Discussion review 1

Math 181B

1 Review

1.1 Some distributions

- **Chi-squared distribution:** If $Z_1, \ldots, Z_n \overset{i.i.d.}{\sim} N(0,1)$, and $X = Z_1^2 + \ldots + Z_n^2$, then X has a Chi-squared distribution with n degrees of freedom. We write $X \sim \chi_n^2$.
- **t-distribution**: if $Z \sim N(0,1)$ and $X \sim \chi_n^2$. Assume X and Z to be independent. Then $T = \frac{Z}{\sqrt{X/n}}$ has a ***t***-distribution with n degrees of freedom. We write $T \sim t_n$.
- If $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2, then$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2$ has a chi squared distribution with $n-1$ degrees of freedom.
- If $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and standard deviation σ, then

$$T_{n-1} = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$$

  has a Student t distribution with $n-1$ degrees of freedom.

1.2 Some tests and intervals

- **Paired t-interval/test:**
  - Assumptions: a) Normality of the differences (or large sample size); b) Independence of differences.
  - Confidence Interval for μ_d:
$$\left[\bar{d} + t_{n-1,\frac{\alpha}{2}}\frac{s_D}{\sqrt{n}}, \bar{d} - t_{n-1,\frac{\alpha}{2}}\frac{s_D}{\sqrt{n}}\right]$$
  - Test statistics:
$$T = \frac{\bar{D} - \mu_0}{\frac{s_D}{\sqrt{n}}}$$
  - **p-value:**
$$P\left(t_{n-1} \le t\right) \text{ or } P\left(t_{n-1} \ge t\right) \text{ or } 2P\left(t_{n-1} \le -|t|\right)$$

    where t is the observation of test statistics T.

- **Two-sample z-interval/test:**
  - Assumptions: a) Normality of each sample (or large sample size); b) Independence within samples; c) Independence between samples; d) Known population variance σ_X^2 and σ_Y^2 of each sample, respectively.
  - Confidence Interval for $\mu_X - \mu_Y$:

$$\left[\bar{x} - \bar{y} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\ ,\ \bar{x} - \bar{y} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right]$$

  - Test statistics:

$$Z = \frac{\bar{X} - \bar{Y} - c}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

  - **p-value:**

$$P(Z \leq z) \text{ or } P(Z \geq z) \text{ or } 2P(Z \leq -|z|)$$

    where z is the observation of test statistics Z.

- **Two-sample t-interval/test unknown equal variance:**
  - Assumptions: a) Normality of each sample (or large sample size); b) Independence within samples; c) Independence between samples; d) Population variance σ_X^2 and σ_Y^2 of each sample are the same, but unknown.
  - Confidence Interval for $\mu_X - \mu_Y$:

$$\left[\bar{x} - \bar{y} + t_{n+m-2,\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n} + \frac{1}{m}}\ ,\ \bar{x} - \bar{y} - t_{n+m-2,\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n} + \frac{1}{m}}\right]$$

  - Test statistics:

$$T = \frac{\bar{X} - \bar{Y} - c}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

    where S_p^2 is the pooled variance:

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2 + \sum_{i=1}^m \left(Y_i - \bar{Y}\right)^2}{n+m-2}$$

    s_p^2 is the corresponding observation.
  - **p-value:**

$$P\left(t_{n+m-2} \leq t\right) \text{ or } P\left(t_{n+m-2} \geq t\right) \text{ or } 2P\left(t_{n+m-2} \leq -|t|\right)$$

    where t is the observation of test statistics T.

- **Two-sample t-interval/test unknown and unequal variance:** There is no exact solution to find a statistic with known density for testing the equality of two means from normally distributed random samples when the standard deviations of the samples are not equal. This is the Behrens-Fisher Problem. People usually apply the test statistics:

$$T_\nu = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{s_X^2}{n} + \frac{S_Y^2}{m}}}$$

where T_ν has approximately a t-distribution with v degrees of freedom:

$$\nu = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{s_X^4}{n^2(n-1)} + \frac{s_Y^4}{m^2(m-1)}}$$

The confidence of $\mu_X - \mu_Y$ and the p-value of HT can be found similarly to aforementioned tests.

1.3 Generalized likelihood ratio test

Definition 1.1 (Generalized Likelihood Ratio)**.** *Let $y_1, y_2, \ldots, y_n$ be a random sample from $f_Y(y; \theta_1, \ldots, \theta_k)$. The generalized likelihood ratio, λ, is defined to be*

$$\lambda = \frac{\max_\omega L(\theta_1, \ldots, \theta_k)}{\max_\Omega L(\theta_1, \ldots, \theta_k)} = \frac{L(\omega_e)}{L(\Omega_e)};$$

where ω is the set of unknown parameter values admissible under H_0; Ω, is the set of all possible values of all unknown parameters; we denote $\max_\omega L(\theta_1, \ldots, \theta_k)$ and $\max_\Omega L(\theta_1, \ldots, \theta_k)$ by $L(\omega_e)$ and $L(\Omega_e)$, respectively.

Then, we can define the procedure of generalized likelihood ratio test:

Definition 1.2 (Generalized Likelihood Ratio Test)**.** *A generalized likelihood ratio test (GLRT) is one that rejects H_0 whenever*

$$0 < \lambda \leq \lambda^*$$

where λ^* is chosen so that

$$P(0 < \Lambda \leq \lambda^* \mid H_0 \textit{ is true }) = \alpha;$$

where Λ denotes the generalized likelihood ratio expressed as a random variable.