LUKÁCS' THEOREM AND CONSEQUENCES

Definition: Let $\xi_1, \dots, \xi_n \sim \mathcal{N}(0,1)$ be i.i.d. rv's. Then the distribution of the rv $\xi = \sum_{i=1}^n \xi_i^2$ is called χ^2 (chi2) distribution with degrees of freedom (d.f.) n.
Definition: Let $\eta \sim \mathcal{N}(0,1)$ and $\xi \sim \chi^2(n)$ be independent rv's. Then the distribution of

$$t = \frac{\eta}{\sqrt{\xi/n}}$$

is called Student t-distribution with degrees of freedom (d.f.) n and denoted by $t(n)$ (Student=V. Gosset).

Lukács' Theorem. Let $X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma)$ be i.i.d. rv's. Then

1. $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$;
2. $nS_n^2/\sigma^2 \sim \chi^2(n-1)$, or equivalently, $(n-1){S_n^*}^2/\sigma^2 \sim \chi^2(n-1)$;
3. $\bar{X}$ and S_n^2 are independent rv's, or equivalently, $\bar{X}$ and ${S_n^*}^2$ are independent rv's.

Consequences:

- Recall that in case of $X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma_0)$ i.i.d. sample, where σ_0 is known, for any $0 < \alpha < 1$, the $1-\alpha$ level confidence interval for μ is

$$I_{1-\alpha} = \bar{X} \pm \frac{z_{\alpha/2}\sigma_0}{\sqrt{n}}, \tag{1}$$

  where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile value of the standard normal distribution.

- In case of $X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma)$ i.i.d. sample, where σ is unknown, by Lukacs' Theorem,

$$t = \frac{\frac{\bar{X}-\mu}{\sigma}\sqrt{n}}{\sqrt{\frac{(n-1){S_n^*}^2}{\sigma^2}/(n-1)}} = \frac{\bar{X}-\mu}{S_n^*}\sqrt{n} \sim t(n-1),$$

  therefore, for any $0 < \alpha < 1$, the $1-\alpha$ level confidence interval for μ is

$$I_{1-\alpha} = \bar{X} \pm \frac{t_{\alpha/2}(n-1)S_n^*}{\sqrt{n}}, \tag{2}$$

  where $t_{\alpha/2}(n-1)$ is the $1-\alpha/2$ quantile value of the $t(n-1)$ distribution.

- Going further, in view of the expectation and variance of the $\chi^2(n-1)$ distribution,

$$\mathbb{E}\left((n-1){S_n^*}^2/\sigma^2\right) = n-1,$$

  so

$$\mathbb{E}\left({S_n^*}^2\right) = \sigma^2.$$

This is another proof that the corrected empirical variance is an unbiased estimator of the true (population) variance of the normal distribution. Also,

$$\mathrm{Var}\left((n-1){S_n^*}^2/\sigma^2\right) = 2(n-1),$$

so

$$\mathrm{Var}({S_n^*}^2) = \frac{2(n-1)}{(n-1)^2}\sigma^4 = \frac{2\sigma^4}{(n-1)} \to 0$$

as $n \to \infty$. Consequently, ${S_n^*}^2$ is an unbiased estimator with "small" variance in the normal case.

- Therefore, for "large" n ($n \geq 30$), even in case of unknown variance the confidence interval of (1) can be updated to

$$I_{1-\alpha} = \bar{X} \pm \frac{z_{\alpha/2} S_n^*}{\sqrt{n}},$$

whereas (2) is mainly applicable for "small" ($n < 30$) sample sizes.