

4 Diskrétní a spojité náhodné veličiny

4.1 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; $i = 1, \dots, N$.
 - $\Pr(X_i = 1) = p$
 - $\Pr(X_i = 0) = 1 - p = q$
- Binomické rozdělení:
 - $X \dots$ 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: $E[X] = Np$; $\text{Var}[X] = Np(1-p)$
- $\text{dbinom}(x, N, p)$, $\text{pbinom}(x, N, p)$

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	\sum
$m_{observed}$	3	24	104	286	670	1033	1343	1112	829	478	181	45	7	6115

Příklad 4.1. Výpočet parametru p binomického modelu

Vezměte údaje z datasetu 1. Předpokládejme, že náhodná veličina X popisující počet chlapců v rodinách s dvanácti dětmi pochází z binomického rozdělení s parametrem $N = 12$. Vypočítejte odhad pravděpodobnosti výskytu chlapců v rodinách s dvanácti dětmi.

Řešení příkladu 4.1

Pravděpodobnost p výskytu chlapců v rodinách s dvanácti dětmi odhadneme pomocí vzorce

$$\hat{p} = \frac{\text{počet narozených chlapců}}{\text{celkový počet narozených dětí}} = \frac{\sum_{n=0}^N n m_{observed}}{NM}. \quad (1)$$

[1] 0.5192

1

Interpretace výsledků: Pravděpodobnost výskytu chlapců v rodinách s dvanácti dětmi je
(..... %).

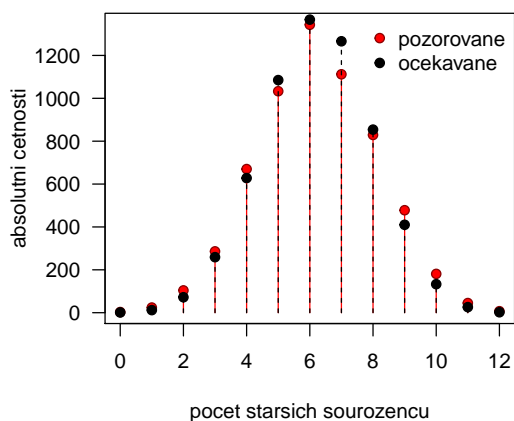
Příklad 4.2. Pozorované a očekávané početnosti v binomickém modelu

Za předpokladu, že počet chlapců v rodinách s dvanácti dětmi pochází z binomického rozdělení s parametry $N = \dots\dots\dots$ a $p = \dots\dots\dots$ odhadněte očekávané početnosti chlapců v rodinách s dvanácti dětmi a porovnejte je s pozorovanými početnostmi.

Řešení příkladu 4.2

	0	1	2	3	4	5	6	7	8	9	10	11	12
m. obs	3	24	104	286	670	1033	1343	1112	829	478	181	45	7
m. exp	1	12	72	259	628	1085	1367	1266	854	410	133	26	2

2
3
4



Obrázek 1: Porovnání pozorovaných a očekávaných početností v Poissonově modelu

Příklad 4.3. Výpočet pravděpodobností za předpokladu binomického modelu

Za předpokladu, že náhodná veličina X popisující počet chlapců v rodinách s dvanácti dětmi pochází z binomického rozdělení s parametry $N = \dots\dots\dots$ a $p = \dots\dots\dots$ vypočítejte pravděpodobnost, že v rodině s dvanácti dětmi bude

- právě devět chlapců,
- nejvýše čtyři chlapci,
- alespoň osm chlapců,
- čtyři, pět, šest, nebo sedm chlapců.

Řešení příkladu 4.3

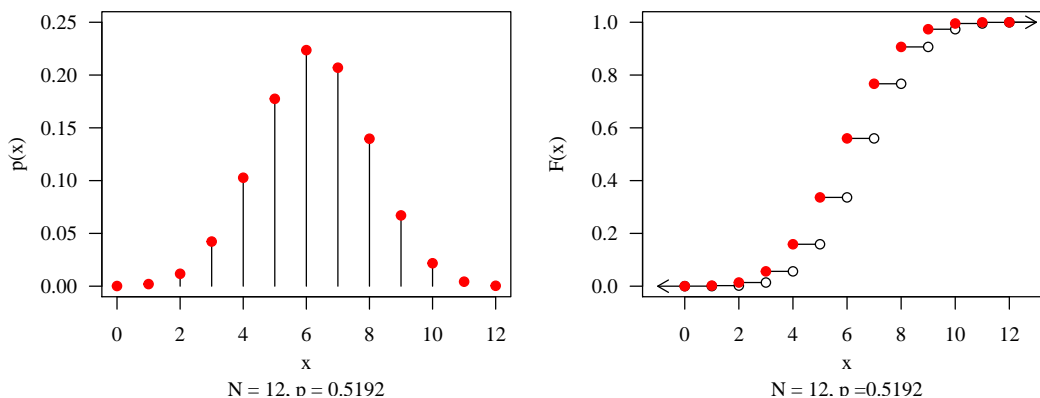
[1] 0.067	5
[1] 0.1589	6
[1] 0.2331	7
[1] 0.7108	8

Interpretace výsledků: Pravděpodobnost, že v rodině bude právě devět chlapců, je%. Pravděpodobnost, že v rodině budou nejvýše čtyři chlapci, je%. Pravděpodobnost, že v rodině bude alespoň osm chlapců, je%. Pravděpodobnost, že v rodině bude čtyři, pět, šest, nebo sedm chlapců, je%.

Příklad 4.4. Graf pravděpodobnostní a distribuční funkce binomického modelu

Nakreslete graf pravděpodobnostní funkce a graf distribuční funkce binomického rozdělení $\text{Bin}(N, p)$, kde $N = 12$ a $p = 0.5192$.

Řešení příkladu 4.4



Obrázek 2: Pravděpodobnostní a distribuční funkce binomického modelu

4.2 Poissonovo rozdělení $\text{Po}(\lambda)$

- X ... počet událostí, které nastanou v jednotkovém časovém intervalu, přičemž k událostem dochází náhodně, jednotlivě a vzájemně nezávisle. Střední počet těchto událostí je vyjádřen parametrem $\lambda > 0$.
- $X \sim \text{Po}(\lambda)$
- $\theta = \lambda$
- pravděpodobnostní funkce:

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x = 0, 1, \dots;$$

- vlastnosti: $E[X] = \lambda$; $\text{Var}[X] = \lambda$
- `dpois(x, lambda)`, `ppois(x, lambda)`

Příklad 4.5. Výpočet parametru λ Poissonova modelu

Načtete datový soubor `17-anova-newborns.txt` a odstraňte z něj neznámá pozorování. Zaměřte se na znak $X = \text{počet starších sourozenců}$ novorozence. Za předpokladu, že náhodná veličina X popisující počet starších sourozenců novorozence pochází z Poissonova rozdělení parametrem λ odhadněte střední hodnotu počtu starších sourozenců λ .

Řešení příkladu 4.5

Střední hodnotu počtu starších sourozenců odhadneme pomocí vzorce

$$\lambda = \frac{\text{počet starších sourozenců}}{\text{počet novorozenců}} = \frac{\sum_{i=1}^N x_i}{N}. \quad (2)$$

[1] 0.9428365

9

Interpretace výsledků: Střední hodnota počtu starších sourozenců novorozenců v datovém souboru $\lambda = \dots\dots\dots$

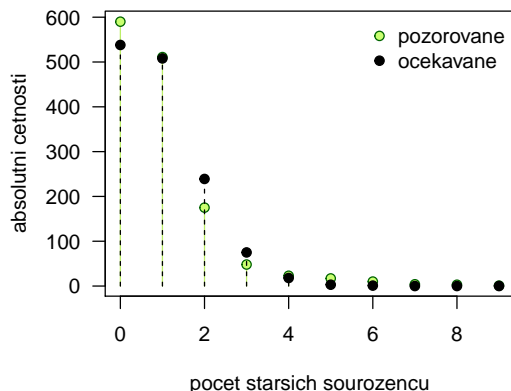
Příklad 4.6. Porovnání pozorovaných a očekávaných početností v Poissonově modelu

Za předpokladu, že počet starších sourozenců novorozenců pochází z Poissonova rozdělení s parametrem $\lambda = \dots\dots\dots$ $\dots\dots\dots$ odhadněte očekávané početnosti starších sourozenců a porovnejte je s pozorovanými početnostmi.

Řešení příkladu 4.6

	0	1	2	3	4	5	6	7	8	9
m. obs	590	511	175	48	23	17	10	4	3	1
m. exp	538	508	239	75	18	3	1	0	0	0

10
11
12



Obrázek 3: Porovnání pozorovaných a očekávaných početností v Poissonově modelu

Příklad 4.7. Výpočet pravděpodobností za předpokladu Poissonova modelu

Vraťem se nyní k příkladu 4.5. Za předpokladu, že data pochází z Poissonova rozdělení s parametrem $\lambda = \dots\dots\dots$ určete pravděpodobnost, novorozenec má

1. dva, tři nebo čtyři starší sourozence
2. alespoň čtyři starší sourozence
3. nejvýše dva
4. právě tři

Řešení příkladu 4.7

[1] 0.2403672 13

[1] 0.01568161 14

[1] 0.9299071 15

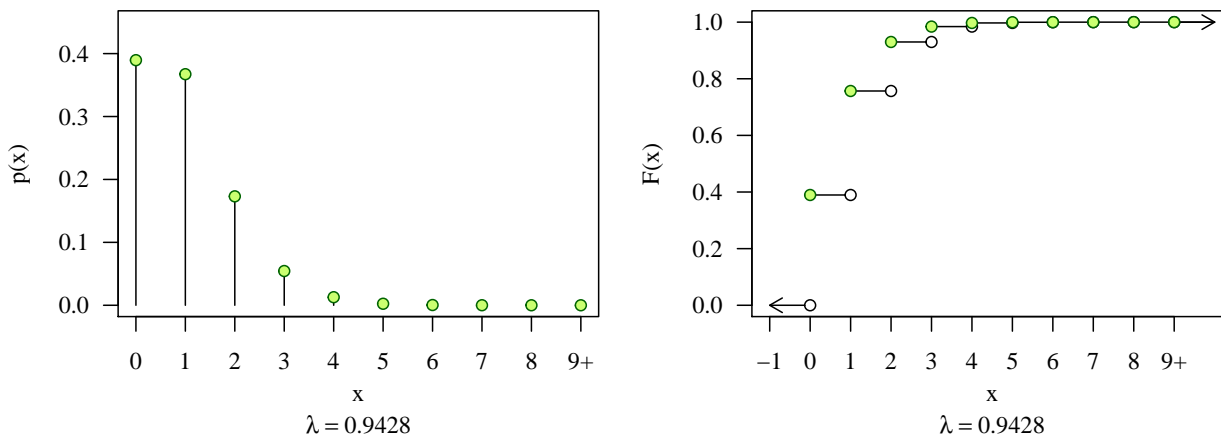
[1] 0.367255 16

Interpretace výsledů: Pravděpodobnost, že novorozenec bude mít dva, tři nebo čtyři starší sourozence je% . Pravděpodobnost, že novorozenec bude mít alespoň čtyři starší sourozence je% . Pravděpodobnost, že novorozenec bude mít nejvýše dva starší sourozence je% . Pravděpodobnost, že novorozenec bude mít jednoho staršího sourozence je% .

Příklad 4.8. Graf pravděpodobnostní a distribuční funkce Poissonova modelu

Nakreslete graf pravděpodobnostní a distribuční funkce Poissonova rozdělení $Po(0.9428)$ v hodnotách $x = 0, 1, 2, 3, 4, 5, 6, 7, 8$, a $x \geq 9$.

Řešení příkladu 4.8



Obrázek 4: Pravděpodobnostní a distribuční funkce Poissonova modelu

4.3 Normální rozdělení $N(\mu, \sigma^2)$

- $X_1, \dots, X_n \dots$ nezávislé náhodné veličiny
- Normální rozdělení

– $X \sim N(\mu, \sigma^2)$

– $\theta = (\mu, \sigma^2)^T$

– hustota

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}.$$

– vlastnosti $E[X] = \mu$; $\text{Var}[X] = \sigma^2$

– `dnorm(x, mu, sigma)`, `pnorm(x, mu, sigma)`, `rnorm(M, mu, sigma)`, `qnorm(alpha, mu, sigma)`

- Standardizované normální rozdělení

– $X \sim N(0, 1)$

– $\theta = (0, 1)^T$

– hustota

$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}, \quad x \in \mathbb{R}.$$

– vlastnosti $E[X] = 0$; $\text{Var}[X] = 1$

– `dnorm(x)`, `pnorm(x)`, `rnorm(M)`, `qnorm(alpha)`

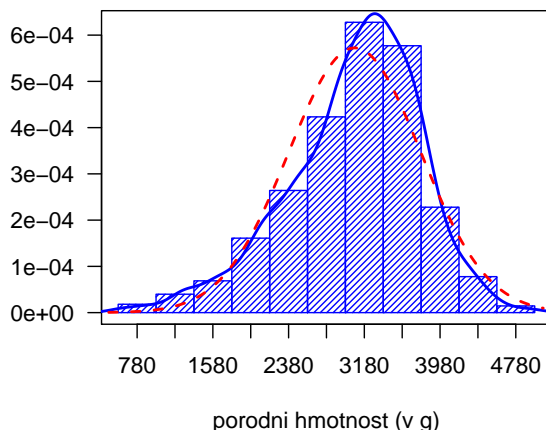
- Vlastnosti normálního rozdělení

– **Věta 1:** Necht' X_1, \dots, X_n jsou nezávislé náhodné veličiny z normálního rozdělení $N(\mu, \sigma^2)$. Potom náhodná veličina $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Příklad 4.9. Výpočet pravděpodobností na základě normálního modelu

Mějme datový soubor obsahujícího údaje o porodní hmotnosti novorozenců v jedné okresní nemocnici za období jednoho roku (Alánová, 2008). Za předpokladu, že náhodná veličina X popisující porodní hmotnost novorozenců pochází z normálního rozdělení $N(\mu, \sigma^2)$ odhadněte parametr střední hodnoty μ a rozptylu σ^2 . Finální rozdělení porovnejte s naměřenými údaji.

Řešení příkladu 4.9



Interpretace výsledků: Náhodná veličina X popisující porodní hmotnost novorozenců pochází z normálního rozdělení se střední hodnotou $\mu = \dots\dots\dots$ a rozptylem $\dots\dots\dots$, tj. $X \sim N(\dots\dots\dots)$;

Příklad 4.10. Výpočet pravděpodobností na základě normálního modelu

Na základě datového souboru obsahujícího údaje o porodní hmotnosti novorozenců v jedné okresní nemocnici za období jednoho roku (Alánová, 2008) byla odhadnuta střední hodnota a směrodatná odchylka porodní hmotnosti novorozenců. Střední hodnota $\mu = 3078.94$ g, směrodatná odchylka $s = 697$ g. Za předpokladu, že data pochází z normálního rozdělení vypočítejte, jaká je pravděpodobnost, že porodní hmotnost novorozence bude (a) menší než 3800 g; (b) v rozmezí 2000–3000 g; (c) větší než 4000 g, (d) rovná 2100 g.

Řešení příkladu 4.10

- [1] 0.8495533 17
- [1] 0.743032 18
- [1] 0.09317345 19
- [1] 0 20

Interpretace výsledků: Pravděpodobnost, že porodní hmotnost novorozenců bude menší než 3800 g, je $\dots\dots\dots\%$. Pravděpodobnost, že porodní hmotnost novorozenců bude v rozmezí 2500–4200 mm, je $\dots\dots\dots\%$. Pravděpodobnost, že porodní hmotnost novorozenců bude větší než 4000 mm, je $\dots\dots\dots\%$. Pravděpodobnost, že porodní hmotnost novorozenců bude rovná 2100 g, je $\dots\dots\dots\%$, protože data pochází z normálního rozdělení, což je $\dots\dots\dots$ typ rozdělení, proto $\Pr(X = 2100) = \dots\dots\dots$

Příklad 4.11. Výpočet pravděpodobností na základě standardizovaného normálního modelu

Vraťme se nyní k předchozímu příkladu 4.9. Za předpokladu, že porodní hmotnost novorozenců pochází z normálního rozdělení $N(3078.94, 697^2)$ vypočítejte pravděpodobnost, že porodní hmotnost novorozence bude (a) menší než 3800 g; (b) v rozmezí 2000–3000 g; (c) větší než 4000 g, (d) rovná 2100 g. Řešení proveďte přes standardizaci náhodné veličiny X .

Řešení příkladu 4.11

[1] 0.8495533 21

[1] 0.743032 22

[1] 0.09317345 23

[1] 0 24

Interpretace výsledků: Pravděpodobnost, že porodní hmotnost novorozenců bude menší než 3800 g je
.....%. Pravděpodobnost, že porodní hmotnost novorozenců bude v rozmezí 2500–4200 mm je%.
Pravděpodobnost, že porodní hmotnost novorozenců bude větší než 4000 mm je%. Pravděpodob-
nost, že porodní hmotnost novorozenců bude rovná 2100 g je%, protože data pochází z normálního
rozdělení, což je typ rozdělení, proto $\Pr(X = 2100) = \dots\dots\dots$.

Příklad 4.12. Výpočet pravděpodobností na základě normálního modelu

Předpokládejme, že velký ročník na VŠ má výsledky ze statistiky normálně rozdělené kolem střední hodnoty 72 bodů se směrodatnou odchylkou 9 bodů. Najděte pravděpodobnost, že průměr výsledků náhodného výběru 10-ti studentů bude větší než 80 bodů.

Řešení příkladu 4.12

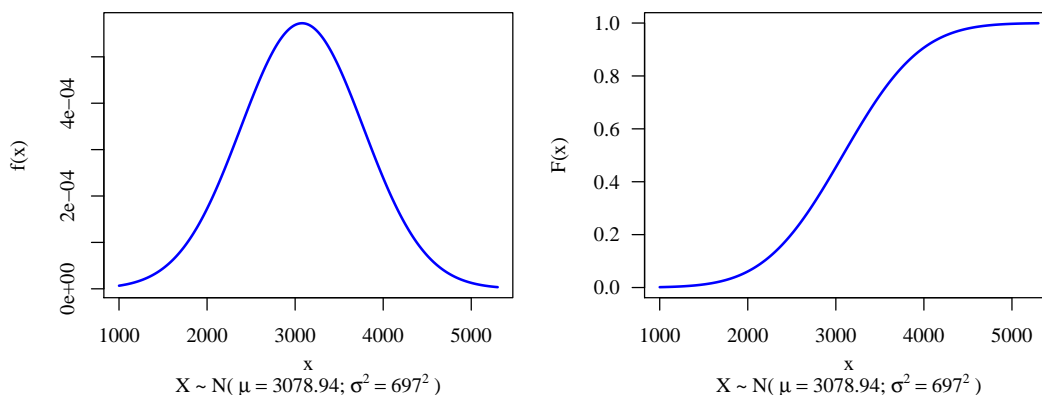
[1] 0.002470053 25

Interpretace výsledků: Pravděpodobnost, že průměr výsledků náhodného výběru 10-ti studentů bude větší než 80 bodů je %.

Příklad 4.13. Graf hustoty a distribuční funkce normálního modelu

Vraťme se k příkladu 4.9. Nakreslete graf hustoty a distribuční funkce náhodné veličiny $X \sim N(3078.94, 697)$.

Řešení příkladu 4.13



4.4 Aproximace binomického modelu normálním modelem

- Normální rozdělení je limitním rozdělením binomického rozdělení $\text{Bin}(N, p)$, tedy pro $N \rightarrow \infty, p \rightarrow 0.5$:

$$X \sim \text{Bin}(N, p) \rightarrow X \sim N(\mu, \sigma^2),$$

kde $\mu = Np$ a $\sigma^2 = Np(1 - p)$.

- Haldova podmínka: Necht' $X \sim \text{Bin}(N, p)$ a platí, že $Np > 5$ a $N(1 - p) > 5$. Potom rozdělení náhodné proměnné X můžeme aproximovat normálním rozdělením $X \sim N(Np, Np(1 - p))$.
- Výše zmíněný poznatek je také znám jako **Moivre-Laplaceova věta**.

Příklad 4.14. Aproximace binomického modelu normálním modelem

Předpokládejme, že pravděpodobnost výskytu dermatoglyfického vzoru *vír* na palci pravé ruky u mužů české populace $p = 0.533$.

- Jaká je pravděpodobnost, že ve vybraném vzorku 10 mužů bude výskyt dermatoglyfického vzoru *vír* na palci pravé ruky (a) alespoň u sedmi mužů; (b) nejvýše u pěti mužů; (c) u osmi nebo devíti mužů.
- Jaká je pravděpodobnost, že ve vybraném vzorku 100 mužů bude výskyt dermatoglyfického vzoru *vír* na palci pravé ruky (a) alespoň u 56; (b) nejvýše u 53 mužů; (c) u 60–85 mužů.

Požadované pravděpodobnosti vypočítejte exaktně na základě binomického rozdělení a aproximačně na základě normálního rozdělení. Výsledné hodnoty navzájem porovnejte.

Řešení příkladu 4.14

	alespon 7	nejvyse 5	8-9	
binomicke	0.2313	0.5396	0.0801	26
normalni	0.1449	0.4172	0.0353	27

	alespon 56	nejvyse 53	60-85	
binomicke	0.3304	0.5151	0.1067	29
normalni	0.2942	0.4760	0.0896	30

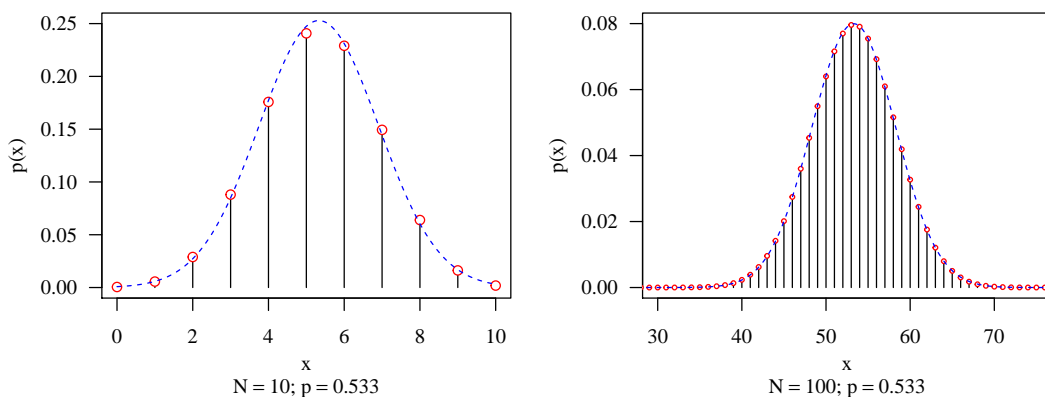
Interpretace výsledků: Pravděpodobnost výskytu dermatoglyfického vzoru *vír* alespoň u sedmi mužů z deseti je% (resp.%). Pravděpodobnost, výskytu vzoru *vír* nejvýše u pěti mužů z deseti je% (resp.%). Pravděpodobnost, výskytu vzoru *vír* u osmi nebo devíti mužů z deseti je% (resp.%). Protože Haldova podmínka dobré aproximace splněna, bychom aproximaci binomického rozdělení normálním rozdělením použít.

Pravděpodobnost výskytu dermatoglyfického vzoru *vír* alespoň u 56 mužů ze sta je% (resp.%). Pravděpodobnost výskytu vzoru *vír* nejvýše u 53 mužů ze sta je% (resp.%). Pravděpodobnost výskytu vzoru *vír* u 60–85 mužů ze sta je% (resp.%). Protože Haldova podmínka dobré aproximace splněna, aproximaci binomického rozdělení normálním rozdělením použít.

Příklad 4.15. Aproximace binomického modelu normálním modelem

Předpokládejme, že pravděpodobnost výskytu dermatoglyfického vzoru *vír* na palci pravé ruky u mužů české populace $p = 0.533$. Pro $N = 10$ a $N = 100$ vykreslete graf pravděpodobnostní funkce binomického rozdělení a aproximujte jej křivkou funkce hustoty normálního rozdělení. Hodnoty obou funkcí porovnejte.

Řešení příkladu 4.15



Obrázek 5: Aproximace binomického modelu normálním modelem