STAT 517
Noncentral $\chi^2$ and $F$-distributions

Prof. Michael Levine

April 12, 2016
Definition

Let $X_1, \ldots, X_n \sim \mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \ldots, n$

Need to find the distribution of the quadratic form $Y = \frac{1}{\sigma^2} \sum_{i=1}^{n} X_i^2$

Direct integration suggests that the mgf of this distribution is

$$M(t) = \frac{1}{(1 - 2t)^{n/2}} \exp \left[ \frac{t \sum_{i=1}^{n} \mu_i^2}{\sigma^2 (1 - 2t)} \right]$$

for any $t < \frac{1}{2}$

In general, a random variable with an mgf $M(t) = \frac{1}{(1-2t)^{r/2}} e^{t\theta/(1-2t)}$ is said to have a noncentral chi-squared distribution with $r$ df and a noncentrality parameter $\theta$
If $\theta = 0$, the mgf is $M(t) = \frac{1}{(1-2t)^{r/2}}$ and we are back to $\chi^2_r$.

We will use the notation $\chi^2_r(\theta)$ for the noncentral chi-squared.

Conclude that the quadratic form $Y = \frac{1}{\sigma^2} \sum_{i=1}^n X_i^2 \sim \chi^2_n(\sum_{i=1}^n \mu_i^2 / \sigma^2)$.

If each $\mu_i = 0$, $Y \sim \chi^2_n$. 
Remarks about non-centrality parameter

- Note that for \( Y = \frac{1}{\sigma^2} \sum_{i=1}^{n} X_i^2 \) its non-centrality parameter can be computed by replacing each \( X_i \) in the quadratic form by its mean.

- This is not a random occurrence; let \( Q = Q(X_1, \ldots, X_n) \) be a quadratic form in normally distributed variables s.t. \( Q/\sigma^2 \sim \chi_r^2(\theta) \)

- Then, its noncentrality parameter is \( \theta = Q(\mu_1, \ldots, \mu_n)/\sigma^2 \)

- Moreover, if \( Q/\sigma^2 \) is a chi-square variable (central or non-central) for a set of values \( \mu_1, \ldots, \mu_n \), it is chi-square for all real values of these means

- Finally, Hogg-Craig’s theorem is valid for both central and non-central chi-square variables
Recall that for independent $U \sim \chi^2_{r_1}$, $V \sim \chi^2_{r_2}$, the ratio

$$F = \frac{r_2 U}{r_1 V} \sim F_{r_1, r_2}$$

Now, take $U \sim \chi^2_{r_1}(\theta)$, $V \sim \chi^2_{r_2}$, $U$ and $V$ are independent

The ratio

$$F = \frac{r_2 U}{r_1 V} \sim F_{r_1, r_2}(\theta)$$

is a noncentral F with noncentrality parameter $\theta$
Consider random variables $X_1, \ldots, X_b \sim N(\mu_j, \sigma^2)$ and independent

For known real constants $k_1, \ldots, k_b$ want to find a confidence interval for $\sum_{j=1}^{b} k_j \mu_j$

Take a random sample size $a$ $X_{1j}, X_{2j}, \ldots, X_{aj}$ from $N(\mu_j, \sigma^2)$

Clearly, $\bar{X}_j = \frac{\sum_{i=1}^{a} X_{ij}}{a} \sim N(\mu_j, \sigma^2/a)$; also $\sum_{i=1}^{a} (X_{ij} - \bar{X}_j)^2 / \sigma^2 \sim \chi^2(a-1)$ and the two are independent

Moreover, all of the $2b$ random variables $\bar{X}_j$, $\sum_{i=1}^{a} (X_{ij} - \bar{X}_j)^2 / \sigma^2$ are independent

Also, all $\bar{X}_1, \bar{X}_2, \bar{X}_b$ and $\sum_{j=1}^{b} \sum_{i=1}^{a} \frac{(X_{ij} - \bar{X}_j)^2}{\sigma^2}$ are independent; the latter is $\chi^2_b(a-1)$
Define $X - \sum_{j=1}^{b} k_j \bar{X}_j$; clearly, $Z$ is normal with mean $\sum_{j=1}^{b} k_j \mu_j$ and variance $\left( \sum_{j=1}^{b} k_j^2 \right) \sigma^2 / a$

Also, $Z$ is independent of

$$V = \frac{1}{b(a - 1)} \sum_{j=1}^{b} \sum_{i=1}^{a} (X_{ij} - \bar{X}_j)^2$$

Verify that

$$T = \frac{\sum_{j=1}^{b} k_j \bar{X}_j - \sum_{j=1}^{b} k_j \mu_j}{\sqrt{(V / a) \sum_{j=1}^{b} k_j^2}}$$

has $t_{b(a-1)}$ distribution
Introduction to multiple comparisons

- With probability $1 - \alpha$,

$$
\sum_{j=1}^{b} k_j \bar{X}_j - c \sqrt{\left( \sum_{j=1}^{b} k_j^2 \right) \frac{V}{a}}
$$

$$
\leq \sum_{j=1}^{b} k_j \mu_j \leq \sum_{j=1}^{b} k_j \bar{X}_j + c \sqrt{\left( \sum_{j=1}^{b} k_j^2 \right) \frac{V}{a}}
$$

- Do we need to derive this procedure for all possible linear combinations of means, e.g. $\mu_2 - \mu_1$, $\mu_3 - \frac{\mu_1 + \mu_2}{2}$ etc

- Or is it possible to have a simultaneous confidence interval for all of them?
Scheffe’s multiple comparison procedure

The following confidence interval works simultaneously for all linear combinations

$$\sum_{j=1}^{b} k_j \bar{X}_j - \sqrt{bd \left( \sum_{j=1}^{b} k_j^2 \right) \frac{V}{a}}$$

$$\leq \sum_{j=1}^{b} k_j \mu_j \leq \sum_{j=1}^{b} k_j \bar{X}_j + \sqrt{bd \left( \sum_{j=1}^{b} k_j^2 \right) \frac{V}{a}}$$

where $d = F_{\alpha, b, b(a-1)}$ is the $\alpha$ percentile of the respective F-distribution

It is typically very conservative