Basic test statistics

Introduce $Y \sim N(0, \sigma^2)$.

1.1 Central distributions

1. Linear combination of $N(\mu_i, \sigma^2)$. Let a_i be known constants. Then, if we define a linear combination of independent normally distributed random variables:

$$\sum_{i=1}^{n} a_i N(\mu_i, \sigma_i^2) \sim N\left(\sum_i a_i \mu_i,\ \sum_i a_i^2 \sigma_i^2\right) \tag{1}$$

$\Rightarrow$ ***Linear combination of normally distributed random variables is also normally distributed.***

2. Standardized normal r.v. Let $Y \sim N(\mu, \sigma^2)$ then:

$$\frac{Y-\mu}{\sigma} \sim N(0,1) \tag{2}$$

3. χ_1^2 distribution with one degree of freedom. Let $Z \sim N(0,1)$ is std. normal r.v., then:

$$Z^2 = N^2(0,1) \sim \chi_1^2 \tag{3}$$

$\Rightarrow$ ***Square of std. normal random variable is χ_1^2 distributed r.v. with one degree of freedom.***

4. χ_n^2 distribution. Let ***iid*** $Z_i \sim N(0,1)$, then:

$$\sum_{i=1}^{n} Z_i^2 \sim \chi_n^2 = \chi^2(n) \tag{4}$$

If $Z_i \sim N(\mu_i, \sigma_i)$, then:

$$\sum_{i=1}^{n} \left(\frac{Z_i - \mu_i}{\sigma_i}\right)^2 \sim \chi_n^2 = \chi^2(n) \tag{5}$$

$\Rightarrow$ ***Sum of n squared std. random variables follows χ_n^2 distribution with n degrees of freedom.***

The primary reason that the chi-square distribution is used extensively in hypothesis testing is its relationship to the normal distribution.

Situation where χ_{n-1}^2 distribution arises from a normally distributed sample. If X_i are iid samples $\sim N(\mu, \sigma^2)$, then

$$SS_{xx} = \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \sim \sigma^2 \chi_{n-1}^2 \tag{6}$$

Sample variance S_{xx}^2 and χ_{n-1}^2:

$$S_{xx}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \sim \frac{\sigma^2}{(n-1)}\ \chi_{n-1}^2 \tag{7}$$

5. The Central Limit Theorem – distribution of the sample mean. Let Y_i be ***iid*** random variables: $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2 < \infty$, ***then its sample mean***

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \tag{8}$$

converges to a normal distribution as $n \to \infty$:

$$\boxed{\bar{Y}_{n\to\infty} \sim N(\mu,\ \sigma^2/n)} \tag{9}$$

Var of the sample mean
we know that $Var\left(\sum_{i=1}^{n} x_i\right) = n\sigma^2$ and that $Var(nx) = n^2\sigma^2$ is to combine this for the sample mean, we get:
$Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = Var\left(\sum_{i=1}^{n} \frac{1}{n}x_i\right) = nVar\left(\frac{1}{n}x_i\right) = n\frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}$

$\Rightarrow$ ***if*** n ***is sufficiently large, then*** $\bar{Y}$ ***approximately follows a normal distribution***. What constitutes sufficiency largely depends on the underlying distribution of the Y_is.

The theorem plays the key role in probability theory because it implies that if statistical methods work for normal distributions than they can be applicable to problems with other distributions.

6. t-distribution (Student's) with ν degrees of freedom. Let r.v. $N(0,1)$ and χ^2_ν are independent, then:

$$\frac{N(0,1)}{\sqrt{\chi^2_\nu/\nu}} \sim t_\nu \tag{10}$$

where $\nu = n - 1$.

How the t-distribution arises. Let $X_1, ..., X_n$ be iid samples $\sim N(\mu,\ \sigma^2)$. The sample mean is:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{11}$$

the Bessel corrected (we loose one degree of freedom in $\bar{X}$) sample variance is:

$$S^2_{xx} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \tag{12}$$

Then the random variable:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \tag{13}$$

has a standard normal distribution (CLT), and the random variable:

$$\boxed{\frac{\bar{X} - \mu}{S_{xx}/\sqrt{n}} \sim t_{n-1}} \tag{14}$$

where σ has been substituted for S, ***has a Student's t-distribution with*** $n-1$ ***degrees of freedom***.

7. F-distribution (Snedecor's) with ν and η degrees of freedom. Let r.v. χ^2_ν and χ^2_η, then:

$$\boxed{\frac{\chi^2_\nu/\nu}{\chi^2_\eta/\eta} \sim F_{\nu,\eta}} \tag{15}$$

$\Rightarrow$ ***ratio of two independent*** χ^2 ***random variables each divided by their respective degrees of freedom, follows the F-distribution***.

1.2 Non-central distributions

1. Non-central $\chi^2_{n,\lambda}$ distribution. If there are n iid $\sim N(\delta_i,\ 1)$, then:

$$\sum_{i=1}^{n} N^2(\delta_i, 1) \sim \chi^{2\prime}_{n,\lambda} \tag{16}$$

– non-central $\chi^{2\prime}_{n,\lambda}$ distribution ***with*** n ***degrees of freedom and non-centrality parameter*** λ:

$$\lambda = \sum_{i=1}^{n} \delta_i^2 \tag{17}$$

2. Non-central $t'_{v,\delta}$ distribution. Let $N(\delta,\ 1)$ and χ^2_ν are independent r.v., then:

$$\frac{N(\delta, 1)}{\sqrt{\chi^2_\nu/\nu}} \sim t'_{v,\delta} \tag{18}$$

– non-central $t'_{v,\delta}$ distribution ***with*** ν ***degrees of freedom*** and ***non-centrality parameter*** δ.

3. Non-central $F'_{\nu,\eta,\lambda}$ distribution. Let r.v. $\chi^{2\prime}_{n,\lambda}$ and χ^2_η are independent, then:

$$\frac{\chi^{2\prime}_{n,\lambda}/\nu}{\chi^2_\eta/\eta} \sim F'_{v,\eta,\lambda} \tag{19}$$

– non-central F-distribution ***with*** ν ***and*** η ***degrees of freedom and non-centrality parameter*** λ.