This is a take home midterm. You can use your notes, my online notes on canvas and the textbooks book. You are supposed to work on your own text without external help. I'll be available to answer question in person or via email. Please, write clealy and legibly and take a readable scan before uploading.

Name (print): ____________________

Question:	1	2	3	4	Total
Points:	20	30	30	30	110
Score:					

Question 1 .... *20 point*

Confidence intervals for variance and covariance of two populations. We say that a r.v. X has a $\beta'(\alpha,\beta)$ distribution if

$$f_X(x) = \frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha,\beta)}\,.$$

We call $B'_{\alpha,\beta}(x)$ the c.d.f. of X.

(a) (10 points) Let U be a χ^2 r.v. with n d.o.f. and V a χ^2 r.v. with m d.o.f. and assume that U and V are independent. Show that $X = U/V$ has a $\beta'(n/2, m/2)$ distribution.(**Hint**: You can reason like in the derivation of the t distribution and use Theorem 5.7.3.)

Solution: Call $Y = V$ and $X = U/V$ so that $V = Y$ and $U = XY$. The joint p.d.f. of X and Y is then

$$f_{X,Y}(x,y) = \frac{1}{2^{n/2}\Gamma(n/2)}\frac{1}{2^{m/2}\Gamma(m/2)}yy^{m/2-1}(xy)^{n/2-1}e^{-y(1-x)/2}$$

so that

$$\begin{aligned}
f_X(x) =& \frac{1}{2^{n/2}\Gamma(n/2)}\frac{1}{2^{m/2}\Gamma(m/2)}x^{n/2-1}\int_0^\infty y^{(m+n)/2-1}e^{-y(1-x)/2}dy \\
=& \frac{1}{2^{n/2}\Gamma(n/2)}\frac{1}{2^{m/2}\Gamma(m/2)}x^{n/2-1}\left(\frac{1-x}{2}\right)^{-(n+m)/2}\Gamma((n+m)/2) \\
=& \frac{x^{n/2-1}(1+x)^{-(n+m)/2}}{B(n/2,m/2)}
\end{aligned}$$

(b) (10 points) It is not easy to find calculator for the β' distribution family. Show that if X had a $\beta'(\alpha,\beta)$ distribution then

$$Y = \frac{X}{1+X}$$

has a $\beta(\alpha,\beta)$ distribution.

Solution: We know that

$$f_X(x) = \frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha,\beta)}$$

Since $x = y/(1-y)$ and $dx/dy = 1/(1-y)^2$ while $(1+x) = 1/(1-y)$ we get

$$f_Y(y) = \frac{1}{B(\alpha,\beta)}\frac{1}{(1-y)^2}\left(\frac{y}{1-y}\right)^{\alpha-1}(1-y)^{\alpha+\beta} = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}$$

Question 2 .... *30 point*

Let X_i, $i = 1, \ldots, N$ be a random sample from a normal population with mean μ_X and variance σ_X^2 and Y_i, $i = 1, \ldots, M$ a random sample from a normal population with mean μ_Y and variance σ_Y^2. Assume that the two samples are independent.

(a) (10 points) Show that the r.v.

$$T = \frac{\sigma_Y^2 \sum_{i=1}^N (X_i - \overline{X})^2}{\sigma_X^2 \sum_{j=1}^M (Y_j - \overline{Y})^2}$$

is a pivotal quantity and find its p.d.f.. Here we used

$$\overline{X} = \frac{1}{N} \sum_{i=1}^N X_i \qquad \overline{Y} = \frac{1}{M} \sum_{j=1}^M Y_j \,.$$

Solution:

We know that $\sum_{i=1}^N (X_i - \overline{X})^2/\sigma_X^2$ has a χ^2 distribution with $N-1$ d.o.f. while $\sum_{j=1}^N (Y_j - \overline{Y})^2/\sigma_Y^2$ has a χ^2 distribution with $M-1$ d.o.f.. From point a) we get that T has a $\beta'((N-1)/2, (M-1)/2)$ distribution.

Observe that the distribution of T does not depend on μ_X, μ_Y or σ_X, σ_Y so that T is pivotal.

(b) (10 points) Write an exact coefficient γ confidence interval for the ratio σ_Y/σ_X of the form

$$A(\mathbf{X}, \mathbf{Y}) \le \frac{\sigma_Y}{\sigma_X} \le B(\mathbf{X}, \mathbf{Y})\,.$$

Express your result in term of the inverse of $B'_{\alpha,\beta}$.

Solution: Let α_1 and α_2 be such that $\alpha_1 + \alpha_2 = 1 - \gamma$ and let b_1 and b_2 be such that

$$B'_{(N-1)/2,(M-1)/2}(b_2) = 1 - \alpha_2 \qquad B'_{(N-1)/2,(M-1)/2}(b_1) = \alpha_1$$

then we can write

$$\mathbb{P}(b_1 \le X \le b_2) = \gamma$$

so that the C.I. is

$$\sqrt{b_1}\sqrt{\frac{\sum_{j=1}^{M}(Y_j - \overline{Y})^2}{\sum_{i=1}^{N}(X_i - \overline{X})^2}} \le \frac{\sigma_Y}{\sigma_X} \le \sqrt{b_2}\sqrt{\frac{\sum_{j=1}^{M}(Y_j - \overline{Y})^2}{\sum_{i=1}^{N}(X_i - \overline{X})^2}}$$

(c) (10 points) Assume now that $N = M$ and that $\gamma = 0.95$. Find the minimum N such that

$$B(\mathbf{X},\mathbf{Y})/A(\mathbf{X},\mathbf{Y}) < 1.1.$$

Here $A(\mathbf{X},\mathbf{Y})$ and $B(\mathbf{X},\mathbf{Y})$ are the statistics you found in point c). You can use any calculator to find the inverse c.d.f. of a $\beta(\alpha,\beta)$ random variable. Here is a possible one.

Solution: We need to find N such that

$$\frac{B'^{-1}_{(N-1)/2,(N-1)/2}((1+\gamma)/2)}{B'^{-1}_{(N-1)/2,(N-1)/2}((1-\gamma)/2)} \leq 1.1^2 = 1.21$$

or

$$\frac{b_N^+}{b_N^-}\,\frac{1-b_N^-}{1-b_N^+} \leq 1.21$$

where $b_N^+ = B^{-1}_{(N-1)/2,(N-1)/2}((1+\gamma)/2)$ and $b_N^- = B^{-1}_{(N-1)/2,(N-1)/2}((1+\gamma)/2)$. Using $\gamma = 0.95$ and Octave I got

$$N = 1695\,.$$

Question 3 . . . . *30 point*

Let the pair of random variable (X, Y) have a bivariate normal distribution with $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, $\mathrm{var}(X) = \mathrm{var}(Y) = 1$, and $\mathrm{cov}(X, Y) = \rho$.

(a) (10 points) Show that $U = (X + Y)/\sqrt{2}$ and $V = (X - Y)/\sqrt{2}$ are independent normal random variables and find their expected values and variances.

Solution: Clearly

$$\mathbb{E}(U) = \mathbb{E}(V) = 0$$

and U and V have a bivariate normal distribution. We have

$$\mathrm{cov}(U, V) = \mathbb{E}(UV) = \mathbb{E}(X^2) - \mathbb{E}(Y^2) = 0$$

so that U and V are independent.

Finally we have

$$\sigma_U^2 = \mathbb{E}(U^2) = \frac{1}{2}\mathbb{E}(X^2) + \frac{1}{2}\mathbb{E}(Y^2) + \mathrm{cov}(X, Y) = 1 + \rho$$
$$\sigma_V^2 = \mathbb{E}(V^2) = \frac{1}{2}\mathbb{E}(X^2) + \frac{1}{2}\mathbb{E}(Y^2) - \mathrm{cov}(X, Y) = 1 - \rho$$

(b) (10 points) Let now (X_i, Y_i) be a random sample from a population with a bivariate normal distribution like that described in point a). Call $U_i = (X_i + Y_i)/\sqrt{2}$ and $V = (X_i - Y_i)/\sqrt{2}$. Use question 1 to show that

$$S = \frac{1-\rho}{1+\rho} \frac{\sum_{j=1}^{N}(U_j - \overline{U})^2}{\sum_{i=1}^{N}(V_i - \overline{V})^2}$$

is a pivotal quantity and find its distribution. Here we set

$$\overline{U} = \frac{1}{N}\sum_i U_i \qquad \overline{V} = \frac{1}{N}\sum_{j=1}^{N} V_i\,.$$

Solution: As before, we have that $\sum_{j=1}^{N}(U_j - \overline{U})^2)/(1+\rho)$ and $\sum_{i=1}^{N}(V_i - \overline{V})^2)/(1-\rho)$ are independent χ^2 r.v. with $N-1$ degree of freedom so that S has a $\beta'((N-1)/2, (N-1)/2)$ distribution.

(c) (10 points) Use point b) to find a coefficient γ confidence lower bound for ρ of the form

$$\rho \geq A(\mathbf{X}, \mathbf{Y})\,.$$

Express your result in term of the appropriate inverse c.d.f..

Solution: Let b be such that $B'((N-1)/2, (N-1)/2) = \gamma$ so that

$$\mathbb{P}(S \leq b) = \gamma\,.$$

Calling

$$\widetilde{S}(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{i=1}^{N}(X_i + Y_i - \overline{X} - \overline{Y})^2}{\sum_{i=1}^{N}(X_i - Y_i - \overline{X} + \overline{Y})^2}$$

from point b) we get that with probability γ we have

$$\frac{1-\rho}{1+\rho} \leq b\,\widetilde{S}(\mathbf{X}, \mathbf{Y})$$

or

$$\rho \geq \frac{1 - b\,\widetilde{S}(\mathbf{X}, \mathbf{Y})}{1 + b\,\widetilde{S}(\mathbf{X}, \mathbf{Y})}\,.$$

Question 4 .... *30 point*

Let T_i, $i = 1, \ldots, N$ be a random sample from a population with exponential distribution with parameter λ, so that $\mathbb{E}(T_i) = 1/\lambda$.

(a) (10 points) Show that

$$Z = \lambda \sum_{i=1}^{N} T_i$$

is a pivotal quantity and find its distribution. Use it to find a coefficient γ confidence interval for λ of the form

$$A(\mathbf{T}) \leq \lambda \leq B(\mathbf{T})$$

such that $\mathbb{P}(\lambda < A(\mathbf{X})) = \mathbb{P}(\lambda > B(\mathbf{X}))$. Express your result in term of the inverse of the appropriate c.d.f.

Solution: We know that Z is a $\Gamma(N, 1)$ random variable. Calling $G_N(z)$ the c.d.f. of Z let

$$b_{1,N} = G_N^{-1}((1-\gamma)/2) \qquad b_{2,N} = G_N^{-1}((1+\gamma)/2)$$

so that

$$\mathbb{P}(Z < b_{1,N}) = \mathbb{P}(Z > b_{2,N}) = (1-\gamma)/2$$

Thus with probability γ we have

$$\frac{b_{1,N}/N}{\overline{T}} \leq \lambda \leq \frac{b_{2,N}/N}{\overline{T}}.$$

(b) (10 points) Use the Central Limit Theorem to derive a large sample approximate coefficient γ confidence interval for λ of the form

$$A'(\overline{T}) \leq \lambda \leq B'(\overline{T})$$

such that $\mathbb{P}(\lambda < A'(\overline{T})) = \mathbb{P}(\lambda > B'(\overline{T}))$. Observe that A' and B' depend only on

$$\overline{T} = \frac{1}{N} \sum_{i=1}^{N} T_i \,.$$

Express your result in therm of the inverse probability integral Φ^{-1}.

Solution: The Central Limit Theorem implies that $\overline{T}$ has an approximate normal distribution with mean $1/\lambda$ and variance $1/(N\lambda^2)$ so that

$$W = \sqrt{N}\lambda(\overline{T} - 1/\lambda)$$

is standard normal. Let

$$c = \Phi^{-1}((1+\gamma)/2)$$

so that with probability close to γ we have

$$\frac{1 - c/\sqrt{N}}{\overline{T}} \leq \lambda \leq \frac{1 + c/\sqrt{N}}{\overline{T}}.$$

(c) (10 points) Do the confidence intervals you found in point a) and b) coincide for N large? Justify your answer.

Solution: Since Z is the sum of N independent exponential r.v. with parameter 1, from the Law of Large Numbers we get that if $\delta > 0$

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{Z}{N} - 1 > \delta\right) = 0 \qquad \lim_{N\to\infty} \mathbb{P}\left(\frac{Z}{N} - 1 < -\delta\right) = 1\,.$$

Thus, fixed δ we can choose n so that for $N > n$ we have

$$G_N((1+\delta)N) > (1+\gamma)/2 \qquad\qquad G_N((1-\delta)N) < (1-\gamma)/2$$

or

$$1 - \delta \le \frac{b_{1,N}}{N} \le \frac{b_{2,N}}{N} \le 1 + \delta$$

so that

$$\lim_{N\to\infty} \frac{b_{1,N}}{N} = \lim_{N\to\infty} \frac{b_{2,N}}{N} = 1\,.$$

Clearly also

$$\lim_{N\to\infty} \left(1 - \frac{c}{\sqrt{N}}\right) = \lim_{N\to\infty} \left(1 + \frac{c}{\sqrt{N}}\right) = 1\,.$$

We can do better. The CLT tell us that

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{Z-N}{\sqrt{N}} \le c\right) = \Phi(c) = (1+\gamma)/2$$

so that

$$G_N(N + \sqrt{N}c) = (1+\gamma)/2 + \epsilon_N$$

with $\lim_{N\to\infty} \epsilon_N = 0$. Thus there exist $\eta_{2,N}$ with $\lim_{N\to\infty} \eta_{2,N} = 0$ so that

$$N + \sqrt{N}(c + \eta_{2,N}) = b_{2,N} \qquad \text{or} \qquad \frac{b_{2,N}}{N} = 1 + \frac{c + \eta_{2,N}}{\sqrt{N}}$$

and similarly we can find $\eta_{1,N}$ with $\lim_{N\to\infty} \eta_{1,N} = 0$ so that

$$N - \sqrt{N}(c + \eta_{1,N}) = b_{1,N} \qquad \text{or} \qquad \frac{b_{1,N}}{N} = 1 - \frac{c + \eta_{1,N}}{\sqrt{N}}$$

so that the two CIs really coincide.