Normal distribution and t-distribution

Let $X \sim N(\mu, \sigma)$. Then

$$\frac{X-\mu}{\sigma} \sim N(0,1)$$

Let us now have a sample, $X_1, ..., X_n$, from the distribution $N(\mu, \sigma)$. (X_i's are independent and have $N(\mu, \sigma)$-distribution.) The mean μ can be estimated by the sample mean

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

and the variance σ^2 by the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

The sample mean $\bar{X}$ is $N(\mu, \sigma/\sqrt{n})$-distributed. Therefore,

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

If the variance is not know, it is estimated by its sample standard deviation S, and

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1},$$

where $n-1$ is the number of degrees of freedom, the parameter of the t-distribution.

T-test

Let us assume that $X \sim N(\mu_X, \sigma)$ and $Y \sim N(\mu_Y, \sigma)$, and we have a sample $X_1, ..., X_{n_X}$ from the distribution of X and a sample $Y_1, ..., Y_{n_Y}$ from the distribution of Y. In addition, let the two samples be independent.

The test statistic of the one sample T test is

$$\frac{\bar{X}-\mu_0}{S/\sqrt{n}},$$

where μ_0 is the expected value under the null hypothesis.

The test statistic of the two sample T-test, where we test whether there is difference in means between the two samples, is

$$\frac{\bar{X}-\bar{Y}}{S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}},$$

where $S_p^2 = \frac{(n_X-1)S_X^2+(n_Y-1)S_Y^2}{n_X+n_Y-2}$ and S_X^2 and S_Y^2 are the variance estimators for X and Y, respectively.

χ^2-distribution

Let $Z_1, ..., Z_n$ be a sample from $N(0,1)$-distribution. Then

$$\sum_{i=1}^{n} Z_i^2 = Z_1^2 + ... + Z_n^2 \sim \chi_n^2,$$

where n is called the number of degrees of freedom, the parameter of χ^2-distribution.

This implies that

$$\frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi_n^2$$

and

$$\frac{\sum\limits_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2.$$

The sample variance can be written as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{\sigma^2}{n-1}\frac{\sum\limits_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2},$$

where $\sum\limits_{i=1}^{n}(X_i-\bar{X})^2/\sigma^2 \sim \chi_{n-1}^2$. Therefore, $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$.

F-distribution

F-distribution is defined as a ratio of two χ^2-distributed random variables divided by their number of degrees of freedom, i.e.

$$\frac{\chi_n^2/n}{\chi_m^2/m} \sim F(n,m).$$

Let us now have two independent samples: $X_1, ..., X_n$ from $N(\mu_X, \sigma_X)$ and $Y_1, ..., Y_m$ from $N(\mu_Y, \sigma_Y)$. Furthermore, let $\bar{X}$ and $\bar{Y}$ be the sample means and S_X^2 and S_Y^2 the samples variances of the two samples, respectively. Then

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim \frac{\chi_{n-1}^2/(n-1)}{\chi_{m-1}^2/(m-1)} \sim F(n-1, m-1).$$

Connection between $N(0,1)$ and t-distribution

t-distribution is defined as a ratio of two independent random variables: a $N(0,1)$-distributed random variable Z and a square root of a χ^2-distributed random variable V divided by the number of its degrees of freedom n, i.e.

$$\frac{Z}{\sqrt{V/n}} \sim t_n.$$

Let $X \sim N(\mu, \sigma)$ and S^2 its estimated variance based on a sample of size n, and $V \sim \chi^2_n$ as above. Then,

$$\frac{X-\mu}{S} = \frac{X-\mu}{\sigma} \cdot \frac{\sigma}{S} = \frac{(X-\mu)/\sigma}{\sqrt{S^2/\sigma^2}} \sim t_{n-1}$$

since $(X-\mu)/\sigma \sim N(0,1)$ and $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$.