MULTIVARIATE STATISTICS, Practice to Lesson 3.
Nonparametric tests based on limit distributions

Basic idea: find the limit distribution under the zero hypothesis of a conveniently constructed statistic, that does not depend on the parameters of the underlying distributon.

- *Pearson's χ^2-test*

  Test statistic:

$$\chi^2 = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i}$$

  where O_i's are the Observed and E_i's are the Expected frequencies. Under the zero hypothesis it asymptotically follows $\chi^2(df)$-distribution, where $df = r-1-b$ (b is the number ofestimated parameters). So the critical region corresponding to a test with size ε is the upper ε-point of the $\chi^2(df)$-distribution ($1-\varepsilon$ quantile value, see Table).

  1. χ^2-test for **goodness of fit**:

$$\chi^2 = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i}$$

  2. χ^2-test for **homogeneity**:

$$\chi^2 = nm \sum_{i=1}^{r} \frac{(\frac{\nu_i}{n} - \frac{\mu_i}{m})^2}{\nu_i + \mu_i}$$

  3. χ^2-test for **independence**:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\nu_{ij} - n\frac{\nu_{i\cdot}}{n}\frac{\nu_{\cdot j}}{n})^2}{n\frac{\nu_{i\cdot}}{n}\frac{\nu_{\cdot j}}{n}} = n \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(\nu_{ij} - \frac{\nu_{i\cdot}\nu_{\cdot j}}{n})^2}{\nu_{i\cdot}\nu_{\cdot j}}$$

  with $df = rs - 1 - [(r-1) + (s-1)] = (r-1)(s-1)$.

- *Kolmogorov–Smirnov-test*

  1. **One-sample K–S-test (for goodness of fit)**:

$$D_n = \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)|.$$

  Under the zero hypothesis (the ordered sample is from the continuous F- distribution)

$$\lim_{n\to\infty} \mathbb{P}(\sqrt{n} D_n < z) = K(z), \quad \forall z \in \mathbb{R},$$

  that is the test statistic $\sqrt{n}D_n$ asymptotically follows Kolmogorov-distribution.

  2. **Two-sample K–S-test (for homogeneity)**:

$$D_{n,m} = \sup_{x \in \mathbb{R}} |F_n^*(x) - G_m^*(x)|.$$

  Under the zero hypothesis (the two ordered samples are from the same continuous distribution), the test statistic $\sqrt{\frac{nm}{n+m}} D_{n,m}$ asymptotically follows Kolmogorov-distribution. So the critical region corresponding to a test with size ε is the upper ε-point of the Kolmogorov-distribution ($1-\varepsilon$ quantile value, see Table). Observe that the above suprema are, in fact, maxima of finitely many terms, and the distributions can be transformed into the $\mathcal{U}(0,1)$.