Skórová funkce rozdělení a její použití pro studium podobnosti náhodných veličin

Zdeněk Fabián
ÚI AVČR Praha

Buď $\chi \subseteq R$ otevřený interval. Rozdělení F s nosičem χ je popsáno distribuční funkcí $F(x)$ a pravděpodobnostní hustotou $f(x)$. Zatím nepříliš známou možností popisu F je skórová funkce rozdělení $S_F(x)$. Pro dané $x_0 \in \chi$ představuje $f(x_0)$ pravděpodobnost výskytu pozorované hodnoty v (malém intervalu kolem) x_0, hodnota $S_F(x_0)$ udává relativní vliv x_0 na těžiště (typickou hodnotu) x^* rozdělení, kteréžto je řešením rovnice $S_F(x) = 0$. ES_F^2 lze chápat jako Fisherovu informaci vzhledem k x^*. Efektivní poloměr rozdělení je pak popsán veličinou ω, kde $\omega^2 = 1/ES_F^2$ je skórový rozptyl.

V první části příspěvku funkci S_F definuji a ukážu, že za předpokladu, že data $\mathbb{X}_n = (X_1, ..., X_n)$ jsou náhodným výběrem z F, které je členem parametrické rodiny $\{F(x;\theta), \theta \in \Theta\}$, charakterizují výběr kromě odhadnuté vlivové funkce rozdělení $S_F(\mathbb{X}_n; \hat{\theta}_n)$ i výběrové těžiště $\hat{x}^* = x^*(\hat{\theta}_n)$ a výběrový skórový rozptyl $\hat{\omega}^2 = \omega^2(\hat{\theta}_n)$, kde $\hat{\theta}_n$ je odhadem θ.

Skórová funkce rozdělení $\mathcal{N}(\mu,\sigma)$ je $S_F(x;\mu,\sigma) = \frac{x-\mu}{\sigma^2}$, $x^* = \mu$ a $\omega^2 = \sigma^2$. Skórová funkce $S_F(x;\tau) = \frac{1}{\tau}(\frac{1}{\tau} - 1)$ exponenciálního rozdělení s hustotou $f(x;\tau) = \frac{1}{\tau}e^{x/\tau}$ je shodná s (Fisherovou) skórovou funkcí pro τ ($x^* = \tau, \omega^2 = \tau^2$). Obecně je však S_F novou, skalární funkcí charakterizující F, charakteristiky x^* a ω^2 novými charakteristikami rozdělení a jejich výběrové hodnoty novým popisem výběrů z F. Tak např. pro Paretovo rozdělení s hustotou $f(x,\alpha) = \frac{\alpha}{x^{\alpha+1}}$, jehož střední hodnota a rozptyl pro $\alpha < 1$ neexistují, je $S_F(x;\alpha) = \alpha(1 - x^*/x)$, $x^* = 1 + 1/\alpha$ je těžiště a $\omega^2 = (\alpha+2)/\alpha^3$ skórový rozptyl. Protože obecně $ES_F = 0$, výběrové těžiště Paretova rozdělení lze určit z rovnice $\sum_{i=1}^n (1 - x^*/x_i) = 0$, odkud plyne, že typická hodnota výběru z Pareta je harmonický průměr.

Ve druhé části uvedu výsledky simulací použití odhadnuté skórové funkce rozdělení pro zjišťování podobnosti dvou náhodných veličin. Skórový korelační koeficient náhodných veličin X a Y definuji jako

$$r_F = corr(S_X, S_Y)$$

kde $corr(X,Y)$ je Pearsonův korelační koeficient a $S_X(\mathbb{X};\hat{\theta})$ a $S_Y(\mathbb{X};\hat{\theta})$ odhadnuté skórové funkce rozdělení F_X a F_Y. Výsledky porovnám s výsledky běžně používaných metod odhadu korelačního koeficientu.