Příklady - Bodový odhad

15. října 2013

Pražské metro

Přijdu v pražském metru na nástupiště a tam zjistím, že metro v mém směru jelo před 2:30 a metro v opačném směru před 4:20. Udělejte bodový odhad, jak dlouho budu čekat na metro.

1) Metoda momentů

Označím si:

$X = 2 : 30 = 150s$,

$Y = 4 : 20 = 260s$.

Předpokládejme, že metra jezdí v pravidelných intervalech po B sekundách. Pokud přijdu v náhodný okamžik, má doba, jakou budu čekat, rovnoměrné spojité rozdělení od 0 do B.

Odhaduji jeden parametr, bude mi tedy stačit první moment, tedy střední hodnota. Zjistím tedy střední hodnotu, jak dlouho bych čekal na metro a střední hodnotu z rovnoměrného spojitého rozdělení. Obojí porovnám.

Podle zadání budu v jednom směru čekat $x_1 = B - X$ a v druhém směru $x_2 = B - Y$. Střední hodnota těchto časů je: $= \frac{B-X+B-Y}{2} = B - \frac{X+Y}{2} = B - \frac{150+260}{2} = B - 205$.

Střední hodnota pro rovnoměrné rozdělení od 0 do B je: $\mu = \frac{0+B}{2} = \frac{B}{2}$.

Porovnáme:

$$\begin{aligned} \mu &= \overline{x} \\ \frac{B}{2} &= B - 205 \\ B &= 2B - 410 \\ B &= 410 \end{aligned}$$

Máme tedy bodový odhad pro dobu, po jaké jezdí metra. Bodový odhad pro dobu, jakou budu čekat v prvním směru je tedy: $T = B - X = 410 - 150 = 260$.

Můžeme si všimnout, že nám vyšla stejná hodnota jako je Y. Není to náhoda, protože když příklad necháme v písmenkách X a Y, vyjde nám

$$\begin{aligned} \frac{B}{2} &= B - \frac{X+Y}{2} \\ B &= X + Y \end{aligned}$$

$$T = B - X = Y$$

Získáváme tedy poněkud paradoxní výsledek, že nejlepší bodový odhad doby, jakou budu čekat na metro, získám tak, že se podívám na hodiny v opačném směru.

Je tento podivný výsledek správně?

Ano, je. Musíme si ale uvědomit, že děláme bodový odhad, který nám nic neříká o přesnosti odhadu. A protože celou metodu nasazujeme jen na dva údaje, je odhad velmi, velmi nepřesný.

2) Metoda maximální věrohodnosti

Vyřešíme stejný příklad metodou maximální věrohodnosti. Pro rovnoměrné rozdělení se to dělá spíše úvahou než výpočtem, což si ukážeme.

Metra jezdí v pravidelných intervalech délky B. Pro dobu, před jakou odjelo metro, mám opět rovnoměrné rozdělení od 0 do B.

Znám dvě hodnoty pro doby, před jakou odjelo metro:

$X = 2:30 = 150s$,

$Y = 4:20 = 260s$.

Určitě chci, aby tyto hodnoty ležely v intervalu $\langle 0, B\rangle$. Pak bude hustota pravděpodobnosti pro obě hodnoty rovna:

$f(150) = \frac{1}{B}$,

$f(260) = \frac{1}{B}$.

(Hustota pravděpodobnosti má tvar obdélníku. Jeho délka je B. Integrál přes def. obor musí být 1, takže výška obdélníku je $\frac{1}{B}$.)

Chci mít co největší věrohodnostní funkci, což je: $V(B) = f(150) \cdot f(260) = \frac{1}{B} \cdot \frac{1}{B}$. Proto musí být B co nejmenší. A přesto se obě hodnoty musí do intervalu vejít. To mejlépe udělám tak, že B bude rovno největší hodnotě z dat. Tedy $B = \max(X_i) = 260$.

Odhad pro dobu čekání je tedy $260 - 150 = 110$.

Shrnutí a porovnání metod

Obě metody nám tedy dávají různé odhady. A mají také různé statistiky pro odhad parametru B:

$B = X + Y$,

$B = \max(X_i)$.

Která z těchto statistik je lepší?

Soustřeďme se na druhou statistiku. Pokud bychom znali skutečnou hodnotu B a pak ji odhadovali největší hodnotou, která padla, pak bude odhad vždy menší než skutečná hodnota. Takže statistika nebude nestranná!!! (Bude tzv. „vychýlená".) Takže to není dobrá statistika a dáme přednost první metodě momentů.

Udělejme si takové porovnání obou metod:

Rovnoměrné rozdělení: Lepší je metoda momentů.

Exponenciální rozdělení: Obě metody dávají stejné výsledky. Metoda momentů bývá jednodušší.

Normální = Gaussovo rozdělení: Metoda momentů je jednodušší a dává lepší výsledky.

Složité několikarozměrné rozdělení: Lepší je metoda maximální věrohodnosti.

Normální = Gaussovo rozdělení

Naměřili jsme tato data: 2, 8, 8, 14. Předpokládáme, že tato data mají normální rozdělení. Jaké konkrétní Gaussově křivce nejlépe odpovídají?

1) Metoda maximální věrohodnosti

Tato metoda je poměrně pracná a pro normální rozdělení nedává příliš dobrý odhad pro rozptyl. Přesto jsem ji tu ukázal, abyste se s ní seznámili.

Gaussova křivka je dána parametry μ a σ. Tyto parametry tedy odhadneme.

Hustota pravděpodobnosti normálního rozdělení je dána vzorcem:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Věrohodnostní funkce je tedy dána vzorcem:

$$\begin{aligned}
V(\mu,\sigma) &= \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(2-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(8-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(8-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(14-\mu)^2}{2\sigma^2}} = \\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^4 \cdot e^{-\frac{(2-\mu)^2+(8-\mu)^2+(8-\mu)^2+(14-\mu)^2}{2\sigma^2}} = \\
&= \frac{1}{4\pi^2} \cdot \sigma^{-4} \cdot e^{-\frac{1}{2}\cdot\sigma^{-2}\cdot\left[(2-\mu)^2+(8-\mu)^2+(8-\mu)^2+(14-\mu)^2\right]}.
\end{aligned}$$

Hledáme maximum funkce, tedy body, v nichž jsou obě parciální derivace podle μ i podle σ rovny 0.

Spočtěme první parciální derivaci:

$$\begin{aligned}
\frac{\partial V}{\partial \mu} &= \frac{1}{4\pi^2} \cdot \sigma^{-4} \cdot e^{-\frac{1}{2}\cdot\sigma^{-2}\cdot\left[(2-\mu)^2+(8-\mu)^2+(8-\mu)^2+(14-\mu)^2\right]} \cdot \\
&\quad \cdot \left(-\frac{1}{2}\right) \cdot \sigma^{-2} \cdot \left[-2(2-\mu) - 2(8-\mu) - 2(8-\mu) - 2(14-\mu)\right] = \\
&= \frac{1}{4\pi^2} \cdot \sigma^{-6} \cdot e^{-\frac{1}{2}\cdot\sigma^{-2}\cdot\left[(2-\mu)^2+(8-\mu)^2+(8-\mu)^2+(14-\mu)^2\right]} \cdot \left[(2-\mu)+(8-\mu)+(8-\mu)+(14-\mu)\right].
\end{aligned}$$

Kdy se tento výraz rovná nule? Tehdy, když je nulová hranatá závorka napravo. Máme tedy rovnici:

$$\begin{aligned}
\left[(2-\mu)+(8-\mu)+(8-\mu)+(14-\mu)\right] &= 0 \\
32-4\mu &= 0 \\
\mu &= 8.
\end{aligned}$$

Spočtěme ještě druhou parciální derivaci:

$$\frac{\partial V}{\partial \sigma} = \frac{1}{4\pi^2} \cdot \Big[-4\sigma^{-5} \cdot e^{-\frac{1}{2}\cdot\sigma^{-2}\cdot\left[(2-\mu)^2+(8-\mu)^2+(8-\mu)^2+(14-\mu)^2\right]} +$$

$$
\begin{aligned}
& +\sigma^{-4} \cdot e^{-\frac{1}{2}\cdot\sigma^{-2}\cdot\left[(2-\mu)^2+(8-\mu)^2+(8-\mu)^2+(14-\mu)^2\right]} \cdot \\
& \cdot \left(-\frac{1}{2}\right) \cdot \left(-2\sigma^{-3}\right) \cdot \left[\left[(2-\mu)^2 + (8-\mu)^2 + (8-\mu)^2 + (14-\mu)^2\right]\right] = \\
= \; & \frac{1}{4\pi^2} \cdot e^{-\frac{1}{2}\cdot\sigma^{-2}\cdot\left[(2-\mu)^2+(8-\mu)^2+(8-\mu)^2+(14-\mu)^2\right]} \cdot \sigma^{-7} \cdot \\
& \cdot \left[-4\sigma^2 + (2-\mu)^2 + (8-\mu)^2 + (8-\mu)^2 + (14-\mu)^2\right].
\end{aligned}
$$

Kdy se tento výraz rovná nule? Když je nulová hranatá závorka na posledním řádku. Máme tedy:

$$
\begin{aligned}
-4\sigma^2 + (2-\mu)^2 + (8-\mu)^2 + (8-\mu)^2 + (14-\mu)^2 &= 0 \\
-4\sigma^2 + (2-8)^2 + (8-8)^2 + (8-8)^2 + (14-8)^2 &= 0 \\
-4\sigma^2 + 36 + 0 + 0 + 36 &= 0 \\
\sigma^2 &= 18.
\end{aligned}
$$

Hledané rozdělení je tedy: $N(8, 18)$.

Můžeme si všimnout, že získaný rozptyl je rozptyl, jako kdyby data byla základním souborem. To je chybička, která s rostoucím počtem dat mizí. Odhad rozptylu je vychýlený, proto dáme přednost metodě momentů.

2) Metoda momentů

Střední hodnota dat je: $\overline{x} = 8$.

Výběrový rozptyl dat je: $s^2 = 24$.

Střední hodnotu a rozptyl základního souboru odhadneme střední hodnotou a rozptylem výběrového souboru:

$$
\begin{aligned}
\mu &= \overline{x} = 8, \\
\sigma^2 &= s^2 = 24.
\end{aligned}
$$

Hledané rozdělení je: $N(8, 24)$.

Metoda je podstatně jednodušší a dává pro rozptyl správný výsledek.

Příklady ke spočtení

1) Několik atomových jader zkoumaného prvku se rozpadlo po 8, 4, 12, 2, 19, 3 minutách. Odhadněte funkci, která udává pravděpodobnost, že se jádro v určitém čase rozpadne.

Exponenciální rozdělení, $f(t) = \frac{1}{8} \cdot e^{-\frac{t}{8}}$

2) Hmotnost myší má normální rozdělení. Získali jsme následující hmotnosti myší: 6, 12, 7, 10, 18, 9, 12, 10, 15. Odhadněte rozdělení hmotnosti myší.

Metoda momentů, $f(m) \sim N(11;\, 14,25)$

3) Má tramvaj jezdí ráno v pravidelných intervalech. Zlý vandal ale strhl jízdní řád. Na tramvaj jsem minulý týden čekal 2, 14, 8, 11 a 5 minut. Odhadněte, v jakém intervalu jezdí tramvaj.

Metoda momentů, rovnoměrné spojité rozdělení, 16 minut.

4) Jsem opravář na železniční trati. Ráno jsem zjistil, že musím provést opravu na 111., 123. a 89. kilometru. Odhadněte, jaký úsek trati mám na starosti.

Metoda momentů, rovnoměrné spojité rozdělení, 75. až 139. kilometr.