Poznámky k předmětu Aplikovaná statistika, 6. téma

6. Asymptotické chování součtu velkého počtu náhodných veličin

To, proč v praktické statistice prochází některé základní postupy, se kterými jste se již setkali, je důsledkem tzv. **centrální limitní věty** (CLV), která, volně řečeno, říká, že vhodně znormovaný součet velkého počtu náhodných veličin má přibližně normované normální rozdělení.

Věta 1 (Centrální limitní věta)**.** *Buď* $X_1, X_2, \dots$ *posloupnost nezávislých stejně rozdělených náhodných veličin se střední hodnotou* $\mu \in \mathbb{R}$ *a rozptylem* $\sigma^2 \in (0, +\infty)$*. Označme*

$$Y_n := \frac{1}{\sqrt{n\sigma^2}} \left(\sum_{i=1}^{n} X_i - n\mu \right), \; n \in \mathbb{N},$$

a F_n *distribuční funkci náhodné veličiny* Y_n*. Pak pro každé* $x \in \mathbb{R}$ *platí*

$$\lim_{n \to +\infty} F_n(x) = \Phi(x), \tag{1}$$

kde Φ *je distribuční funkce náhodné veličiny s* $\mathcal{N}(0,1)$*.*

Poznámka 1. *Vztah (1) se obvykle čte následovně:*

(i) Y_n *konvergují v distribuci k náhodné veličině* U *s rozdělením* $\mathcal{N}(0,1)$ *pro* n *jdoucí do nekonečna, či*

(ii) náhodné veličiny Y_n *mají asymptoticky (pro* $n \to +\infty$*) rozdělení* $\mathcal{N}(0,1)$*, což se zapisuje ve tvaru*

$$Y_n \overset{\text{as.}}{\sim} \mathcal{N}(0,1), \; n \to +\infty.$$

Z Věty 1 plyne tvrzení o asymptotickém chovaní součtu a aritmetickém průměru náhodných veličin.

Poznámka 2. *Pro posloupnost náhodných veličin* $X_1, X_2, \dots$ *splňující předpoklady Věty 1 platí*

$$\sum_{i=1}^{n} X_i \overset{\text{as.}}{\sim} \mathcal{N}(n\mu, n\sigma^2), \; n \to +\infty,$$

$$\bar{X} := \frac{1}{n} \sum_{i=1}^{n} X_i \overset{\text{as.}}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \; n \to +\infty.$$

Jak je vidět, je možné s využitím centrální limitní věty přibližně modelovat celkovou chybu jako součet velkého počtu drobných náhodných chyb.

Příklad 1. *V jednom kole hry lze na automatu prohrát* 2 *Kč (s pravděpodobností* $0,29$*) nebo* 1 *Kč (s pravděpodobností* $0,40$*), anebo vyhrát* 3 *Kč (s pravděpodobností* $0,31$*). Pomocí CLV určete přibližnou pravděpodobnost, že po absolvování* 5 000 *kol této hry budeme mít zisk.*

Příklad 2. *Buďte* $X_i, i = 1, \dots, 10\,000$*, nezávislé stejně rozdělené náhodné veličiny popisující počty volání na telefonní ústřednu během i-té minuty. Víme, že průměrný počet volání na ústřednu za jednu minutu je* 4*. S využitím CLV zjistěte přibližnou pravděpodobnost, že během příštích* 10 000 *minut bude průměrný počet volání za minutu menší než* $3,99$*.*