## Binomické rozdělení

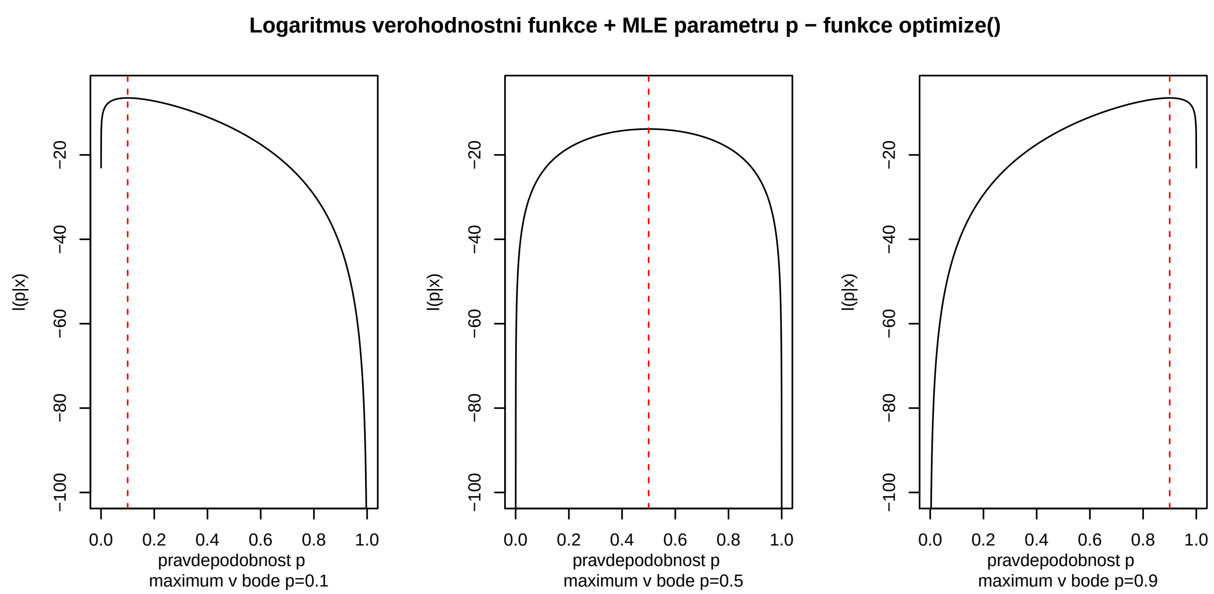
**Příklad 1** (**binomické rozdělení; maximálně věrohodný odhad** $p$**).** Nechť $X \sim Bin(N, p)$ a realizace $X$ jsou $x = n$. Předpokládejme, že jsme pozorovali (a) $x = 2$, (b) $x = 10$ a (c) $x = 18$ úspěchů v $N = 20$ pokusech.

a) Pomocí R vypočítejte maximálně věrohodný odhad $p$. Výsledek zobrazte do grafu spolu s logaritmickou funkcí věrohodnosti.

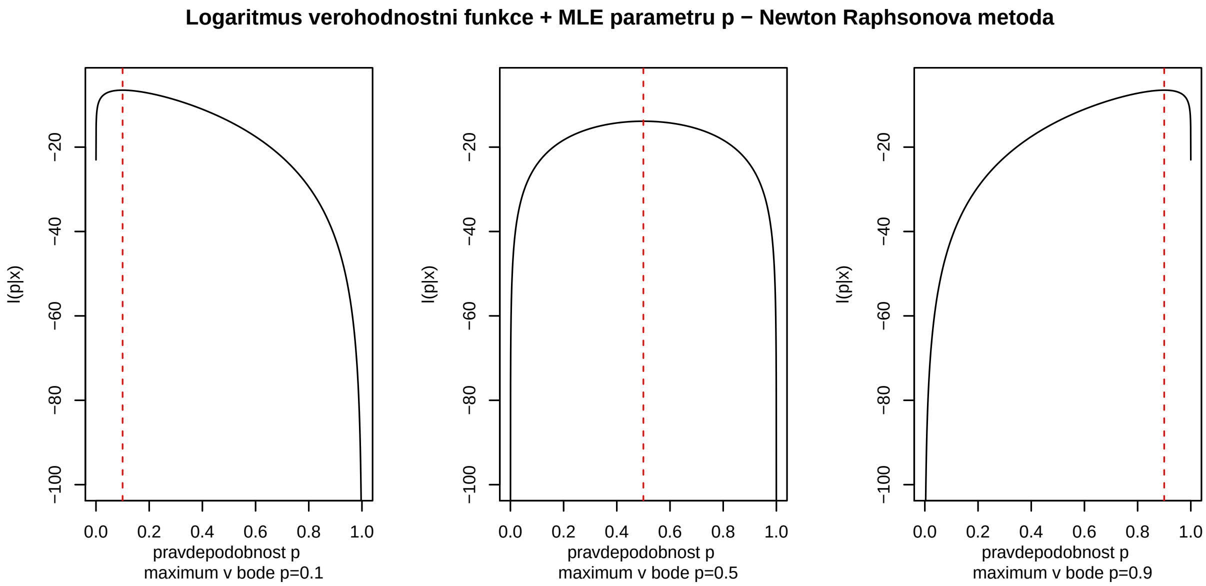
b) Naprogramujte Newton-Raphsonovu iterační metodu. Touto metodou nahraďte funkci `optimize()`, nalezněte maximálně věrohodný odhad parametru $p$, výsledek zaneste do grafu spolu s logaritmickou funkcí věrohodnosti (grafy budou stejné, jako grafy vygenerované v části (a)).

c) Pomocí R vypočítejte maximálně věrohodný odhad $p$. Výsledek zobrazte do grafu spolu s funkcí věrohodnosti.

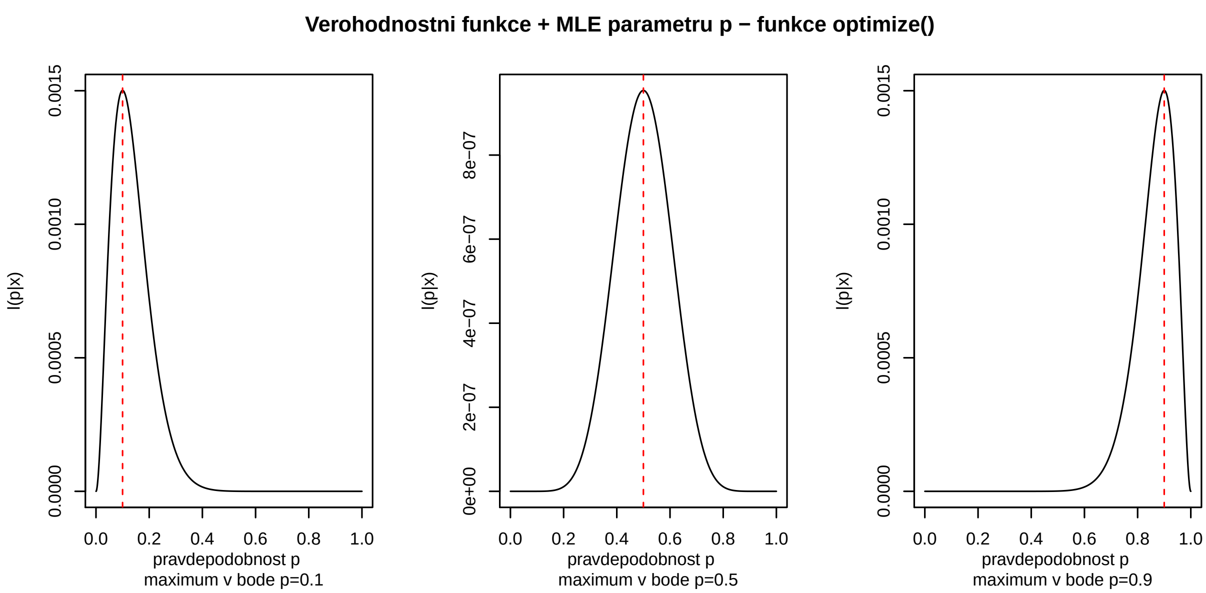
ad a)
```
## [1] 0.1000006
## [1] 0.5
## [1] 0.8999994
```




ad b)
```
## [1] 0.1
## [1] 0.5
## [1] 0.9
```




ad c)
```
## [1] 0.09998216
## [1] 0.5
## [1] 0.9000178
```




**Příklad 2** **($\mathcal{I}(\widehat{p})$ a rozptyl pro $p$; $X \sim Bin(N,p)$).** Z funkce věrohodnosti odvoďte pozorovanou Fisherovu míru informace $\mathcal{I}(\widehat{p})$ a rozptyl $\widehat{Var}[\widehat{p}]$.

**Příklad 3 (maximálně věrohodné odhady; binomické rozdělení).** Za předpokladu, že náhodná proměnná $X$ má binomické rozdělení, vypočítejte maximálně věrohodný odhad $\widehat{p}$ pomocí logaritmu funkce věrohodnosti $l(p|\mathbf{x})$. Porovnejte tento odhad s výrazem $\sum_{i=1}^{N} x_i/N$. Realizacemi náhodné proměnné $X$ jsou následující binární proměnné:

(a) pohlaví (sex; data: one-sample-probability-sexratio.txt, kde označení pohlaví 'dívka' ('f') přeznačíme na 1 a označení pohlaví 'chlapec' ('m') přeznačíme na 0;

(b) pohlaví (sex; data: two-samples-probabilities-sexratio.txt), kde označení pohlaví 'muž' ('m') přeznačíme na 1 a označení pohlaví 'žena' ('f') přeznačíme na 0.

V případě (a) počítáme pravděpodobnost výskytu děvčat a v případě (b) pravděpodobnost výskytu chlapců.

```
## [1] "a) Odhad parametru p= 0.4804"
## [1] "b) Odhad parametru p= 0.5274"
```

**Příklad 4.**

1. Nakreslete škálovaný logaritmus funkce věrohodnosti binomického rozdělení. Na $x$-ové ose bude $p$ a na $y$-ové ose $\ln \mathcal{L}(p) = l(p|\mathbf{x}) - \max(l(p|\mathbf{x}))$. Porovnejte $\ln \mathcal{L}(p)$ s kvadratickou aproximací vypočítanou pomocí Taylorova rozvoje $\ln \mathcal{L}(p) = \ln\left(\frac{L(p|\mathbf{x})}{L(\widehat{p}|\mathbf{x})}\right) \approx -\frac{1}{2}\mathcal{I}(\widehat{p})(p-\widehat{p})^2$.

2. Nechť skóre funkce $S(p) = \frac{\partial}{\partial p} \ln L(p|\mathbf{x})$. Vezmeme-li derivaci kvadratické aproximace uvedené výše, dostaneme $S(p) = -\mathcal{I}(\widehat{p})(p-\widehat{p})$ anebo $-\mathcal{I}^{-1/2}(\widehat{p})S(p) \approx \mathcal{I}^{1/2}(\widehat{p})(p-\widehat{p})$. Potom zobrazením pravé strany na $x$-ové ose a levé strany na $y$-ové ose dostaneme asymptoticky lineární funkci s jednotkovým sklonem. Asymptoticky také platí $\mathcal{I}^{1/2}(\widehat{p})(p-\widehat{p}) \sim N(0,1)$, Je postačující mít rozsah $x$-ové osy $\langle -2; 2\rangle$, protože funkce je asymptoticky (lokálně) lineární na tomto intervalu. Rozumně škálujte $y$-vou osu. Zobrazte pro (a) $n = 8$, $N = 10$, (b) $n = 80$, $N = 100$ a (c) $n = 800$, $N = 1000$ ($p \in (0.5; 0.99)$). Okomentujte rozdíly mezi (a), (b) a (c).