## Osnove matematične statistike

Iz *populacije* $S$ vzamemo *vzorec* $n$ objektov. Na vzorcu izvedemo meritve naključne spremenljivke $X$, ki jih združimo v naključni vektor $\mathbf{V} = (X_1, X_2, ..., X_n)$. *Statistika* je vsaka merljiva funkcija vzorca $Z(\mathbf{V})$. V splošnem je statistika naključna spremenljivka.

**Vzorčno povprečje** za vzorec z $n$ elementi je:

$$\langle X \rangle_n = \frac{1}{n} \sum_{i=1}^{n} X_i \,. \tag{1}$$

Vzorčno povprečje je nepristranska in dosledna cenilka populacijskega povprečja $m$:

$$\mathrm{E}\left[\langle X \rangle_n\right] = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left[X_i\right] = \frac{1}{n} nm = m \,, \qquad P\left(|\langle X \rangle_n - m| \geq \varepsilon\right) \leq \frac{\sigma_x^2}{n\varepsilon^2} \underset{n\to\infty}{\longrightarrow} 0 \,, \text{ za } \varepsilon > 0. \tag{2}$$

Če je spremenljivka $X$ normalno porazdeljena z gostoto $\mathcal{N}\left(X; m, \sigma\right)$, potem je vzorčno povprečje tudi normalno porazdeljeno, in sicer z gostoto $\mathcal{N}\left(\langle X \rangle_n; m, \sigma/\sqrt{n}\right)$. Če pa $X$ ni normalno porazdeljena, potem se porazdelitev vzorčnega povprečja z naraščajočo velikostjo vzorca približuje normalni:

$$\lim_{n\to\infty} f_{\langle X \rangle_n}(\langle x \rangle_n) = \mathcal{N}\left(\langle X \rangle_n; m, \sigma/\sqrt{n}\right) \,. \tag{3}$$

Običajno predpostavimo, da je za $n \geq 30$ porazdelitev $\langle X \rangle_n$ približno normalna, ne glede na porazdelitev spremenljivke $X$.

**Vzorčna varianca** za vzorec z $n$ elementi je:

$$s^2 = \left\langle (X - \langle X \rangle_n)^2 \right\rangle_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \langle X \rangle_n)^2 \,. \tag{4}$$

Vzorčna varianca je pristranska cenilka populacijske variance $\sigma^2$:

$$\mathrm{E}\left[s^2\right] = \left(1 - \frac{1}{n}\right)\sigma^2 = \sigma^2 + \mathcal{O}(1/n) \,. \tag{5}$$

**Popravljena vzorčna varianca** za vzorec z $n$ elementi je:

$$S^2 = \frac{ns^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \langle X \rangle_n)^2 \,. \tag{6}$$

Popravljena vzorčna varianca je nepristranska in dosledna cenilka populacijske variance $\sigma^2$:

$$\mathrm{E}\left[S^2\right] = \sigma^2 \qquad \mathrm{Var}\left[S^2\right] \underset{n\to\infty}{\longrightarrow} 0 \,. \tag{7}$$

## Osnove teorije cenilk - točkovno ocenjevanje

Parametre porazdelitev (npr. $m$, $\sigma$, $p$, $\lambda$ itn.) lahko ocenjujemo točkovno ali intervalno. Pri točkovni oceni sta na voljo dve metodi: *metoda momentov* in *metoda največje zanesljivosti.*

Pri **metodi momentov** izraze za teoretične momente verjetnostne porazdelitve, katere parametre želimo oceniti, izenačimo z izrazi za ustrezne vzorčne momente. Potrebujemo toliko parov momentov, kolikor parametrov želimo oceniti. V dobljenem sistemu enačb so ocenjevani parametri neznanke. Z rešitvijo sistema enačb dobimo cenilke za parametre. Cenilko parametra $\theta$ označimo z $\hat{\theta}$.

Pri **metodi največje zanesljivosti** tvorimo *funkcijo zanesljivosti*:

$$L(\mathbf{v};\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \cdot \cdots \cdot f(x_n;\theta)\,, \tag{8}$$

kjer je $f(x;\theta)$ gostota verjetnosti naključne spremenljivke $X$ s parametrom $\theta$. Funkcija zanesljivosti je torej enaka verjetnosti, da pri vzorčenju dobimo vzorec v prostornini $d\mathbf{v}$ okoli $\mathbf{v}$. Parameter $\theta$ verjetnostne porazdelitve določimo tako, da ima vrednost funkcije $L(\mathbf{v};\theta)$ na konkretnem vzorcu meritev $\mathbf{v}$ maksimum, to je, da smo izmerili vzorec $\mathbf{v}$ zato, ker je najbolj verjeten. V ta namen $L(\mathbf{v};\theta)$ odvajamo po $\theta$ in odvod izenačimo z nič. Iz dobljene enačbe dobimo izraz za parameter $\theta$, ki predstavlja izraz za cenilko $\hat{\theta}$. Če je ocenjevanih parametrov več, potem funkcijo zanesljivosti odvajamo po vsakem parametru posebej, s čimer dobimo sistem enačb za parametre. Rešitve sistema so cenilke za parametre.