Kapitola 7

Případ německých tanků

7.1 Odhad počtu prvků

Nechť $X_1, \ldots, X_n$ je výběr bez vracení z *diskrétního rovnoměrného rozdělení* na množině $\{1, 2, \ldots, N\}$, kde $n \leq N$. Označme $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ uspořádaný náhodný výběr. Ukážeme, že odhad parametru N lze založit na veličině $X_{(n)}$. Budeme postupovat podle článků Johnson (1994 a, b).

Lemma 7.1 *Nechť* $p_j = \mathsf{P}(X_{(n)} = j)$. *Pak platí*

$$p_j = \frac{\binom{j-1}{n-1}}{\binom{N}{n}}, \qquad j = n, n+1, \ldots, N. \tag{7.1}$$

Důkaz. Jev $\{X_{(n)} = j\}$ implikuje, že hodnoty $X_{(1)}, X_{(2)}, \cdots, X_{(n-1)}$ musí být vybrány z množiny $\{1, 2, \ldots, j-1\}$. Takových možností je $\binom{j-1}{n-1}$. Všech možností, jak vybrat hodnoty $X_1, \ldots, X_n$, je $\binom{N}{n}$. Odtud plyne vzorec (7.1). □

Lemma 7.2 *Platí*

$$\sum_{j=n}^{N} \binom{j}{n} = \binom{N+1}{n+1}, \qquad \sum_{j=n}^{N} \binom{j+1}{n+1} = \binom{N+2}{n+2}. \tag{7.2}$$

Důkaz. Vzorce (7.2) vyplývají z toho, že součet pravděpodobností (7.1) se rovná jedné. Lze je také dokázat přímo matematickou indukcí. □

Věta 7.3 *Platí*

$$\mathsf{E}X_{(n)} = \frac{n(N+1)}{n+1}, \tag{7.3}$$

$$\mathsf{E}X_{(n)}^2 = \frac{n}{n+2}(N+2)(N+1) - \frac{n(N+1)}{n+1}, \tag{7.4}$$

$$\mathsf{var}X_{(n)} = \frac{n(N+1)(N-n)}{(n+1)^2(n+2)}. \tag{7.5}$$

Důkaz. Nejprve dokážeme vzorec (7.3). Máme

$$\mathsf{E}X_{(n)} = \sum_{j=n}^{N} jp_j = \sum_{j=n}^{N} j\frac{\binom{j-1}{n-1}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}}\sum_{j=n}^{N} n\binom{j}{n}.$$

Z (7.2) pak plyne, že

$$\mathsf{E}X_{(n)} = \frac{n}{\binom{N}{n}}\binom{N+1}{n+1} = \frac{n(N+1)}{n+1}.$$

Nyní budeme dokazovat (7.4). Máme

$$\mathsf{E}X_{(n)}^2 = \sum_{j=n}^{N} j^2 p_j = \sum_{j=n}^{N}(j+1)jp_j - \sum_{j=n}^{N} jp_j. \tag{7.6}$$

Víme, že $\sum_{j=n}^{N} jp_j = \mathsf{E}X_{(n)} = n(N+1)/(n+1)$. Dále dostaneme

$$\begin{aligned}\sum_{j=n}^{N}(j+1)jp_j &= \frac{1}{\binom{N}{n}}\sum_{j=n}^{N}\frac{(j+1)j(j-1)!}{(n-1)!(j-n)!}\\ &= \frac{n(n+1)}{\binom{N}{n}}\sum_{j=n}^{N}\frac{(j+1)!}{(n+1)!(j-n)!}\\ &= \frac{n(n+1)}{\binom{N}{n}}\binom{N+2}{n+2}\\ &= \frac{n}{n+2}(N+2)(N+1).\end{aligned}$$

Dosazením do (7.6) dostaneme vzorec (7.3). Vzorec (7.4) plyne z toho, že $\mathsf{var}\, X_{(n)} = \mathsf{E}X_{(n)}^2 - [\mathsf{E}X_{(n)}]^2$. □

Věta 7.4 *Nechť* $\hat{N} = \frac{n+1}{n}X_{(n)} - 1$. *Pak platí*

$$\mathsf{E}\hat{N} = N, \qquad \mathsf{var}\,\hat{N} = \frac{(N+1)(N-n)}{n(n+2)}. \tag{7.7}$$

Důkaz. Tvrzení věty plyne z toho, že

$$\mathsf{E}\hat{N} = \frac{n+1}{n}\mathsf{E}X_{(n)} - 1, \qquad \mathsf{var}\,\hat{N} = \frac{(n+1)^2}{n^2}\mathsf{var}\,X_{(n)}. \qquad \square$$

Dá se dokázat, že $\hat{N}$ je stejnoměrně nejlepší nestranný odhad (uniformly minimum variance unbiased estimate — UMVUE) parametru N. Jiný nestranný odhad je $Z = 2\bar{X} - 1$. Pro něj platí

$$\mathsf{var}\,Z = \frac{(N+1)(N-n)}{3n} > \frac{(N+1)(N-n)}{n(n+2)} = \mathsf{var}\,\hat{N}.$$

Další nestranné odhady uvádí Johnson (1994b).

7.2 Problém německých tanků

Teorie vyložená v odst. 7.1 se používá k tomu, aby se odhadl počet prvků nějaké množiny. Popíšeme to podle části článku Grajalez (2013), která je nazvaná The German tank problem. Představme si, že chceme vědět, kolik je taxíků v určitém městě. Každý taxík má své číslo. Předpokládejme, že jsme viděli taxíky s čísly 17, 250, 337 a 591. Pokud tato čísla můžeme považovat za náhodný výběr z rovnoměrného rozdělení na množině $\{1, 2, \ldots, N\}$, odhad $\hat{N}$ podle vzorce uvedeného ve větě 7.4 bude

$$\hat{N} = \frac{5}{4} \times 591 - 1 = 737.71.$$

V praxi bychom toto číslo zaokrouhlili, a tak bychom odhadli, že taxíků je ve městě 738.

Odhad počtu taxíků se může jevit jenom jako hříčka, ale jsou situace, kdy na takovém odhadu mohou záviset životy mnoha lidí. V roce 1944 spojenci chystali vylodění v Normandii. Zajímalo je, na kolik německých tanků tam mohou narazit. Hlavně se obávali nového německého tanku, kterému se říkalo Panther. Ten měl důkladné pancéřování a velké dělo, takže předčil americké Shermany. Spojenci se s Panthery střetli v Itálii. Jelikož se tyto tanky začaly vyrábět teprve nedávno, byla naděje, že jich ve Francii bude jen několik málo.

Každý tank má několik čísel. Má číslo podvozku, číslo motoru a číslo hlavně. Potíž je, že tato čísla nejsou na tanku napsána tak velkými ciframi jako to mají taxíky v Londýně či v New Yorku. Spojenci měli k dispozici jen dva Panthery. Jeden byl ukořistěn na Sicílii a druhý v Rusku. Jak z výběru o rozsahu 2 odhadnout celkový počet vyrobených tanků, tak to byl úkol, který dostali američtí statistici.

Čísla podvozků nepomohla. Bylo známo, že podvozky vyrábí pět továren a čísla netvoří souvislou řadu. Zato skříně převodovek měla čísla 1, 2, ... Jenže výběr o rozsahu 2 je příliš malý na rozumný odhad počtu N vyrobených tanků. Naštěstí tanky mají podvozková kola, která podpírají housenkové pásy. Tato kola mají gumové pneumatiky a ty se lisují ve formách. Každá forma má své číslo a to se také vylisuje na pneumatiku. Forma však nevydrží nekonečně dlouho. Statistici si zjistili u britských výrobců, po kolika výliscích je nutné použít novou formu s vyšším číslem. Panther měl 8 náprav a na každé 6 podvozkových kol. U dvou tanků tedy bylo 96 pneumatik. Tímto způsobem se dospělo k odhadu, že Němci za jeden měsíc vyrábějí 270 Pantherů. To bylo mnohem víc, než zatím spojenci předpokládali. Plány invaze musely být změněny. Podrobněji o této problematice pojednává článek Ruggles, Brodie (1947).

Po válce se zjistilo, že v únoru 1944 Němci vyrobili 276 Pantherů. Odhad statistiků byl tedy velmi přesný.

K této problematice se pojí následující úsměvná historka. Na středním východě dostal plukovník Trevor Dupuy od izraelského velení souhlas k prohlídce celé výrobní linky Merkava Tank. Plukovník se zeptal, kolik tanků se tu vyrábí. Bylo mu řečeno, že jde o tajnou informaci. Přitom bylo na každém podvozku napsáno sériové výrobní číslo.