

1 Binomické rozdělení $\text{Bin}(N, p)$

- Bernoulliho pokusy X_1, \dots, X_N
 - $X_i = 1 \dots$ událost nastala; $X_i = 0 \dots$ událost nenastala; $i = 1, \dots, N$
 - $\Pr(X_i = 1) = p$
 - $\Pr(X_i = 0) = 1 - p = q$
- $X \dots$ počet událostí v posloupnosti N nezávislých Bernoulliho pokusů, přičemž pravděpodobnost nastání události v každém pokusu je vyjádřena parametrem p
- $\sum_{i=1}^N X_i = X \sim \text{Bin}(N, p)$
- $\theta = (N, p)$
- pravděpodobnostní funkce

$$p(x) = \binom{N}{x} p^x (1-p)^{N-x} \quad x = 0, 1, \dots, N$$

- vlastnosti: $E[X] = Np$; $\text{Var}[X] = Np(1-p)$
- $\text{dbinom}(x, N, p)$, $\text{pbinom}(x, N, p)$, $\text{rbinom}(M, N, p)$
- Data:
 - **Dataset 1: Počet chlapců v rodinách s 12 dětmi**
 - V rámci studie poměru pohlaví u lidí z roku 1889 bylo na základě záznamů z nemocnic v Sasku zaznamenáno rozdělení počtu chlapců v čtrnáctičlenných rodinách. Mezi $M = 6115$ rodinami s $N = 12$ dětmi byla pozorována početnost chlapců. Údaje ze studie jsou uvedeny v následující tabulce.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	Σ
m_{observed}	3	24	104	286	670	1033	1343	1112	829	478	181	45	7	6115

Příklad 1.1. Výpočet očekávaných početností za předpokladu binomického modelu

Vezměte údaje z **datasetu 1**. Vypočítejte očekávané početnosti výskytu chlapců m_{expected} za předpokladu, že početnosti chlapců X v rodinách mají binomické rozdělení $\text{Bin}(N, p)$ s parametry $N = 12$ a

$$\hat{p} = \frac{\sum_{n=0}^N n m_{\text{observed}}}{NM}. \quad (1.1)$$

Řešení příkladu 1.1

Odhad parametru p , tj. $\hat{p} = 0.5192$. Tabulka očekávaných početností m_{expected} je

n	0	1	2	3	4	5	6	7	8	9	10	11	12	Σ
m_{expected}	1	12	72	258	628	1085	1367	1266	854	410	133	26	2	6114

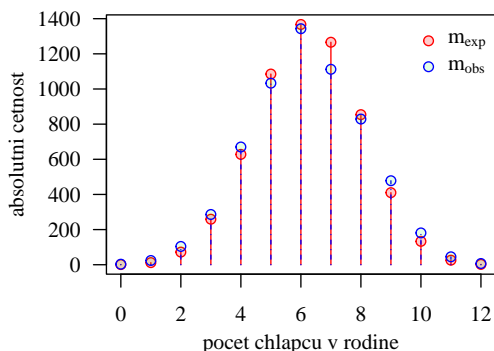
Z tabulky očekávaných početností vidíme, že součet očekávaných početností dává výsledek $M = 6114$. Odchylka od původní hodnoty $M = 6115$ je způsobena přesným výpočtem očekávaných početností a jejich následným zaokrouhlením. Konkrétně v tomto případě, pokud bychom očekávané početnosti spočítali na základě binomického rozdělení $\text{Bin}(N, p)$, kde $N = 12$ a $\hat{p} = 0.5192$ (hodnotu \hat{p} bychom pro další výpočet zaokrouhlili na čtyři desetinná místa) dostali bychom následující tabulku očekávaných početností.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	Σ
m_{expected}	1	12	72	259	628	1085	1367	1266	854	410	133	26	2	6115

Při použití hodnoty odhadu \hat{p} zaokrouhlené na čtyři desetinná místa došlo k mírné úpravě binomického rozdělení a očekávaná početnost výskytu tří chlapců v rodině se z hodnoty 258.47 přehoupla na hodnotu 258.52, čímž jsme po zaokrouhlení získali absolutní početnost 259 namísto 258. Ostatní očekávané početnosti zůstaly nezměněny, jelikož jejich nezaokrouhlená hodnota byla dostatečně blízko k celému číslu. ★

Příklad 1.2. Overdispersion a underdispersion v binomickém modelu

V předchozím příkladu 1.1 jsme stanovili očekávané početnosti výskytu chlapců v rodinách s dvanácti dětmi. Do jednoho grafu zanešte nyní hodnoty pozorovaných početností $m_{observed}$ a hodnoty očekávaných početností $m_{expected}$. Pozorované a očekávané početnosti od sebe barevně odlište. Na základě výsledného grafu stanovte, zda došlo v tomto případě k overdisperzi nebo underdisperzi. Závěr podložte srovnáním rozptylu vypočítaného z pozorovaných dat s rozptylem vypočítaným z očekávaných dat.

Řešení příkladu 1.2

Obrázek 1: Porovnání pozorovaných a očekávaných početností v binomickém modelu

Oproti očekávaným početnostem jsou pozorované početnosti krajních případů vyšší a naopak pozorované početnosti nejčastějších případů jsou nižší. Tato situace ukazuje na vyšší rozptyl pozorovaných početností než teoretických. Jde tedy odisperzi.

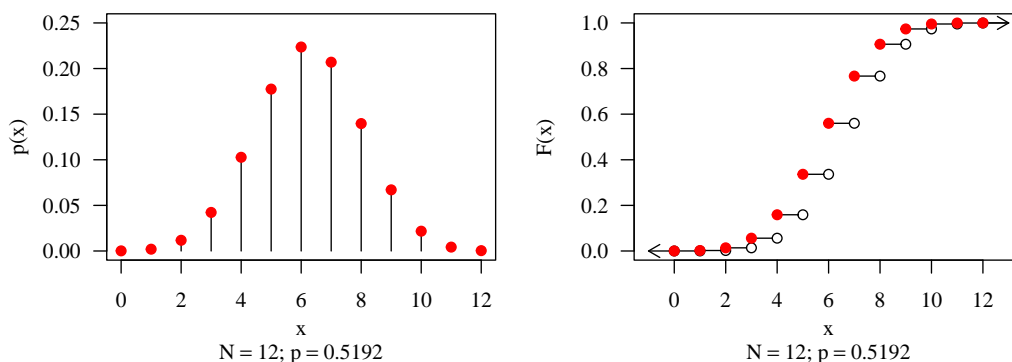
	var. obs	var. exp
1	3.48984	2.995539

1
2

Rozptyl skutečných počtů chlapců v rodinách s dvanácti dětmi je 3.4898, rozptyl očekávaných počtů chlapců je 2.9956. Rozptyl skutečných počtů je tedy vyšší než rozptyl očekávaných počtů chlapců v rodinách s dvanácti dětmi. ★

Příklad 1.3. Graf pravděpodobnostní a distribuční funkce binomického modelu

V příkladu 1.1 jsme odhadli hodnotu parametru p binomického rozdělení $\text{Bin}(N, p)$ jako $\hat{p} = 0.5192$. Hodnota parametru $N = 12$. Nakreslete graf pravděpodobnostní a distribuční funkce binomického rozdělení $\text{Bin}(N, p)$, kde $N = 12$ a $p = 0.5192$.

Řešení příkladu 1.3

Obrázek 2: Pravděpodobnostní a distribuční funkce binomického modelu

★

Příklad 1.4. Výpočet pravděpodobností na základě binomického modelu

Za předpokladu, že náhodná veličina X , udávající počet chlapců v rodině s dvanácti dětmi, pochází z binomického rozdělení, tj. $X \sim \text{Bin}(N, p)$, s parametry $N = 12$ a $p = 0.5192$, vypočítejte pravděpodobnost, že v rodině s dvanácti dětmi bude (a) právě devět chlapců; (b) nejvýše čtyři chlapci; (c) alespoň osm chlapců; (d) čtyři, pět, šest, nebo sedm chlapců.

Řešení příkladu 1.4

	a	b	c	d
1	0.06703911	0.1588736	0.2330869	0.7107605

3
4

Pravděpodobnost, že v rodině bude právě devět chlapců, je 6.7%. Pravděpodobnost, že v rodině budou nejvýše čtyři chlapci, je 15.89%. Pravděpodobnost, že v rodině bude alespoň osm chlapců, je 23.31%. Pravděpodobnost, že v rodině bude čtyři, pět, šest, nebo sedm chlapců, je 71.08%. ★

Příklad 1.5. Střední hodnota a rozptyl náhodné veličiny z binomického modelu

Za předpokladu, že náhodná veličina X , udávající počet chlapců v rodině s dvanácti dětmi, pochází z binomického rozdělení, tj. $X \sim \text{Bin}(N, p)$, s parametry $N = 12$ a $p = 0.5192$, vypočítejte střední hodnotu $E[X]$ a rozptyl $\text{Var}[X]$ náhodné veličiny X . Střední hodnotu a rozptyl porovnejte s jejich odhady vypočítanými na (a) základě očekávaných dat; (b) na základě pozorovaných dat (viz příklad 1.2).

Řešení příkladu 1.5

	E.X	Var.X	E.exp	Var.exp	E.obs	Var.obs
1	6.2304	2.995576	6.229926	2.995539	6.230581	3.48984

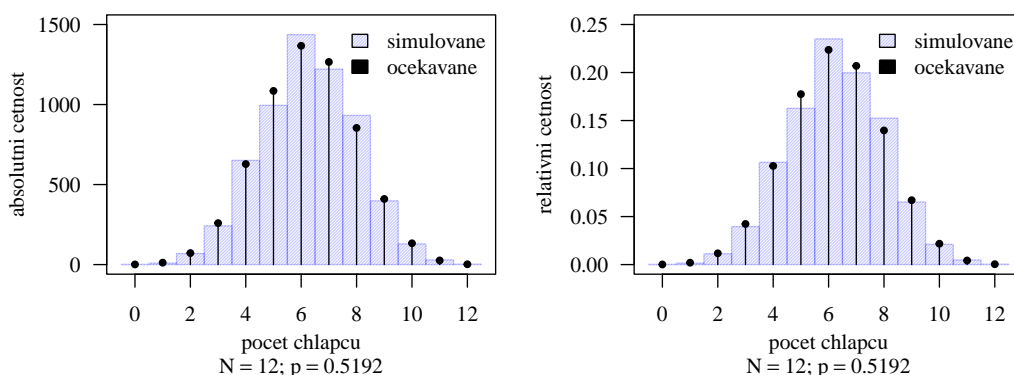
5
6

Střední hodnota počtu chlapců v rodině s dvanácti dětmi je 6.2304 s rozptylem 2.9956. Odhad střední hodnoty počtu chlapců v rodině vypočítaný na základě očekávaných hodnot je 6.2299 s rozptylem 2.9955. Odhad střední hodnoty počtu chlapců v rodině vypočítaný na základě pozorovaných dat je 6.2306 s rozptylem 3.4898. ★

Příklad 1.6. Simulační studie pro binomický model

Sekci o binomickém rozdělení zakončíme simulační studií modelující chování očekávaných početností náhodné veličiny X ... počet chlapců v rodinách s dvanácti dětmi za předpokladu, že $X \sim \text{Bin}(N, p)$, $N = 12$, $p = 0.5192$.

Vygenerujte pseudonáhodná čísla X (početnosti úspěchů) opakovaná M -krát ($M = 6115$) z $\text{Bin}(N, p)$, kde $N = 12$ a $p = 0.5192$. Vytvořte histogram vygenerovaných pseudonáhodných čísel superponovaný hodnotami očekávaných (teoretických) početností za předpokladu, že $X \sim \text{Bin}(N, p)$.

Řešení příkladu 1.6

Obrázek 3: Porovnání pozorovaných a očekávaných početností v binomickém modelu

★