function for the whole sample is

$$L(p_1, p_2) = p_1^{\sum_{i=1}^{n/2} y_i}(1-p_1)^{n/2-\sum_{i=1}^{n/2} y_i} p_2^{\sum_{i=n/2+1}^{n} y_i}(1-p_2)^{n/2-\sum_{i=n/2+1}^{n} y_i}$$

where $y_i = 0$ or 1.

The corresponding log-likelihood function is

$$lnL(p_1, p_2) = ln(p_1)\sum_{i=1}^{n/2} y_i + ln(1-p_1)(n/2 - \sum_{i=1}^{n/2} y_i) + ln(p_2)\sum_{i=n/2+1}^{n} y_i + ln(1-p_2)(n/2 - \sum_{i=n/2+1}^{n} y_i).$$

Differentiating lnL with respect to p_1 and with respect to p_2 and setting both derivatives equal to zero, we obtain

$$\hat{p}_1 = \frac{\sum_{i=1}^{n/2} y_i}{n/2},$$
$$\hat{p}_2 = \frac{\sum_{i=n/2+1}^{n} y_i}{n/2}.$$

(b) $\{Y_i\}_{i=1}^{n/2}$ from the first subsample are *independent identically distributed* random variables with $E(Y_i) = p_1$. Then by the Law of Large Numbers,

$$Z_1 = \frac{\sum_{i=1}^{n/2} Y_i}{n/2} \xrightarrow{P} p_1. \tag{1}$$

Similarly, all $\{Y_i\}_{i=n/2+1}^{n}$ are *i.i.d.* with $E(Y_i) = p_2$, so,

$$Z_2 = \frac{\sum_{i=n/2+1}^{n} Y_i}{n/2} \xrightarrow{P} p_2. \tag{2}$$

(1) and (2) mean that

$$\hat{p}_1 \xrightarrow{P} p_1$$
$$\hat{p}_2 \xrightarrow{P} p_2$$

This suggests that estimator $\hat{p_1}$ is a *consistent* estimator of p_1 and $\hat{p_2}$ is a *consistent* estimator of p_2.

(c) Recall the properties of convergence in probability:

$$plim(X+Y) = plimX + plimY,$$
$$plim(\alpha X) = \alpha plimX, \qquad \text{where } \alpha = const.$$

Then using these properties and the results of (b),

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\left[\frac{n}{2}Z_1 + \frac{n}{2}Z_2\right] = \frac{1}{2}[Z_1 + Z_2] \xrightarrow{P} \frac{1}{2}[p_1 + p_2].$$

Note, that if we used the maximum likelihood estimate of p over the whole sample ignoring the difference between p_1 and p_2, we would get the estimate

$$\hat{p} = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{Y}.$$

(d) When we consider the full sample, $Y_1, .., Y_n$, the assumptions of the Law of Large Numbers

are violated and the Law of Large Numbers cannot be used.
(e) Recalling again Exercise 6 (part (c)) in the Problem set 1, the likelihood-ratio test of $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$ uses the likelihood-ratio statistics Λ:

$$0 \leq \Lambda = \frac{\max L(p_1, p_2 | p_1 = p_2 = p)}{\max L(p_1, p_2 | p_1 \neq p_2)} \leq 1$$

To compute the numerator, we use the *restricted* ML estimate of p_1, p_2 ($\hat{p}_1 = \hat{p}_2 = \hat{p} = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{Y}$). To compute the denominator, we use the *unrestricted* ML estimate of p_1 and p_2 ($\hat{p}_1 = \frac{\sum_{i=1}^{n/2} y_i}{n/2}$, $\hat{p}_2 = \frac{\sum_{i=n/2+1}^{n} y_i}{n/2}$).

Then we compute the statistic LR$=-2log\Lambda$. It has approximate $\chi^2[1]$-distribution.

A statistical test can now be constructed as a decision rule. If Λ is close to 1, and correspondingly, LR is small, the restricted maximum likelihood estimate would be (nearly) as likely as the unrestricted estimate, so in this case we would fail to reject the hypothesis. We therefore choose a critical value $c > 0$, and reject the hypothesis if $LR > c$.

2. The standardized bivariate normal distribution of two random variables X and Y with a correlation of ρ ($|\rho| < 1$) is characterized by the density:

$$f_{X,Y}(x, y; \rho) = (2\pi)^{-1}(1 - \rho^2)^{-\frac{1}{2}} exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1 - \rho^2)}\right],$$

for $-\infty < x, y < +\infty$.
(a) What are $E(XY)$ and $E(Y|X = x)$?
(b) Show that X and Y are independent if $\rho = 0$.
(c) Derive $log(f_{X,Y}(x, y; \rho))$ and differentiate this function with respect to ρ. Can you solve this result for ρ?

Before answering the questions in (a), (b), and (c), note that the numerator in $[\cdot]$ can be rewritten as $[(y - x)^2 + (1 - \rho^2)x^2]$. Since $exp[a + b] = exp(a)exp(b)$,

$$f_{X,Y}(x, y) = (2\pi)^{-1}(1 - \rho^2)^{-\frac{1}{2}} exp\left[-\frac{(y - \rho x)^2}{2(1 - \rho^2)}\right] exp\left[-\frac{(1 - \rho^2)x^2}{2(1 - \rho^2)}\right]. \quad (3)$$

(a)
The fast way to find $E(XY)$ is to notice that by definition of the correlation of two random variables,

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{var(X)}\sqrt{var(Y)}}.$$

Since both variables, X and Y, have standard normal distribution,

$$\rho = E(XY).$$

Alternatively, $E(XY)$ can be found using the direct computation of the following integral:

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_{X,Y}(x,y)dydx = \\ &= (2\pi)^{-1}(1-\rho^2)^{-\frac{1}{2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyexp\left[-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right]exp\left[-\frac{(1-\rho^2)x^2}{2(1-\rho^2)}\right]dydx = \\ &= (2\pi)^{-1/2}\int_{-\infty}^{+\infty} xexp\left[-\frac{x^2}{2}\right]\left(\int_{-\infty}^{+\infty} y(2\pi)^{-1/2}(1-\rho^2)^{-\frac{1}{2}}exp\left[-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right]dy\right)dx \end{aligned}$$

The term multiplying y inside the internal integral is the density function of a normal random variable with mean ρx and standard deviation $(1-\rho^2)^{1/2}$. Hence, the internal integral is the expectation of this random variable that is it is equal to ρx.

Then we obtain:

$$E(XY) = (2\pi)^{-1/2}\rho\int_{-\infty}^{+\infty} x^2 exp\left[-\frac{x^2}{2}\right]dx \qquad (4)$$

Changing the variable of integration from x to u such that

$$u = x^2$$
$$du = 2xdx$$
and so
$$x = \pm\sqrt{u},$$

we obtain:

$$\begin{aligned} \int_{-\infty}^{+\infty} x^2 exp\left[-\frac{x^2}{2}\right]dx &= \int_{-\infty}^{0} x^2 exp\left[-\frac{x^2}{2}\right]dx + \int_{0}^{+\infty} x^2 exp\left[-\frac{x^2}{2}\right]dx = \\ &= 2\int_{0}^{+\infty} x^2 exp\left[-\frac{x^2}{2}\right]dx = \int_{0}^{+\infty}\sqrt{u}exp\left[-\frac{u}{2}\right]du. \end{aligned} \qquad (5)$$

Recall that

$$f(x,k) = \begin{cases} \frac{1}{2^{k/2}\Gamma(\frac{k}{2})}x^{k/2-1}exp\left[-\frac{x}{2}\right] & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

is a p.d.f. of $\chi^2(k)$ distribution. For $k = 3$ and $x > 0$

$$f(x,3) = \frac{1}{2^{3/2}\Gamma(\frac{3}{2})}x^{1/2}exp\left[-\frac{x}{2}\right].$$

By definition of the p.d.f.

$$\int_{0}^{+\infty} f(x,3)dx = 1.$$

Therefore,

$$\int_{0}^{+\infty} x^{1/2}exp\left[-\frac{x}{2}\right]dx = 2^{3/2}\Gamma\left(\frac{3}{2}\right) = 2^{3/2}\frac{\pi^{1/2}}{2} = (2\pi)^{1/2}.$$

Using this result for our integral in (5), we get

$$\int_0^{+\infty} \sqrt{u} exp\left[-\frac{u}{2}\right] du = (2\pi)^{1/2},$$

and so,

$$E(XY) = (2\pi)^{-1/2}\rho(2\pi)^{1/2} = \rho.$$

Now, compute $E(Y|X = x)$.

$$E(Y|X = x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y, x)dy = \int_{-\infty}^{+\infty} y\frac{f_{Y,X}(y, x)}{f_X(x)}dy.$$

We need to find the marginal density of X.

$$f_X(x) = \int_{-\infty}^{+\infty} f_{Y,X}(y, x)dy$$

Using representation of $f_{Y,X}$ in (3), we get:

$$f_X(x) = (2\pi)^{-1/2}exp\left[-\frac{x^2}{2}\right]\int_{-\infty}^{+\infty}(2\pi)^{-1/2}(1-\rho^2)^{-\frac{1}{2}}exp\left[-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right]dy \qquad (6)$$

Once again, the term inside the integral is the density function of a normal random variable with mean ρx and standard deviation $(1 - \rho^2)^{1/2}$. Hence, this integral is unity. So,

$$f_X(x) = (2\pi)^{-1/2}exp\left[-\frac{x^2}{2}\right]$$

and $X \sim N(0, 1)$.

Then

$$f_{Y|X}(y, x) = (2\pi)^{-1/2}(1-\rho^2)^{-\frac{1}{2}}exp\left[-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right].$$

This is exactly the density function inside the integral in (6), so that

$$(Y|X = x) \sim N(\rho x, (1 - \rho^2))$$

and thus,

$$E(Y|X = x) = \rho x.$$

(b) If $\rho = 0$, then

$$f_{X,Y}(x, y) = (2\pi)^{-1}exp\left[-\frac{x^2 + y^2}{2}\right].$$

From (a) we know that

$$f_X(x) = (2\pi)^{-1/2}exp\left[-\frac{x^2}{2}\right].$$

Since X and Y enter $F_{X,Y}$ symmetrically, the same argument as in (a) would lead to the same marginal density for Y:

$$f_Y(y) = (2\pi)^{-1/2} exp\left[-\frac{y^2}{2}\right].$$

So, we have

$$f_{X,Y}(x,y) = f_X(x) f_Y(y).$$

Hence, X and Y are independent.

(c)

$$log(f_{X,Y}(x,y;\rho)) = log((2\pi)^{-1}) - \frac{1}{2} log(1-\rho^2) - \frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}$$

$$\begin{aligned}\frac{\partial log(f_{X,Y}(x,y;\rho))}{\partial \rho} &= \frac{\rho}{1-\rho^2} - \frac{-2(1-\rho^2)2xy + (x^2 - 2\rho xy + y^2)4\rho}{4(1-\rho^2)^2} = \\ &= \frac{\rho}{1-\rho^2} + \frac{xy}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2}(x^2 - 2\rho xy + y^2).\end{aligned}$$

Find the ML estimate of ρ – equate $\frac{\partial log(f_{X,Y}(x,y;\rho))}{\partial \rho}$ with zero:

$$\begin{aligned}(\rho + xy)(1-\rho^2) - \rho(x^2 - 2\rho xy + y^2) &= 0 \\ \rho^3 - \rho^2 xy - \rho(1 - y^2 - x^2) - xy &= 0\end{aligned}$$

This is a cubic polynomial in ρ. The two roots of this polynomial are complex (not real) and therefore, ruled out as a solution. The other one is a solution if one can show that it is inside the interval $[-1,1]$ (since ρ is a correlation).

3. $[X,Y]' \sim f(x,y) = \frac{1}{\pi}\exp(-\frac{1}{2}(x^2+y^2))\mathbf{1}_A(x,y)$, $A = \{(x,y)|\ xy > 0\}$. Find marginal df of X and Y. What general conclusions can you draw from the exercise and the fact, that marginal distributions of a bivariate Normal are Normal?

First, compute the marginal distribution for $x > 0$

$$\begin{aligned}f_X(x) &= \int_0^\infty \frac{1}{\pi}\exp[-\frac{1}{2}(x^2+y^2)]\,dy = \frac{1}{\pi}\exp(-x^2/2)\int_0^\infty \exp(-y^2/2)\,dy = \\ &= \frac{1}{\pi}\exp(-x^2/2)\sqrt{2\pi}\int_0^\infty \frac{1}{\sqrt{2\pi}}\exp(-y^2/2)\,dy = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\end{aligned}$$

Similarly for $x \leqslant 0$ we have

$$\begin{aligned}f_X(x) &= \int_{-\infty}^0 \frac{1}{\pi}\exp[-\frac{1}{2}(x^2+y^2)]\,dy = -\int_0^\infty \frac{1}{\pi}\exp[-\frac{1}{2}(x^2+y^2)]\,dy = \\ &\int_0^\infty \frac{1}{\pi}\exp[-\frac{1}{2}(x^2+(-y)^2)]\,dy = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\end{aligned}$$

We can see, therefore, that marginal distribution of X is $N(0,1)$. The same holds for Y. Therefore, we can see that not only bivariate Normal distribution can have Normal marginal distributions, but also a non Normal bivariate distribution. More generally we can say, that marginals do not define joint distribution uniquely.

4. A sample of n random variables $\{X_i\}$ is independently distributed as $N(\mu_X, \sigma_X^2)$.

a) Let $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ denote the sample average. Show that $E(\bar{X}) = \mu_X$, $E(\bar{X} - \mu_X)^2 = \frac{\sigma_X^2}{n}$ and hence that $\bar{X} \sim N(\mu_X, \sigma_X^2/n)$; discuss how this finding helps estimate μ_X when it is unknown.

b) Let $\nu^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ be the sample variance of X around the sample mean. Show that $E(\nu^2) = \sigma_X^2$ and that $(n-1)\nu^2/\sigma_X^2$ is distributed as chi square with (n-1) degrees of freedom.

c) In general if variables $\{X_i\}$ are iid from a distribution with finite variance, and if $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, then $cov(\bar{X}, X_i - \bar{X}) = 0$ for any $i = 1, 2, \ldots, n$. Infer from this, that mean and sample variance of an iid sample from $N(\mu, \sigma^2)$ are independent, and thus the statistic $t = \frac{\bar{X} - \mu_X}{\sqrt{\nu^2}}$ has Student's t distribution with $(n-1)$ degrees of freedom.

d) Build an exact, small-sample confidence interval for μ_X.

a) By definition, $E(X_i) = \mu_X$ $\forall i \in 1 : n$. Then, by linearity of the expectation operator, we have

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sum_{i=1}^n E(X_i)}{n} = \frac{n\mu_X}{n} = \mu_X.$$

When μ_X is unknown, it can be estimated using ML estimation. One can check that for the normal distribution, ML estimate of μ_X is

$$\hat{\mu}_X = \frac{\sum_{i=1}^n X_i}{n} = \bar{X},$$

that is a sample average. Then the finding that $E(\bar{X}) = \mu_X$ tells us that the ML estimate of μ_X is *unbiased* and *consistent*. Note that consistency follows from the Weak Law of Large Numbers: since

$$\bar{X} \xrightarrow{P} E(X_i) = \mu_X,$$

we get that

$$\hat{\mu}_X \xrightarrow{P} \mu_X.$$

Now, let us show that $E(\bar{X} - \mu_X)^2 = \frac{\sigma_X^2}{n}$. Indeed, since $E(\bar{X}) = \mu_X$,

$$E(\bar{X} - \mu_X)^2 = var(\bar{X}) = \frac{\sum_{i=1}^n var(X_i)}{n^2} = \frac{\sigma_X^2}{n},$$

where the second equality uses the fact that random variables $\{X_i\}$ are independent.

Alternatively, the same result could be obtained using the direct computation of $E(\bar{X} - \mu_X)^2$:

$$\begin{aligned} E(\bar{X} - \mu_X)^2 &= E(\bar{X}^2 - 2\bar{X}\mu_X + \mu_X^2) = E(\bar{X}^2) - 2\mu_X^2 + \mu_X^2 = E(\bar{X}^2) - \mu_X^2 = \\ &= \frac{1}{n^2} E\left(\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j\right) - \mu_X^2. \end{aligned}$$

X_i and X_j are independent, hence, $E(X_i X_j) = E(X_i)E(X_j)$. Besides, they are identically distributed, therefore, $E(X_i) = E(X_j) = \mu_X$ and $E(X_i^2) = E(X_j^2)$. Using these facts, we get:

$$\begin{aligned} E(\bar{X} - \mu_X)^2 &= \frac{1}{n^2}\left[\sum_{i=1}^{n} E(X_i^2) + \sum_{i \neq j} E(X_i)E(X_j)\right] - \mu_X^2 = \\ &= \frac{1}{n^2}\left[nE(X_i^2) + (n^2 - n)(E(X_i))^2\right] - \mu_X^2 = \\ &= \frac{1}{n^2}[n\sigma_X^2 + n^2(E(X_i))^2] - \mu_X^2 = \frac{1}{n^2}[n\sigma_X^2 + n^2\mu_X^2] - \mu_X^2 = \frac{\sigma_X^2}{n}. \end{aligned}$$

So, to sum up, we obtain by the fact that linearity preserves Normality, that

$$\bar{X} \sim N(\mu_X, \tfrac{\sigma_X^2}{n}).$$

b) Finally,

$$\begin{aligned} E(\nu^2) &= \frac{1}{n-1} E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{1}{n-1} E\left(\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right) = \\ &= \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 + n\bar{X}^2 - 2\bar{X}\sum_{i=1}^{n} X_i\right) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n} E(X_i^2) - nE(\bar{X}^2)\right) = \frac{1}{n-1}\left(n(\sigma_X^2 + \mu_X^2) - n\left(\frac{\sigma_X^2}{n} + \mu_X^2\right)\right) = \\ &= \frac{1}{n-1}(n-1)\sigma_X^2 = \sigma_X^2. \end{aligned}$$

So, the sample variance ν^2 of X is an unbiased estimate of σ_X^2. It is also a consistent estimate of σ_X^2:

$$plim(\nu^2) = \sigma_X^2.$$

Indeed,

$$\begin{aligned} plim(\nu^2) &= plim\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = plim\left(\frac{n}{n-1}\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}\right) = \\ &= 1 \cdot plim\left(\frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}^2\right). \end{aligned}$$

By the Law of Large Numbers,

$$plim\left(\frac{\sum_{i=1}^{n} X_i^2}{n}\right) = E(X_i^2)$$

and

$$plim\left(\bar{X}\right) = E(X_i) = \mu_X.$$

Then

$$plim\left(\frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}^2\right) = E(X_i^2) - \mu_X^2 = \sigma_X^2.$$

Therefore,

$$plim(\nu^2) = \sigma_X^2.$$

Note, that ν^2 is *not a ML estimate* of σ_X^2. ML estimate of σ_X^2 is $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$. This estimate is *biased* but also *consistent*.

We move on to show the small-sample distribution of ν^2. The proof has several steps, we start from a theorem about diagonalisation, that will be given without proof.

•**THR** Let $A = A'$ be an $n\times n$ matrix, the eigenvalues and the corresponding eigenvectors are λ_i and c_i for $i = 1, \ldots, n$. c_is may be chosen so that $C = [c_1, \ldots, c_n]$ is such that $C'C = I_n$; moreover, for $\Lambda = diag\{\lambda_1, \ldots, \lambda_n\}$, A can be written as $A = C\Lambda C'$.

•**THR** Let $A = A'$. The rank of A is the number of nonzero eigenvalues of A.

It is an immediate consequence of the previous theorem. $A = C\Lambda C'$, so $rank(A) = rank(C\Lambda C') = rank(\Lambda)$ where the second equality follows from the nonsingularity of C. But the rank of a diagonal matrix is obviously the number of its nonzero elements. Therefore, the proof is finished.

•**THR** Assume that $A = A'$ and $A = AA$, then λ_i is 0 or 1.

This follows from the first theorem. $A = C\Lambda C'$, at the same time $A = AA = C\Lambda C'C\Lambda C' = C\Lambda^2 C'$; therefore $C(\Lambda - \Lambda^2)C' = 0$ which, by nonsingularity of C, implies that $\Lambda = \Lambda^2$ – this can happen iff $\lambda_i \in \{0, 1\}$.

•**THR** Assume that $A = A'$ and $A = AA$, then $rank(A) = trace(A)(= \sum \lambda_i)$.

Recall that $rank(A)$ is the maximum number of linearly independent rows/columns of A, and that $trace(A)$ is the sum of the diagonal elements of A. Again we use the diagonalisation. $rank(A) = rank(C\Lambda C') = rank(\Lambda) = \sum \lambda_i$, where the second equality holds because C is nonsingular, and the last equality follows from the previous theorem. Next, by the properties of *trace* operation, we get $trace(A) = trace(C\Lambda C') = trace(C'C\Lambda) = trace(\Lambda) = \sum \lambda_i$.

After this initial steps, let me define $X = [X_1, \ldots, X_n]'$ and $\mu = [\mu_X, \ldots, \mu_X]'$, $B = [n^{-1}, \ldots, n^{-1}]$ ($1\times n$ row matrix) and $H = [n^{-1}]$ ($n\times n$ matrix). Then $BX = \bar{X}$, $(I_n - H)X = [X_1 - \bar{X}, \ldots, X_n - \bar{X}]$ and $(I_n - H)\mu = 0$. Denote $A = (I_n - H)$, it is easy to see that $A = A'$

is symmetric and that $A = AA$ is idempotent; therefore we can use all the theorems above. $trace(A) = \sum(1-1/n) = (n-1)$, therefore A has $(n-1)$ eigenvalues equal to 1 (suppose $\lambda_1 = \ldots = \lambda_{n-1} = 1$) and one eigenvalue equal to 0 (let $\lambda_n = 0$). Moreover,

$$\frac{(n-1)\nu^2}{\sigma_X^2} = \frac{1}{\sigma_X^2}X'A'AX = \frac{1}{\sigma_X^2}(X-\mu)'A(X-\mu) =$$
$$\frac{1}{\sigma_X^2}(X-\mu)'C\Lambda C'(X-\mu) = [C'(X-\mu)/\sigma_X]'\Lambda[C'(X-\mu)/\sigma_X] = Z'\Lambda Z,$$

where $Z = C'(X-\mu)/\sigma_X$ (division by σ_X is element-wise) and it is easy to see that $((X-\mu)/\sigma_X) \sim N_n(0, I_n)$ and therefore $Z \sim N_n(0, C'C) = N_n(0, I_n)$. The proof of the fact that $\frac{(n-1)\nu^2}{\sigma_X^2} \sim \chi^2_{n-1}$ is finished by noting that $Z'\Lambda Z = \sum_1^{n-1} Z_i^2$ and, as Z_is are independent $N(0,1)$, thus has χ^2_{n-1} distribution.

c)

$$cov(\bar{X}, X_i - \bar{X}) = E(\bar{X}(X_i - \bar{X})) - E(\bar{X})E(X_i - \bar{X}) = E(\bar{X}(X_i - \bar{X})) - E(\bar{X})0 =$$
$$E(\bar{X}(X_i - \bar{X})) = 1/n\sum_j E(X_j(X_i - \bar{X})) = 1/n\sum_j(E(X_iX_j) - E(X_j\bar{X})) =$$
$$1/n((n-1)\mu_X^2+\sigma_X^2+\mu_X^2-\sum_j 1/n\sum_k E(X_jX_k)) = 1/n(n\mu_X^2+\sigma_X^2-1/n\sum_j((n-1)\mu_X^2+\sigma_X^2+\mu_X^2)) =$$
$$1/n(n\mu_X^2 + \sigma_X^2 - 1/n(n^2\mu^2 + n\sigma_X^2)) = 1/n(n\mu_X^2 + \sigma_X^2 - n\mu_X^2 - \sigma_X^2) = 0$$

If $\{X_i\} \overset{iid}{\sim} N(\mu_X, \sigma_X^2)$, then $\bar{X}$ and $X_i - \bar{X}$ are not only uncorrelated but also independent. Therefore their functions are independent as well, in particular $\bar{X}$ is independent of $\sum(X_i - \bar{X})^2$ and thus, independent of ν^2.

Therefore we have two independent random variables: $Z = \frac{\bar{X}-\mu^X}{\sqrt{\sigma_X^2/n}}$ and $V_{n-1} = \frac{(n-1)\nu^2}{\sigma_X^2}$; thus, the statistic

$$t = Z/\sqrt{V_{n-1}/(n-1)} = \frac{\sqrt{n}(\bar{X}-\mu^X)}{\sigma_X}/\sqrt{\frac{(n-1)\nu^2}{\sigma_X^2}/(n-1)} = \sqrt{n}(\bar{X}-\mu_X)/\nu$$

is, by definition, distributed as Student's t statistic with $(n-1)$ degrees of freedom.

d) We know that $t \sim t_{n-1}$ irrespective of the true value of μ_X or σ_X^2, therefore we can use this statistic to build a confidence interval for the mean μ_X. We are interested in a confidence interval with coverage probability of $1-\alpha$ (or $(1-\alpha)100\%$), which means we are looking for an interval $\hat{A}$ such that $1-\alpha = P_{t_{n-1}}(\hat{A} \ni t)$. t is unimodal symmetric, therefore it is intuitively clear that it is best to look for a symmetric interval centered around 0, $1-\alpha = P_{t_{n-1}}(-\hat{a} < t < \hat{a})$. But $1-\alpha = P_{t_{n-1}}(-\hat{a} < t < \hat{a}) = P_{t_{n-1}}(t < \hat{a}) - P_{t_{n-1}}(t < -\hat{a}) = F_{t_{n-1}}(\hat{a}) - (1 - F_{t_{n-1}}(\hat{a})) = 2F_{t_{n-1}}(\hat{a}) - 1$, which is equivalent to $\hat{a} = F^{-1}_{t_{n-1}}(1-\alpha/2)$ or that $\hat{a}$ is an $(1-\alpha/2)$ percentile of a t_{n-1} distribution.

To sum up, the $(1-\alpha)$ confidence interval for μ_X is given by

$$CI_{1-\alpha}(\mu_X) = (\bar{X} - \hat{a}\nu/\sqrt{n}, \bar{X} + \hat{a}\nu/\sqrt{n}), \quad \text{where} \quad \hat{a} = F_{t_{n-1}}^{-1}(1-\alpha/2).$$