Definition 1 (Student's t-distribution) *Suppose that $X_0, X_1, \ldots, X_{n-1}$ are independent standard normal variables. Then the distribution of random variable $\frac{X_0}{\sqrt{\frac{X_1^2+\ldots+X_{n-1}^2}{n-1}}}$ is Student's t-distribution with $n-1$ degrees of freedom.*

Example 1 *In the case when the standard deviation σ of normal distribution is not known, we estimate it before constructing the confidence interval for m expected value. It is recognizable in the confidence interval*

$$\left(\overline{X} - t_{n-1,\frac{\alpha}{2}} \frac{\sqrt{\frac{(X_1-\overline{X})^2+\ldots+(X_n-\overline{X})^2}{n-1}}}{\sqrt{n}}; \overline{X} + t_{n-1,\frac{\alpha}{2}} \frac{\sqrt{\frac{(X_1-\overline{X})^2+\ldots+(X_n-\overline{X})^2}{n-1}}}{\sqrt{n}}\right),$$

where $t_{n-1,\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quantile of t-distribution with $n-1$ degrees of freedom.

Definition 2 (chi-squared distribution) *Suppose that $X_1, \ldots, X_n$ are i.i.d. standard normal variables. Then the distribution of random variable $Y = X_1^2 + \ldots + X_n^2$ is χ^2 with n degrees of freedom. Notation: $Y \sim \chi_n^2$.*

Example 2 (confidence interval for the variance of normal distribution) *It can be proved that the distribution of random variable $\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}$ is χ_{n-1}^2. This follows the confidence interval for σ^2:*

$$P\left(\sigma^2 \in \left(\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{h_{1-\frac{\alpha}{2},n-1}}; \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{h_{\frac{\alpha}{2},n-1}}\right)\right) = 1-\alpha,$$

where $h_{p,n-1}$ is the p quantile of χ_{n-1}^2 distribution.

Covariance, correlation

Definition 3 *The covariance of random variables X and Y is $cov(X,Y) := E((X-E(X))(Y-E(Y)))$.*

Remark 1

1. *in practice, $cov(X,Y) = E(XY) - E(X)E(Y)$, because by definition $cov(X,Y) = E[XY - XE(Y) - YE(X) + E(X)E(Y)] = E(XY) - E(X)E(Y)$*
2. *it is easy to show that if X and Y are independent random variables, then $cov(X,Y) = 0$*
3. *if $cov(X,Y) = 0$, then X and Y are not necessarily independent, i.e., the converse previous statement is not true (counterexample: $P(X=1) = P(X=-1) = P(X=0) = \frac{1}{3}$ and $Y = X^2$.)*
4. *covariance is a symmetric function: $cov(X,Y) = cov(Y,X)$*
5. *$cov(X,X) = Var(X)$*
6. *let a, b be real numbers; $cov(aX, bY) = ab \cdot cov(X,Y)$*

Proposition 1 *Suppose that $X_1, \ldots, X_n$ are (not necessarily independent) random variables. The variance of their sum can be written as*

$$Var(X_1 + \ldots + X_n) = \sum_{i=1}^{n} Var(X_i) + 2 \cdot \sum_{1 \le i < j \le n} cov(X_i, X_j).$$

In the special case, when the elements are pairwise independent, we get the form $Var(\sum X_i) = \sum Var(X_i)$.

Definition 4 (correlation coefficient) $R(X,Y) = \frac{cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$.

Properties:

1. if X and Y are independent, then $R(X,Y) = 0$ (conversely not true)
2. for all $a > 0$ and b real numbers, $R(X, aX+b) = 1$, because $cov(X, aX+b) = a \cdot Var(X)$

3. $|R(X,Y)| \leq 1$, furthermore, $|R| = 1$ if and only if $X = aY + b$ with probability 1 ($a \neq 0, b \in \mathbb{R}$) (Proof. use Cauchy-Schwarz-Bunakovsky inequality for $\langle X, Y \rangle := \mathrm{cov}(X,Y)$.)

The estimation of covariance based on an n-element sample is the empirical covariance: $\frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n}$. Remark that the sample means (x_i, y_i) pairs.

On the other hand, the estimation of correlation coefficient is the empirical correlation coefficient: $\frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2}}$.