UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

# BAKALÁŘSKÁ PRÁCE

## Bayesovský přístup k stanovení rozsahu výběru v klinických studiích

**Katedra matematické analýzy a aplikací matematiky**

Vedoucí práce: **doc. Mgr. Ondřej Vencálek, Ph.D.**

Vypracovala: **Elizabeth Princová**

Studijní program: Aplikovaná matematika

Studijní obor: Data Science

Forma studia: prezenční

Rok odevzdání: 2025

# BIBLIOGRAFICKÁ IDENTIFIKACE

**Autor:** Elizabeth Princová

**Název práce:** Bayesovský přístup k stanovení rozsahu výběru v klinických studiích

**Typ práce:** Bakalářská práce

**Pracoviště:** Katedra matematické analýzy a aplikací matematiky

**Vedoucí práce:** doc. Mgr. Ondřej Vencálek, Ph.D.

**Rok obhajoby práce:** 2025

**Abstrakt:** Bayesovský přístup ke statistickému uvažování nachází stále širší uplatnění. Výjimkou není ani oblast klinického výzkumu. Tato práce představuje využití bayesovských metod při návrhu klinických studií, konkrétně při stanovení rozsahu výběru. Úvodní část vysvětluje pojem klinické studie a popisuje, jakým způsobem je vhodné v tomto kontextu formulovat hypotézy. Dále jsou představeny základní principy a pojmy bayesovské inference. Stěžejní část práce je věnována konkrétním metodám pro stanovení rozsahu výběru. Tyto metody jsou vysvětleny a představeny s využitím statistického softwaru R na ilustrativním příkladu klinické studie vytvořeném pro účely této práce.

**Klíčová slova:** klinická studie, bayesovská statistika, apriorní rozdělení, velikost výběru, síla testu

**Počet stran:** 60

**Počet příloh:** 1

**Jazyk:** český

# BIBLIOGRAPHICAL IDENTIFICATION

**Author:** Elizabeth Princová

**Title:** A Bayesian approach to determining sample size in clinical trials

**Type of thesis:** Bachelor's

**Department:** Department of Mathematical Analysis and Application of Mathematics

**Supervisor:** doc. Mgr. Ondřej Vencálek, Ph.D.

**The year of presentation:** 2025

**Abstract:** The bayesian approach to statistical reasoning is finding increasing application, including in the field of clinical trials. This thesis presents the use of Bayesian methods in the design of clinical trials, specifically for determining sample size. The introductory part explains the concept of a clinical trial and describes how hypotheses should be formulated in this context. Next, the basic principles and concepts of Bayesian inference are presented. The main part of the thesis is dedicated to specific methods for determining sample size. These methods are explained and demonstrated using the statistical software R on an illustrative example of a clinical trial designed for the purposes of this thesis.

**Key words:** clinical trial, bayesian statistics, prior distribution, sample size, power of a test

**Number of pages:** 60

**Number of appendices:** 1

**Language:** Czech

**Prohlášení**

Prohlašuji, že jsem bakalářskou práci zpracovala samostatně pod vedením pana doc. Mgr. Ondřeje Vencálka, Ph.D., a všechny použité zdroje jsem uvedla v seznamu literatury.

V Olomouci dne..... .....

podpis

# Obsah

# Seznam obrázků

**Poděkování**

Mé poděkování patří vedoucímu této práce, panu doc. Mgr. Ondřeji Vencálkovi, Ph.D., za velkou podporu, ochotu, odborné vedení a veškerý čas, který mi při zpracování této bakalářské práce věnoval. Dále bych ráda poděkovala své rodině a přátelům za jejich podporu.

# Úvod

Zdraví představuje jednu z nejcennějších hodnot, které člověk má. V době, kdy čelíme civilizačním chorobám, nově se objevujícím onemocněním a dalším zdravotním komplikacím, je důležité zaměřit pozornost na vývoj nových léčiv a efektivních léčebných postupů. Možnost tato léčiva zkoumat a ověřovat jejich účinnost nám poskytují klinické studie.

Uskutečnění klinické studie je dlouhý proces, jehož nedílnou součástí je i volba vhodného počtu účastníků. Podobně jako se člověk během života spoléhá na své zkušenosti, můžeme i ve statistice využívat informace získané v rámci dříve provedených klinických studií – a to nejen při volbě počtu účastníků, ale v rámci celé klinické studie obecně. Tyto informace nám umožňují získat reálnější pohled na danou problematiku. Statistika, která tento princip využívá, se nazývá bayesovská.

Cílem této práce je představit metody pro stanovení rozsahu výběru v klinických studiích v kontextu bayesovské statistiky.

V první kapitole přiblížíme, co se skrývá pod pojmem klinická studie a jakým způsobem můžeme měřit účinnost nového léčiva. Ve druhé kapitole vysvětlíme princip testování hypotéz v kontextu klasické frekventistické statistiky a představíme možné výsledky tohoto testování. Ve třetí kapitole čtenáře seznámíme s postupem bayesovské inference a klíčovými pojmy, jako jsou apriorní rozdělení, věrohodnost a aposteriorní rozdělení. Dále si představíme ilustrativní klinickou studii, na jejímž základě ve čtvrté kapitole ukážeme několik metod, které nám mohou pomoci při určování počtu pacientů v rámci klinické studie. Uvedené metody budeme demonstrovat pomocí statistického softwaru R.

# 1. Klinická studie

V této kapitole si představíme, co je to klinická studie a jak v rámci ní můžeme měřit a porovnávat účinnost zkoumaného léčiva. Vycházet budeme především z knihy *Klinické studie v praxi* (viz [5]).

Klinická studie je výzkumný proces zaměřený na hodnocení účinnosti a bezpečnosti léčebných procedur, terapeutických postupů a léčivých přípravků, mezi něž patří různé látky s léčebnými účinky, vakcíny, séra, alergenní přípravky a jiné (viz [6]). Studie zahrnuje skupiny účastníků, které mohou tvořit nemocní pacienti, zdraví dobrovolníci nebo osoby vystavované určitému jevu, jako je například znečištěné ovzduší, kontaminovaná voda, nadměrné sluneční záření, hluková zátěž, chemické látky v pracovním prostředí nebo pasivní kouření. Na těchto skupinách sledujeme efekt léčiva.

K tomu, abychom mohli vyhodnotit, zda je efekt pozitivní či negativní, potřebujeme ke zkoumané skupině i skupinu kontrolní. Této skupině je podávána odlišná dávka či forma nového léku, jiná standardně používaná léčba, placebo, nebo není podávána léčba žádná. Placebem rozumíme přípravek, který přímo neobsahuje aktivní látky působící negativně či pozitivně na lidský organismus. Nastat však může placebo efekt. To znamená, že dochází ke zlepšení či zhoršení stavu pacienta na základě toho, že se domnívá, že mu byl aktivní lék podán.

O novém léku pak můžeme říci, zda je prospěšnější, horší, nebo zda má srovnatelné účinky vzhledem k léčbě použité v kontrolní skupině. K měření účinnosti daného léčiva přistupujeme mnoha způsoby. Předpokládejme, že máme dvě skupiny lidí s vysokým krevním tlakem. Skupině A budeme podávat placebo a skupině B zkoumaný lék. V prvním případě můžeme u skupin zjistit a porovnat procentuální zastoupení jedinců, jejichž stav se zlepšil, do-

šlo u nich ke zmírnění příznaků, případně se u nich stav zhoršil. Sledujeme tedy procento výskytu nějakého jevu. Ve druhém případě můžeme vyjádřit rozdíl mezi skupinami porovnáním odhadů jejich číselných charakteristik (např. střední hodnoty), v tomto případě krevního tlaku.

Na základě výsledků klinické studie můžeme následně rozhodnout o využití zkoumaného léčiva.

# 2. Testování hypotéz

Cílem této kapitoly je objasnit konstrukci hypotéz v klinických studiích a představit možné výsledky při jejich testování. Čerpat budeme z kapitoly 5.5 knihy *Základy počtu pravděpodobnosti a metod matematické statistiky* (viz [3]).

Představme si dva případy klinických studií. V obou případech budou zahrnuty dvě skupiny lidí trpících vysokým tlakem. Skupina A bude skupinou kontrolní k testované skupině B, jejíž pacienti dostanou nové zkoumané léčivo. V obou případech půjde o to zjistit, jak si vede nový lék v porovnání s léčbou použitou ve skupině A. V prvním případě dostane kontrolní skupina A standardně používané léčivo a ve druhém případě dostane skupina A placebo.

Před samotným hodnocením léčiva je důležité rozhodnout, kdy bude měření, v tomto konkrétním případě tlaku, probíhat. Jednou z možností je, že změříme každému pacientovi tlak opakovaně. První měření provedeme před začátkem léčby a druhé po jeho ukončení. Dále budeme pracovat se změnami, kterých bylo u jednotlivých pacientů dosaženo. Můžeme se ale také omezit pouze na jedno měření. Na začátku studie budeme předpokládat, že je rozdělení tlaku v obou skupinách stejné. Po ukončení léčby změříme každému pacientovi v dané skupině tlak a tyto hodnoty dále využijeme pro zkoumání účinnosti. Přesná doba měření – zda proběhne bezprostředně před či po zahájení nebo ukončení léčby, případně s určitým časovým odstupem – záleží na konkrétním designu studie. V uvedeném příkladu vycházíme z druhého zmíněného způsobu.

Rozdíl mezi skupinami A a B budeme v obou případech měřit prostřednictvím teoretických středních hodnot krevního tlaku daných skupin. Pro skupinu A tuto charakteristiku označíme jako $\bar{X}_1$ a pro skupinu B jako $\bar{X}_2$. Hy-

potéza se bude týkat rozdílu těchto dvou charakteristik $\bar{X}_1 - \bar{X}_2$, který si označíme jako $\theta$.

Při testování konstruujeme vždy dvě hypotézy, hypotézu nulovou a alternativní. Tyto hypotézy stanovujeme s ohledem na to, že při testování hypotéz nejsme schopni nulovou hypotézu přijmout. Pokud nemáme dostatek důkazů, které by mluvily v její neprospěch, můžeme ji pouze nezamítat. V opačném případě ji můžeme zamítnout ve prospěch alternativy. Jako nulovou hypotézu budeme volit $H_0 \colon \theta \leq 0$. V prvním případě je interpretace této hypotézy taková, že nový lék má stejný účinek jako lék běžně používaný nebo je jeho účinek horší. V případě použití placeba tato hypotéza značí žádný nebo negativní účinek nového léku. Pro číselné charakteristiky skupin bude v obou případech pro uvedenou hypotézu platit $\bar{X}_1 \leq \bar{X}_2$. V kontextu problému s vysokým krevním tlakem vnímáme jako pozitivní účinek léku snížení tlaku, proto pokud nastane uvedená nerovnost, tedy charakteristika krevního tlaku skupiny A bude menší než skupiny B, nebude se jednat o pozitivní účinek nového léku. Zamítnutí nulové hypotézy značí přijetí alternativní hypotézy $H_A \colon \theta > 0$. V prvním případě by to znamenalo, že nový lék má lepší účinek než lék běžný, a v případě druhém by se mluvilo o obecně pozitivním efektu nového léku. Vyjádřeno pomocí číselných charakteristik by se jednalo o případ $\bar{X}_1 > \bar{X}_2$.

Při testování hypotéz v klasické inferenci mohou nastat čtyři následující případy:

| | | **Skutečnost** | |
|---|---|---|---|
| | | $H_0$ platí | $H_A$ platí |
| **Závěr testu** | $D_0$ = nezamítám $H_0$ | správné rozhodnutí $p(D_0\|H_0)$ | chyba 2. druhu $p(D_0\|H_A)$ |
| | $D_A$ = zamítám $H_0$ | chyba 1. druhu $p(D_A\|H_0)$ | správné rozhodnutí $p(D_A\|H_A)$ |

Tabulka 1: Možné výsledky při testování statistických hypotéz a jejich pravděpodobnosti.

Pro oba možné závěry testu (zamítnutí či nezamítnutí nulové hypotézy) sledujeme jejich pravděpodobnosti při platnosti, resp. neplatnosti nulové hypotézy.

Klíčovou roli hraje především chyba prvního a druhého druhu. Chyba prvního druhu nastává, když nulová hypotéza platí, ale my ji na základě testu zamítáme. V demonstrovaném příkladu by tato chyba nastala v případě, když by nový lék neměl lepší účinek než lék běžný, resp. by měl v příkladu s placebem negativní nebo žádný účinek, ale závěr studie by byl takový, že je lék prospěšnější, resp. obecně účinný. Protože tuto chybu chápeme ve statistice jako závažnější, omezujeme její pravděpodobnost vždy nějakým malým číslem $\alpha$, $0 < \alpha < 1$, které nazýváme hladina významnosti. Nejčastěji se za toto číslo volí hodnoty 0,05, 0,01 a 0,001. Chyba druhého druhu nastává, když platí alternativní hypotéza a my nulovou hypotézu nezamítáme. Nový lék by tedy měl ve skutečnosti lepší účinek, resp. by byl obecně prospěšný, ale naše studie by to neprokázala.

Dalším důležitým aspektem je síla testu $1 - \beta = p(D_A|H_A)$, kde je $\beta$ rovna podmíněné pravděpodobnosti chyby druhého druhu. Za $\beta$ nejčastěji volíme hodnotu 0,2. Silou testu rozumíme pravděpodobnost, že při neplatnosti nulové hypotézy dojde k jejímu zamítnutí, jinak řečeno, jak moc je test schopný

detekovat efekt léku, pokud skutečně existuje. V prvním ze zmiňovaných případů jde tedy o pravděpodobnost, že zamítáme nulovou hypotézu o jiném než lepším efektu oproti léku běžnému za podmínky, že je lék ve skutečnosti opravdu prospěšnější. Obdobně tomu bude i v případě s placebem, kde jde o pravděpodobnost, že zamítáme nulovou hypotézu o jiném než pozitivním efektu za podmínky, že je lék ve skutečnosti prospěšný.

# 3. Bayesovský přístup ke klinické studii

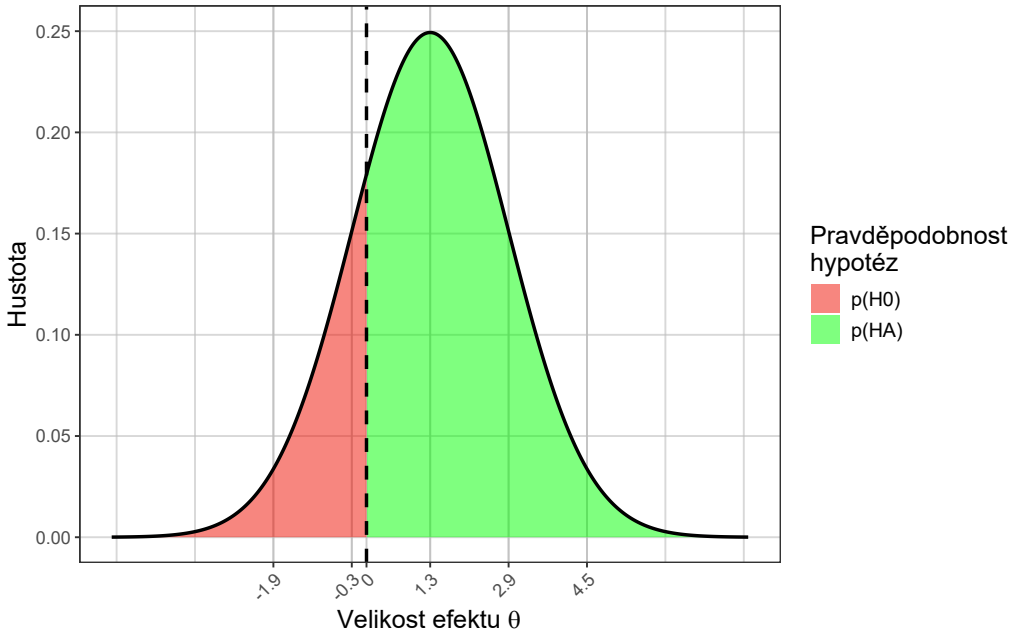
V této části představíme postup bayesovské inference. Dále vysvětlíme klíčové pojmy, jimiž jsou apriorní rozdělení, věrohodnost a aposteriorní rozdělení. V závěru představíme klinickou studii, na níž budeme dále v práci demonstrovat metody pro zohlednění apriorní představy.

## 3.1. Zapojení apriorní představy

Následující kapitola přiblíží, jakým způsobem můžeme do testování hypotéz zapojit informaci získanou z předchozích studií. Vycházet budeme především z kapitoly 6.5 knihy *Bayesian approaches to clinical trials and health-care evaluation* (viz [4]).

V předchozí kapitole jsme pracovali s tím, že je parametr $\theta$ větší, anebo menší či roven 0. Hypotézy jsme konstruovali pouze na základě našeho subjektivního přesvědčení či názoru, jinak řečeno na základě neformální apriorní informace. Tento typ apriorní představy je typický pro klasickou frekventistickou statistiku. Bayesovský přístup ke statistice umožňuje oproti frekventistickému přístupu zapojení apriorní představy o hodnotě zkoumaného parametru, resp. o platnosti různých hypotéz týkajících se tohoto parametru. Apriorní představa je dána tím, jak jsou ještě před zahájením studie jednotlivé hodnoty parametru $\theta$ pravděpodobné. Apriorní představu běžně získáváme z předchozích výzkumů, statistické analýzy, popřípadě dat týkajících se stejné problematiky. Tyto představy se promítají do pravděpodobnosti jednotlivých hypotéz. Nově se tedy bude navíc uvažovat $p(H_0)$ a $p(H_A)$, viz obr. 1.

V této práci se zaměříme na kombinaci obou přístupů – frekventistického a bayesovského. Půjde tedy o přístup hybridní. Pro design studie využijeme




Obrázek 1: Graf normálního rozdělení se střední hodnotou 1,3 a směrodatnou odchylkou 1,6 znázorňující apriorní rozdělení velikosti efektu $\theta$. Přerušovanou čarou je znázorněna hodnota 0, která od sebe odděluje nulovou a alternativní hypotézu. Červená plocha znázorňuje pravděpodobnost $H_0\colon \theta \leq 0$ a zelená pravděpodobnost $H_A\colon \theta > 0$. Na ose $x$ jsou znázorněny $\pm$ jedna a dvě směrodatné odchylky.

formální apriorní představy typické pro bayesovský přístup, ale samotné testování bude probíhat pouze na úrovni statistiky klasické.

Soustřeďme se dále na „nejpříznivější“ výsledek studie, tedy ten, že zamítáme nulovou hypotézu ve prospěch alternativní hypotézy a ta také ve skutečnosti platí. V případě klasické inference stavíme pravděpodobnosti možných výsledků při testování na pravděpodobnosti podmíněné – zamítnutí $H_0$ podmiňujeme skutečností, že $H_A$ platí. Z definice podmíněné pravděpodobnosti (viz [3], definice 3.9) pro tento případ můžeme psát, že

$$p(D_A|H_A) = \frac{p(D_A, H_A)}{p(H_A)},$$

a tedy

$$p(D_A, H_A) = p(D_A|H_A)p(H_A) = (1-\beta)p(H_A). \tag{1}$$

Přechodem od podmíněné pravděpodobnosti k pravděpodobnosti sdružené zohledňujeme členem $p(H_A)$ apriorní představu. Tímto můžeme dosáhnout reálnějšího návrhu studie.

V případě, kdy bude apriorní představa podporovat zamítnutí $H_0$, pravděpodobnost $p(H_0)$ bude malá a $p(H_A)$ naopak velká, bude apriorem upravená síla testu $(1-\beta)p(H_A)$ z rovnice 1 blízká pravděpodobnosti zamítnutí nulové hypotézy

$$p(D_A) \approx p(D_A, H_A) = (1-\beta)p(H_A).$$

## 3.2. Bayesovská inference o střední hodnotě normálního rozdělení při známém rozptylu

V této části třetí kapitoly si na jednoduchém příkladu ukážeme, jakým způsobem můžeme bayesovsky pohlížet na odhad střední hodnoty normálního rozdělení nějaké veličiny za předpokladu, že známe rozptyl tohoto rozdělení. Princip bayesovské statistiky spočívá v tom, že kombinujeme apriorní rozdělení, tedy nějakou představu o odhadovaném parametru (viz 3.2.1), a věrohodnost, informaci o parametru získanou ze současných dat (viz 3.2.2), do podoby aposteriorního rozdělení (viz 3.2.3).

### 3.2.1. Apriorní rozdělení

Předpokládejme náhodný výběr $X_1, X_2, \ldots, X_{n_0}$ z rozdělení $N(\theta, \sigma^2)$. Náhodná veličina $X_i$ může například značit naměřený krevní tlak $i$-tého náhodně vybraného pacienta. Přitom $\sigma^2 > 0$ známe a $\theta \in \mathbb{R}$ považujme za neznámý parametr. Za pomoci těchto měření budeme chtít odhadnout střední hodnotu, tedy průměrný krevní tlak, rozdělení, ze kterého tento výběr pochází. Jelikož nemáme o hodnotě parametru žádnou informaci, která by nám mohla napovědět, jakých hodnot krevní tlak pravděpodobně nabývá, budeme zatím k odhadu přistupovat jako v klasické (frekventistické) statistice.

Střední hodnotu rozdělení můžeme odhadnout pomocí výběrového průměru $\bar{X}_{n_0} = \frac{1}{n_0} \sum_{i=1}^{n_0} X_i$. Následující vlastnosti tohoto odhadu shrnuje věta 6.3 a její důkaz z knihy [3]. Jak z této věty plyne, je střední hodnota veličiny $\bar{X}_{n_0}$ rovna $\theta$, a tento odhad je tedy nestranným odhadem střední hodnoty veličiny $X$. Rozptyl veličiny $\bar{X}_{n_0}$ je přímo úměrný rozptylu veličiny $X$ a současně nepřímo úměrný počtu pozorování, v našem případě $n_0$. Odhad $\bar{X}_{n_0}$ bude stejně jako výběr, ze kterého je počítán, normálně rozdělený, takže

můžeme psát $\bar{X}_{n_0} \sim N\left(\theta, \frac{\sigma^2}{n_0}\right)$.

Tento odhad můžeme nyní brát jako apriorní představu o parametru $\theta$ pro budoucí studie. Rozdělení parametru bude pocházet opět z normálního rozdělení s nejpravděpodobnější hodnotou rovnou právě našemu odhadu $\bar{X}_{n_0}$. Pro tyto účely si označíme hodnotu $\bar{X}_{n_0}$ jako konstantu $\mu_{prior}$. Nejistota parametru bude vyjádřena výše uvedeným rozptylem veličiny $\bar{X}_{n_0}$. Konečně tedy získáme $\theta \sim N\left(\mu_{prior}, \frac{\sigma^2}{n_0}\right)$.

### 3.2.2. Věrohodnost

V této části přejdeme ke studii současné. Opět budeme mít rozdělení $N(\theta, \sigma^2)$ a z něj náhodný výběr $X_1, X_2, \ldots, X_n$, tentokrát s velikostí výběru $n$. Stejně jako v předchozím případě budeme chtít ze získaných dat odhadnout parametr $\theta$, tedy střední hodnotu krevního tlaku, za pomoci veličiny $\bar{X}_n$. Díky větě 6.3 z knihy [3] můžeme opět odvodit rozdělení pro tento odhad, tedy $\bar{X}_n|\theta \sim N\left(\theta, \frac{\sigma^2}{n}\right)$. Toto rozdělení nám umožní definovat věrohodnostní funkci (likelihood), která vyjadřuje, jak věrohodné jsou různé hodnoty parametru $\theta$ při daných datech. Záměrně nemluvíme o „pravděpodobnosti“ parametru při daných datech (z důvodu uvedeného v následujícím odstavci).

Pro lepší interpretaci budeme přistupovat přímo k realizacím veličin, tedy $x_1, x_2, \ldots, x_n$. Protože jsou veličiny $X_1, X_2, \ldots, X_n$ nezávislé, můžeme sdruženou hustotu $p(x_1, x_2, \ldots, x_n|\theta)$ vyjádřit součinem $\prod_{i=1}^{n} p(x_i|\theta)$ (viz [3], věty 3.9 a 3.10). Tuto funkci můžeme označit jako $g(x_1, x_2, \ldots, x_n, \theta)$, tedy funkci pozorovaných dat a parametru $\theta$. Na tuto funkci se můžeme dívat dvojím způsobem:

- $g$ je funkce pozorovaných dat $x_1, x_2, \ldots, x_n$ při daném parametru $\theta$,
- $g$ je funkce parametru $\theta$ při pevně daných datech $x_1, x_2, \ldots, x_n$.

Nejprve se zaměříme na případ funkce pozorovaných dat $x_1, x_2, \ldots, x_n$. Pokud bude veličina $X$ diskrétně rozdělená, budeme každé možné realizaci náhodného výběru přiřazovat pravděpodobnost. Tyto pravděpodobnosti jednotlivých realizací se pak budou sčítat na 1 (viz [3], definice 2.4). V případě spojité proměnné $X$ budeme mluvit o funkci $g$ jako o hustotě. Integrací funkce $g$ přes všechny možné hodnoty $x_1, x_2, \ldots, x_n$ dostaneme opět hodnotu 1 (viz [3], věta 2.3). V tomto kontextu tedy mluvíme o pravděpodobnosti.

Nyní se zaměříme na případ, kdy je $g$ funkcí parametru $\theta$ při pevně daných hodnotách $x_1, x_2, \ldots, x_n$, což je právě námi uvažovaný případ věrohodnosti. Pokud budeme chtít integrovat funkci $g$ podle parametru $\theta$, jako jsme to udělali podobně pro $x_1, x_2, \ldots, x_n$ v předchozí situaci, nezískáme tentokrát hodnotu 1. Věrohodnostní funkce $g$ totiž není pravděpodobnostní hustota v $\theta$, protože nevyjadřuje pravděpodobnostní rozdělení parametru (tuto funkci zastává až aposterior, viz 3.2.3). Proto mluvíme o tom, jak je konkrétní hodnota parametru $\theta$ při daných datech věrohodná, a nikoliv pravděpodobná.

### 3.2.3. Aposteriorní rozdělení

Z Bayesovy věty (viz [3], věta 1.5) víme, že

$$p\left(\theta|\bar{X}_n\right) \propto p\left(\theta\right) p\left(\bar{X}_n|\theta\right),$$

kde $p\left(\theta\right)$ představuje apriorní hustotu, $p\left(\bar{X}_n|\theta\right)$ značí věrohodnost a $p\left(\theta|\bar{X}_n\right)$ představuje aposteriorní hustotu, která vyjadřuje hustotu parametru $\theta$ při zohlednění získaných dat. Vzhledem k apriorní představě tedy zohledníme nově naměřené hodnoty a tím ji upravíme do podoby aposteriorní představy. Hustotu aposteriorního rozdělení získáme díky Bayesově větě jako normovaný součin hustot apriorního rozdělení $N\left(\mu_{prior}, \frac{\sigma^2}{n_0}\right)$ a věrohodnostní funkce,

která odpovídá hustotě normálního rozdělení $N\left(\theta, \frac{\sigma^2}{n}\right)$. Pomocí analytických úprav tedy můžeme psát

$$\begin{aligned}
p\left(\theta|\bar{X}_n\right) &\propto p\left(\theta\right) p\left(\bar{X}_n|\theta\right) \\
&= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n_0}}} \exp\left(-\frac{\left(\theta-\mu_{prior}\right)^2}{2\frac{\sigma^2}{n_0}}\right) \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \exp\left(-\frac{\left(\bar{X}_n-\theta\right)^2}{2\frac{\sigma^2}{n}}\right) \\
&= \frac{1}{2\pi\frac{\sigma^2}{\sqrt{n_0 n}}} \exp\left(\frac{-\left(\theta-\mu_{prior}\right)^2}{2\frac{\sigma^2}{n_0}} + \frac{-\left(\bar{X}_n-\theta\right)^2}{2\frac{\sigma^2}{n}}\right) \\
&\propto \exp\left(\frac{-\left(\theta-\mu_{prior}\right)^2}{2\frac{\sigma^2}{n_0}} + \frac{-\left(\bar{X}_n-\theta\right)^2}{2\frac{\sigma^2}{n}}\right) \\
&\propto \exp\left(-\frac{1}{2}\frac{\left(\theta^2 - 2\theta\frac{n\bar{X}_n+n_0\mu_{prior}}{n+n_0}\right)}{\frac{\sigma^2}{n+n_0}}\right) \\
&= \exp\left(-\frac{1}{2}\frac{\left(\theta^2-2\theta\mu_{post}\right)}{\sigma^2_{post}}\right) \\
&\propto \exp\left(-\frac{1}{2}\frac{\left(\theta-\mu_{post}\right)^2}{\sigma^2_{post}}\right),
\end{aligned}$$

kde

$$\mu_{post} = \frac{n\bar{X}_n + n_0\mu_{prior}}{n+n_0}, \qquad \sigma^2_{post} = \frac{\sigma^2}{n+n_0}.$$

Tímto následně dostáváme aposteriorní rozdělení $\theta|\bar{X}_n \sim N\left(\mu_{post}, \sigma^2_{post}\right)$. Aposteriorní střední hodnota $\mu_{post}$, jak můžeme vidět výše, je váženým průměrem střední hodnoty $\mu_{prior}$ a průměrné hodnoty zjištěné z dat $\bar{X}_n$. Rozptyl aposteriorního rozdělení $\sigma^2_{post}$ je nepřímo úměrný celkovému počtu pozorování v aktuálních datech a v datech, ze kterých jsme vycházeli při volbě apriorního rozdělení. Oba parametry tedy zohledňují počty využitých dat a podle toho se aposteriorní hustota blíží spíše k apriorní hustotě, nebo k vě-

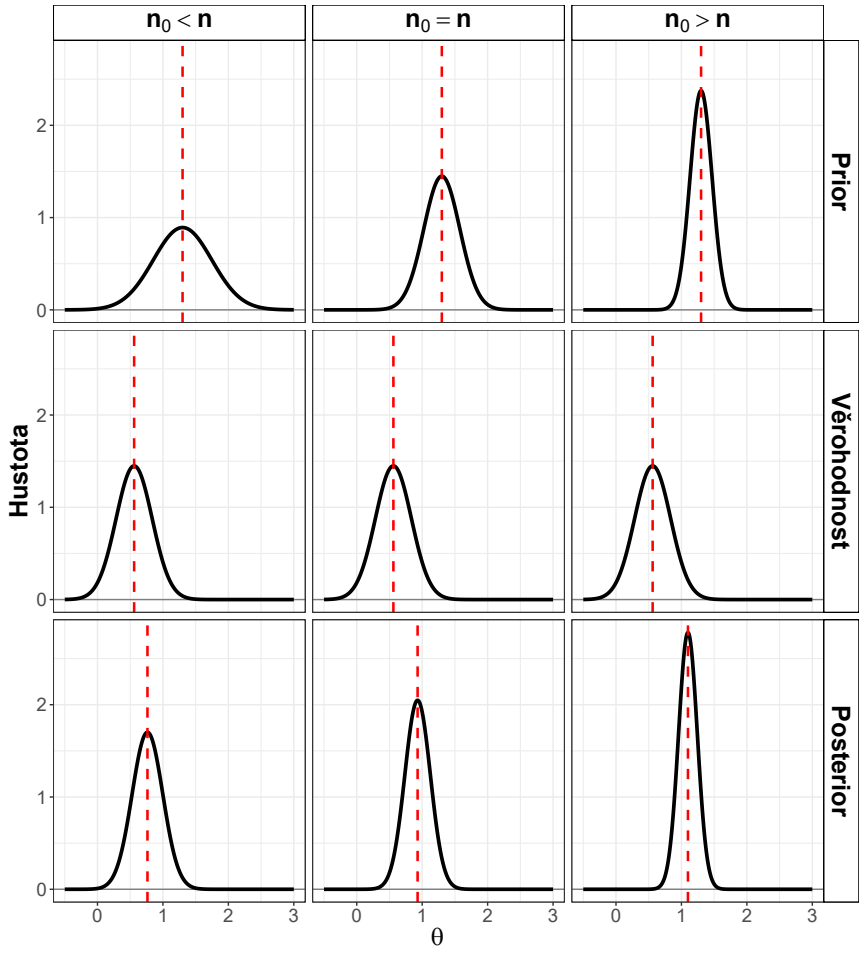
rohodnosti. Pokud bude například současných dat více než těch, která byla použita pro získání apriorního rozdělení, tedy $n_0 < n$, bude střední hodnota aposteriorního rozdělení blíže průměrné hodnotě aktuálních dat. Budou-li rozsahy dat $n$ a $n_0$ stejně velké, tedy $n_0 = n$, bude se střední hodnota aposteriorního rozdělení nacházet uprostřed mezi střední hodnotou apriorního rozdělení a průměrnou hodnotou aktuálních dat. Takto můžeme podobně uvažovat i další situace, viz obr. 2.

Máme-li tedy k dispozici apriorní rozdělení střední hodnoty zkoumaného normálního rozdělení i věrohodnostní funkci, jsme pomocí nich schopni dopočítat aposteriorní rozdělení parametru $\theta$, které vyjadřuje naši představu o možných hodnotách tohoto parametru poté, co jsme zohlednili aktuální data.

## 3.3. Bayesovská inference v obecnější situaci

V kapitole 3.2 jsme vycházeli z toho, že parametr $\theta$ interpretujeme jako střední hodnotu normálního rozdělení. Nyní uvažujme obecně parametr $\theta$, který je funkcí parametrů rozdělení, z něhož (či z nichž) máme náhodný výběr. Příklady takových funkcí uvádíme níže v této kapitole.

Na tuto funkci budeme klást několik požadavků. Prvním z nich je, aby byl parametr $\theta$ nestranně odhadnutelný, stejně jako tomu bylo v předchozí kapitole, tj. existuje nestranný odhad parametru $\theta$. Tento odhad budeme značit $T$. Platí tedy $\mathrm{E}\,(T|\theta) = \theta$. Dále budeme požadovat, aby měl tento odhad normální rozdělení. Rozptyl tohoto rozdělení bude navíc nepřímo úměrný velikosti rozsahu výběru, obecně tedy $\frac{\sigma^2}{n}$, kde $\sigma^2$ reprezentuje nějakou známou kladnou konstantu. Apriorní rozdělení budeme opět předpokládat ve tvaru $\theta \sim N\left(\mu_{prior}, \frac{\sigma^2}{n_0}\right)$. Věrohodnost bude určovat rozdělení splňující výše uvedená kritéria, tedy $T_n|\theta \sim N\left(\theta, \frac{\sigma^2}{n}\right)$, kde $T_n$ bude znamenat nestranný




Obrázek 2: Graf znázorňující apriorní rozdělení $\theta \sim N\left(\mu_{prior}, \frac{\sigma^2}{n_0}\right)$, věrohodnost a jejich kombinací vzniklé aposteriorní rozdělení v závislosti na velikostech vzorků pro $\sigma^2 = 2,2$. V levém sloupci se nachází apriorní rozdělení $N(1,3; 0,2)$ s rozptylem odpovídajícím vzorku o velikosti $n_0 = 11$. Uprostřed se nachází apriorní rozdělení $N(1,3; 0,08)$ s rozptylem odpovídajícím vzorku o velikosti $n_0 = 29$. V pravém sloupci se nachází apriorní rozdělení $N(1,3; 0,03)$ s rozptylem odpovídajícím vzorku o velikosti $n_0 = 78$. Věrohodnost je ve všech třech případech z rozdělení $N(0,56; 0,08)$ se vzorkem o velikosti $n = 29$, tedy shodně s druhým priorem. V levém sloupci vzniká aposteriorní rozdělení $N(0,76; 0,06)$, uprostřed $N(0,93; 0,04)$ a vpravo $N(1,1; 0,02)$. Červenou přerušovanou čarou je znázorněna střední hodnota uvedených rozdělení.

odhad parametru $\theta$ při velikosti rozsahu $n$. Kombinací apriorního rozdělení a věrohodnosti můžeme obdobně jako v odvození v kapitole 3.2.3 získat normálně rozdělené aposteriorní rozdělení.

Tuto situaci uvažuje v *Bayesian approaches to clinical trials and health-care evaluation* Spiegelhalter (viz strana 22 v [4]).

Předpokládejme dva náhodné výběry. První náhodný výběr $X_1, X_2, \ldots, X_n$ bude z rozdělení $X \sim N(\mu_1, \sigma^2)$ a druhý $Y_1, Y_2, \ldots, Y_n$ bude z rozdělení $Y \sim N(\mu_2, \sigma^2)$. Dále si označíme výběrové průměry $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ a druhý $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. V následujícím odstavci si uvedeme funkce, jež splňují výše uvedená kritéria a jimiž můžeme interpretovat odhadovaný parametr $\theta$:

- jeden náhodný výběr $X_1, X_2, \ldots, X_n$
  - $\theta$ je přímo střední hodnotou náhodné veličiny $X$, tedy $\theta = \mu_1$. Funkce výběru bude mít tvar $T_n = \bar{X}$ (viz 3.2).
- dva náhodné výběry $X_1, X_2, \ldots, X_n$ a $Y_1, Y_2, \ldots, Y_n$
  - $\theta$ je rozdíl středních hodnot $X$ a $Y$, tedy $\theta = \mu_1 - \mu_2$. Funkce bude mít tvar $T_n = \bar{X} - \bar{Y}$.
  - $\theta$ je normovaný rozdíl středních hodnot veličin $X$ a $Y$, přičemž normující konstantou je směrodatná odchylka $\sigma$. Tuto charakteristiku $\theta = \frac{\mu_1 - \mu_2}{\sigma}$ nazýváme Cohenovo delta a můžeme ji využít při popisu míry efektu nějaké intervence. Funkce výběru bude ve tvaru $T_n = \frac{\bar{X} - \bar{Y}}{\sigma}$. Tento případ rozvedeme více v kapitole 3.4.
  - uvažovat můžeme i situaci, kde budou mít rozdělení různé rozptyly, tedy $X \sim N(\mu_1, \sigma_1^2)$ a $Y \sim N(\mu_2, \sigma_2^2)$. Parametr $\theta$ je v tomto případě v podobě $\theta = \frac{\mu_1}{\sigma_1} - \frac{\mu_2}{\sigma_2}$ a funkce výběru $T_n = \frac{\bar{X}}{\sigma_1} - \frac{\bar{Y}}{\sigma_2}$. V případě rovnosti rozptylů dostáváme opět Cohenovo delta.

## 3.4. Bayesovská inference pro odhad velikosti efektu na střední hodnotu normálního rozdělení při známém rozptylu

Cílem této části třetí kapitoly je přiblížit situaci, ve které interpretujeme parametr $\theta$ jako Cohenovo delta. Pro tento případ si představíme ilustrativní studii, na které si dále v práci prakticky ukážeme různé možnosti zapojení apriorní představy.

Zaměřme se na klinickou studii, jejímž smyslem je zjistit, zda je nový lék pro pacienty trpící vysokým krevním tlakem obecně prospěšný. Uvažovat budeme dva výběry, tedy dvě různé skupiny pacientů na sobě nezávislé. Pacientům ve skupině A budeme během doby léčby podávat placebo a pacientům ve skupině B nový lék. Podobně, pouze s jinou závěrečnou interpretací, bychom mohli uvažovat situaci, kdy bychom porovnávali dva různé léky, případně různé dávky jednoho léku. Obě skupiny budou obsahovat stejný počet pacientů $n$. U každého pacienta změříme tlak před začátkem léčby a po ukončení léčby a spočítáme jejich rozdíl, se kterým budeme dále pracovat. Pro skupinu A budeme předpokládat náhodný výběr $X_1, X_2, \ldots, X_n$, kde náhodná veličina $X_i$ představuje změnu krevního tlaku u $i$-tého (náhodně vybraného) pacienta. Náhodné veličiny budou nezávislé, stejně rozdělené (i.i.d.), pocházející z normálního rozdělení se střední hodnotou $\mu_1$ a známým rozptylem $\sigma_*^2$. V našem fiktivním příkladu může být např. $\sigma_* = 5$. Pro skupinu B budeme předpokládat náhodný výběr $Y_1, Y_2, \ldots, Y_n$, kde náhodná veličina $Y_i$ představuje změnu krevního tlaku u $i$-tého (náhodně vybraného) pacienta. Náhodné veličiny budou i.i.d pocházející taktéž z normálního rozdělení, tentokrát se střední hodnotou $\mu_2$ a stejným rozptylem $\sigma_*^2$ jako ve skupině A.

Míru efektu nového léku oproti placebu, kterou budeme chtít odhadovat,

budeme měřit pomocí Cohenova delta

$$\theta_d = \frac{\mu_1 - \mu_2}{\sigma_*}.$$

Testovat budeme to, zda je lék obecně účinný, nebo ne, tedy $H_0 \colon \theta_d \leq 0$ vůči alternativě $H_A \colon \theta_d > 0$. V této práci tedy budeme pracovat s jednostrannou alternativou. Všechny uvedené metody se však dají použít i v případě alternativy oboustranné, např. $H_0 \colon \theta_d = 0$ a $H_A \colon \theta_d \neq 0$. Zde bychom zamítnutím nulové hypotézy pouze zjistili, že efekt není 0, ale nevěděli bychom nic o tom, zda je lék účinný, anebo nebezpečný, a proto se na použití této alternativy nebudeme soustředit. Hladinu významnosti zvolíme $\alpha = 0,05$. Pokud tedy nulová hypotéza platí, připouštíme 5% pravděpodobnost, že bude mylně zamítnuta.

Pro agregaci naměřených dat využijeme testovou statistiku $Z$ dvouvýběrového $z$-testu. $Z$-test porovnává průměry dvou nezávislých náhodných výběrů za předpokladu normality, nezávislosti a stejných, známých rozptylů. Výběrový průměr skupiny A označíme jako $\bar{X} \sim N(\mu_1, \frac{\sigma_*^2}{n})$ a skupiny B jako $\bar{Y} \sim N(\mu_2, \frac{\sigma_*^2}{n})$. Rozdělení rozdílu výběrových průměrů získáme opět stejným postupem, jako jsme uváděli v kapitole 3.2.1. Testovou statistiku vyjádříme následovně

$$Z = \frac{\left(\bar{X} - \bar{Y}\right) - (\mu_1 - \mu_2)}{\sqrt{\operatorname{var}\left(\bar{X} - \bar{Y}\right)}} \overset{H_0}{=} \frac{\bar{X} - \bar{Y}}{\sqrt{\operatorname{var}\left(\bar{X}\right) + \operatorname{var}\left(\bar{Y}\right)}} = \frac{\bar{X} - \bar{Y}}{\sigma_*}\sqrt{\frac{n}{2}}$$

Členem $T_n = \frac{\bar{X} - \bar{Y}}{\sigma_*}$ rozumíme odhad míry efektu, tedy výběrovou funkci. Dvouvýběrový $Z$-test budeme následně využívat při výpočtu velikosti síly.

Z již dříve proběhlé studie máme k dispozici apriorní představu o míře efektu

nového léku $p(\theta_d)$, kde $\theta_d \sim N(\theta_0, \sigma_0^2)$, $\theta_0$ je střední hodnota a $\sigma_0^2$ rozptyl parametru $\theta_d$, alternativně bychom mohli psát $\sigma_0^2 = \frac{\sigma^2}{n_0}$. Věrohodnost vychází z rozdělení statistiky $\bar{X} - \bar{Y}$:

$$\left(\bar{X} - \bar{Y}\right) \sim N\left(\mu_1 - \mu_2, \frac{2\sigma_*^2}{n}\right).$$

Po vydělení směrodatnou odchylkou $\sigma_*$ dostáváme

$$\frac{\bar{X} - \bar{Y}}{\sigma_*} = T_n \sim N\left(\frac{\mu_1 - \mu_2}{\sigma_*}, \frac{2}{n}\right) = N\left(\theta_d, \frac{2}{n}\right).$$

V obecném případě (viz kapitola 3.3) jsme požadovali, aby měl rozptyl výběrové funkce tvar $\frac{\sigma^2}{n}$. Můžeme tedy vidět, že pro případ Cohenova delta je konstanta $\sigma^2$ rovna 2.

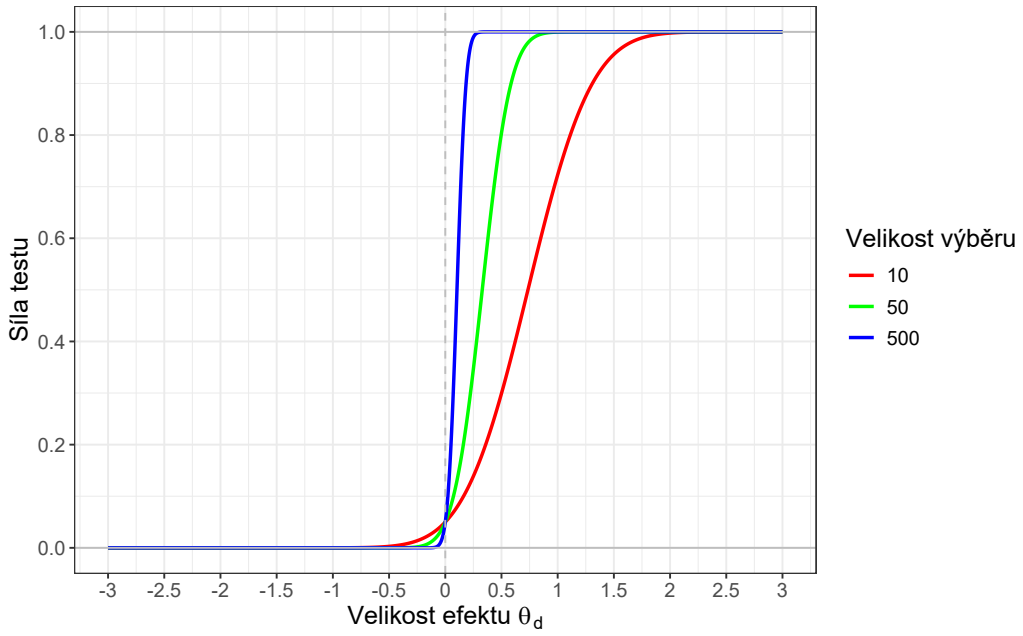
Díky tomu, že umíme odvodit rozdělení určující věrohodnost, jsme opět schopni z Bayesovy věty (viz [3], věta 1.5) získat aposteriorní rozdělení. Odvození bude ekvivalentní postupu uvedeném v kapitole 3.2.3, proto si zde uvedeme pouze výsledné rozdělení, které bude ve tvaru $\theta_d|T_n \sim N\left(\mu_{post}, \sigma_{post}^2\right)$, kde

$$\mu_{post} = \frac{\theta_0 + \frac{n}{2}\sigma_0^2 T_n}{1 + \frac{n}{2}\sigma_0^2}, \qquad \sigma_{post}^2 = \frac{\sigma_0^2}{1 + \frac{n}{2}\sigma_0^2}.$$

Po nahrazení $\sigma_0^2$ alternativním označením $\frac{\sigma^2}{n_0}$ můžeme přepsat střední hodnotu $\mu_{post}$ jako vážený průměr střední hodnoty $\theta_0$ a normovaného rozdílu výběrových průměrů mezi kontrolní a testovací skupinou $T_n$. Rozptyl $\sigma_{post}^2$ bude nepřímo úměrný celkovému počtu pozorování v aktuálních datech a v datech, ze kterých jsme vycházeli při volbě apriorního rozdělení:

$$\mu_{post} = \frac{n_0}{n_0 + n}\theta_0 + \frac{n}{n_0 + n}T_n, \qquad \sigma^2_{post} = \frac{2}{n_0 + n}.$$

Nyní si v naší ukázkové studii rozdělíme situaci na tři různé případy, které se budou lišit volbou rozsahu výběru. Půjde nám o to ukázat, jaký vliv má vhodné zvolení velikosti výběru na odhalení případného efektu. V prvním případě zvolíme počet pacientů na skupinu 10, ve druhém 50 a ve třetím 500. Pro všechny tyto rozsahy si jako první vykreslíme odpovídající silofunkce, zatím bez zohlednění apriorní představy. Silofunkcí rozumíme funkci, která každé hodnotě efektu $\theta_d$ přiřadí sílu testu. Pro každou hodnotu efektu tedy vizuálně uvidíme, s jakou pravděpodobností test tento případný efekt odhalí. Průběh silofunkcí pro zmíněné tři rozsahy je znázorněn v grafu na obrázku 3.

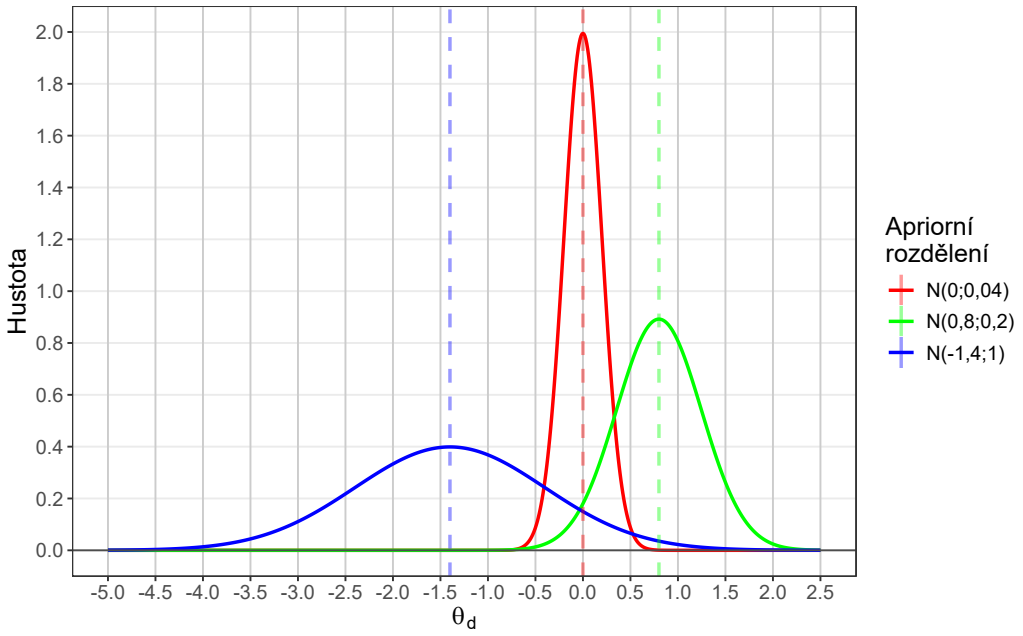



Obrázek 3: Graf silofunkcí $z$-testu při jednostranné alternativě pro rozsahy výběru 10 (červeně), 50 (zeleně) a 500 (modře).

Na grafu můžeme například pozorovat, že pokud by byla velikost skutečného efektu rovna 0,5, se vzorkem velikosti 10 bychom tento efekt odhalili pouze s pravděpodobností 0,3. Tento efekt by tedy s velkou pravděpodobností odhalen nebyl. Velký vzorek, čítající v každé z obou skupin 500 pacientů, by tento efekt odhalil s pravděpodobností rovnou skoro 1. Většinou požadujeme, aby se síla testu pohybovala kolem 80 %, velký vzorek by tedy efekt odhaloval se zbytečně velkou pravděpodobností. Z námi zvolených rozsahů se zdá nejvhodnější rozsah velikosti 50, který případný efekt o velikosti 0,3 odhalí s pravděpodobností 0,8. Podobně můžeme najít v grafu interpretaci pro každé $\theta_d$.

Na závěr této kapitoly si představíme tři apriorní rozdělení parametru $\theta_d$ pocházející z normálního rozdělení, která budeme dále využívat při návrhu vhodného rozsahu výběru a designu studie. Každé z nich bude demonstrovat jiný efekt a jinak velký rozptyl, viz obr. 4:

- $\theta_{d_1} \sim N\left(\mu_{prior_1}, \sigma^2_{prior_1}\right)$, kde $\mu_{prior_1} = 0$ a $\sigma^2_{prior_1} = 0,04$
  – favorizace nulového efektu léku
- $\theta_{d_2} \sim N\left(\mu_{prior_2}, \sigma^2_{prior_2}\right)$, kde $\mu_{prior_2} = 0,8$ a $\sigma^2_{prior_2} = 0,2$
  – favorizace pozitivního efektu léku
- $\theta_{d_3} \sim N\left(\mu_{prior_3}, \sigma^2_{prior_3}\right)$, kde $\mu_{prior_3} = -1,4$ a $\sigma^2_{prior_3} = 1$
  – favorizace negativního efektu léku




Obrázek 4: Graf apriorních rozdělení velikosti efektu $\theta_d$ – červeně $N\left(0; 0,04\right)$, zeleně $N\left(0,8; 0,2\right)$ a modře $N\left(-1,4; 1\right)$.

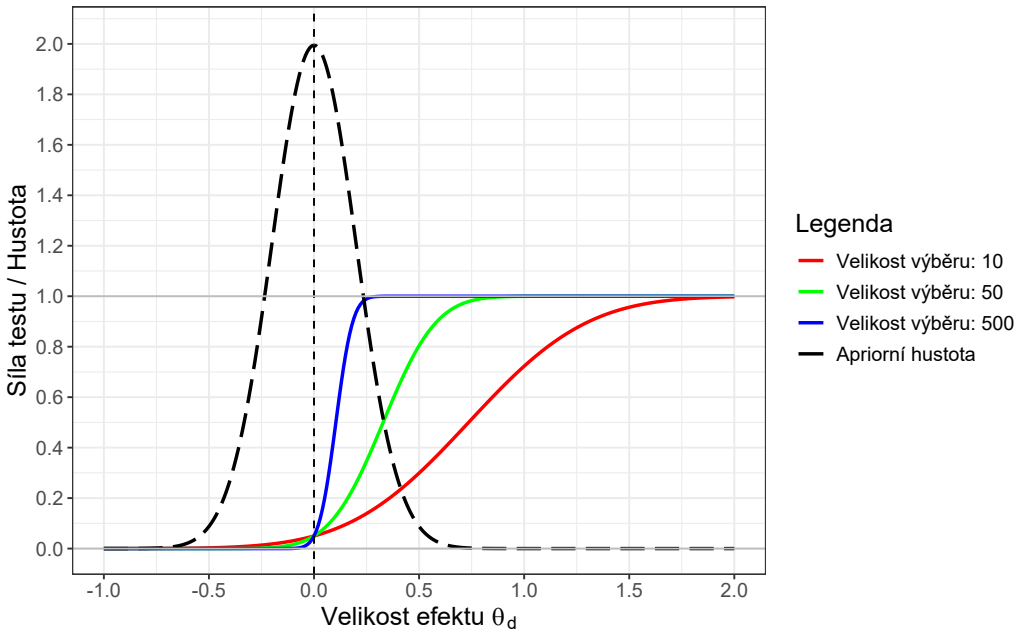
# 4. Metody pro zohlednění apriorní představy při stanovení rozsahu výběru v klinické studii

V této části práce představíme metody, které zohledňují dostupné apriorní představy pro vhodnější a realističtější návrh studie. Každý způsob budeme demonstrovat na klinické studii uvedené v předchozí kapitole. Vycházet budeme z metod popsaných v knize *Bayesian approaches to clinical trials and health-care evaluation* (viz [4]).

## 4.1. Vizualizace silofunkce a apriorního rozdělení

Jednou z nejjednodušších možností, jak zohlednit apriorní představu, je přes sebe vizualizovat silofunkci odpovídající zvolené velikosti výběru a hustotu apriorního rozdělení. Následně můžeme neformálně posoudit, jak velký vzorek by mohl být pro klinickou studii nejvhodnější s ohledem na to, jak velkou máme sílu testu pro tyto různé rozsahy výběru v závislosti na různě pravděpodobných velikostech efektu $\theta$ podle apriorní představy. V ideálním případě požadujeme mít dostatečně velkou sílu testu pro detekci efektu velikosti nejpravděpodobnější hodnoty apriorního rozdělení, tedy apriorní střední hodnoty. Pro námi zavedené rozsahy výběru a apriorní rozdělení by vypadala situace následovně:

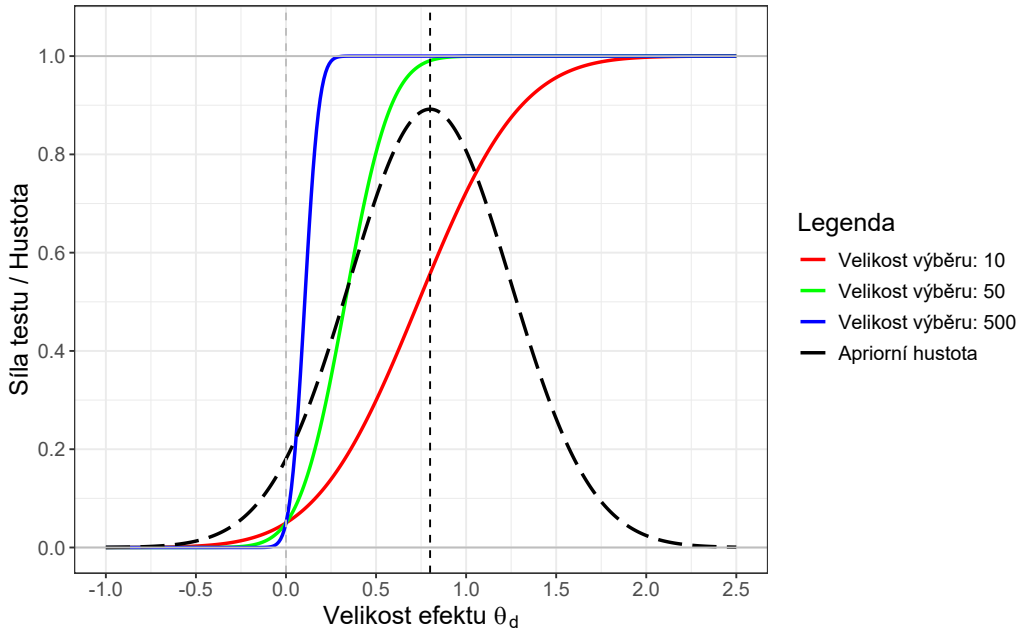
- apriorní rozdělení $N\,(0; 0,04)$, viz obr. 5




Obrázek 5: Graf hustoty apriorního rozdělení $N\,(0; 0,04)$ a silofunkcí $z$-testu při jednostranné alternativě pro rozsahy výběru: červeně – 10, zeleně – 50 a modře – 500.

V případě tohoto apriorního rozdělení, které má střední hodnotu v 0, můžeme vzhledem k námi zvolené nulové hypotéze dosáhnout v bodě maxima apriorní hustoty pro všechny velikosti výběru síly testu rovné nejvýše hladině významnosti. Pozorovat proto budeme hodnoty efektu blízké střední hodnotě. Například pro efekt velikosti 0,06 dosahuje test nejvyšší síly při rozsahu výběru 500, konkrétně okolo 24 %. Tato síla však není dostačující a v praxi by to tedy znamenalo navýšení počtu pacientů pro realističtější návrh studie. V opačném případě by nemuselo k detekci případně velmi malého efektu dojít. Na tomto příkladu ilustrujeme, že nemá smysl požadovat velkou „sílu“ v bodě maxima apriorní hustoty, je-li tento bod v množině hodnot obsažených v nulové hypotéze.

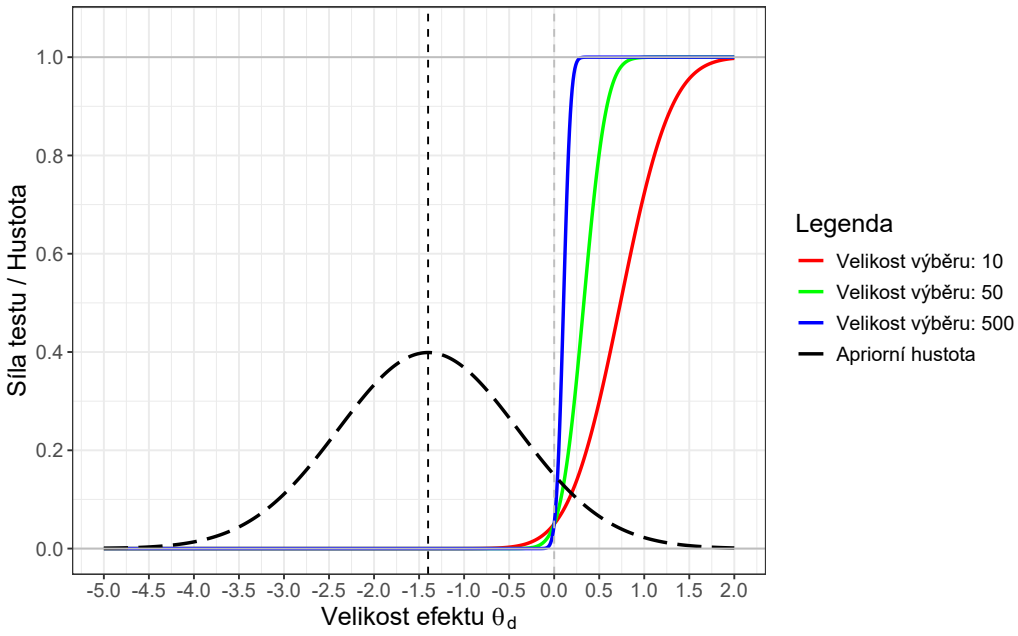
- apriorní rozdělení $N\,(0,8;0,2)$, viz obr. 6




Obrázek 6: Graf hustoty apriorního rozdělení $N\,(0,8;0,2)$ a silofunkcí $z$-testu při jednostranné alternativě pro rozsahy výběru: červeně – 10, zeleně – 50 a modře – 500.

V této situaci by byl rozsah o velikosti 500 už zbytečně příliš velký, apriorní efekt už není tak malý, a tedy není potřeba tak velké množství pacientů. Dokonce ani rozsah výběru čítající 50 pacientů není zapotřebí. Menší rozsah čítající 10 pacientů je však už příliš malý; pro efekt 0,8 má sílu testu pouze okolo 56 %. Nejvhodnější bude tedy volit velikost rozsahu výběru mezi 10 a 50.

- apriorní rozdělení $N\,(-1,4;1)$, viz obr. 7




Obrázek 7: Graf hustoty apriorního rozdělení $N\,(-1,4;1)$ a silofunkcí $z$-testu při jednostranné alternativě pro rozsahy výběru: červeně – 10, zeleně – 50 a modře – 500.

V tomto případě apriorně očekáváme záporný efekt, ovšem alternativní hypotéza odpovídá efektu kladnému. Na grafu můžeme vidět, že i když bychom dál zvětšovali počet pacientů, k zamítnutí nulové hypotézy by pravděpodobně nedošlo. Dosáhnout můžeme pouze síly testu rovné 5 %, tedy nejvýše námi zvolené hladiny významnosti. Zde přichází v úvahu přehodnocení volby konstrukce hypotéz a studie obecně.

## 4.2. Hodnota $\mu_{prior}$ jako bodová alternativa

V případě, kdy má apriorní rozdělení malý rozptyl a hodnota $\theta$ je tedy pravděpodobně blízko $\mu_{prior}$, můžeme tuto střední hodnotu využít jako bodovou

alternativu. Místo toho, abychom použili $H_0 \colon \theta \leq 0$ a $H_A \colon \theta > 0$, přepíšeme hypotézy na: $H_0 \colon \theta \leq 0$ a $H_A \colon \theta = \mu_{prior}$.

Malý rozptyl vyžadujeme především proto, že vybráním pouze jednoho bodu, tedy střední hodnoty, nezohledníme nejistotu ohledně tohoto bodu a ostatních možných efektů. Ochuzujeme se tedy o podstatnou informaci, což však v případě malého rozptylu nečiní tak velký problém. Tato možnost zapojení apriorní představy se využívá především v případě, kdy chceme rychlý a jednoduchý odhad.

V námi zmíněných příkladech by se mohla zdát tato metoda vhodná pro apriorní rozdělení s malým rozptylem 0,04. V tomto případě je ale střední hodnota apriorního rozdělení zahrnuta již v intervalu určujícím nulovou hypotézu, a proto tuto metodu nelze použít. Mohli bychom ji využít například v situaci, kdy by byla střední hodnota apriorního rozdělení rovna 1,3, tedy $N\left(1,3; 0,04\right)$. Nevhodná by naopak mohla být pro situaci se stejnou střední hodnotou 1,3 a rozptylem o velikosti 1.

## 4.3. Průměrná síla testu

Jako další ukazatel toho, zda jsme zvolili vhodnou velikost výběru, můžeme využít průměrnou sílu testu. Metodu si představíme nejen v kontextu hybridní statistiky, které se v této práci věnujeme primárně, ale také následně v případě čistě bayesovské statistiky. Zůstávat budeme stále v kontextu hypotéz $H_0 \colon \theta \leq 0$ a $H_A \colon \theta > 0$. Vycházet budeme z veličiny reprezentující nově naměřená data $T_n|\theta \sim N\left(\theta, \frac{\sigma^2}{n}\right)$, kterou jsme definovali v kapitole 3.3. Apriorní rozdělení budeme taktéž předpokládat stejné jako v uvedené kapitole, tedy $\theta \sim N\left(\mu_{prior}, \frac{\sigma^2}{n_0}\right)$.

### 4.3.1. Klasická síla testu

Síle testu jsme se věnovali již v kapitole 2, kde jsme vysvětlili, že se jedná o pravděpodobnost zamítnutí nulové hypotézy při její neplatnosti. Jak je blíže popsáno v knize *Bayesian approaches to clinical trials and health-care evaluation* v kapitole 2.5 (viz [4]), nulovou hypotézu budeme zamítat v případě, kdy $T_n > -z_\alpha \frac{\sigma}{\sqrt{n}}$, kde $z_\alpha$ značí $\alpha$ kvantil normálního rozdělení. Tento jev označíme jako $S_\alpha^C$. Jev $S_\alpha^C$ nastane s následující pravděpodobností:

$$\begin{aligned} p\left(S_\alpha^C|\theta\right) &= p\left(\frac{T_n-\theta}{\sigma}\sqrt{n} > \frac{-z_\alpha\frac{\sigma}{\sqrt{n}}-\theta}{\sigma}\sqrt{n}\right) \\ &= 1 - p\left(\frac{T_n-\theta}{\sigma}\sqrt{n} < \frac{-z_\alpha\frac{\sigma}{\sqrt{n}}-\theta}{\sigma}\sqrt{n}\right) \\ &= 1 - \Phi\left(-z_\alpha - \frac{\theta\sqrt{n}}{\sigma}\right) \\ &= \Phi\left(z_\alpha + \frac{\theta\sqrt{n}}{\sigma}\right), \end{aligned} \tag{2}$$

kde $\Phi$ značí distribuční funkci normovaného normálního rozdělení. Tímto tedy získáváme klasickou sílu testu $p\left(S_\alpha^C|\theta\right) = \Phi\left(z_\alpha + \frac{\theta\sqrt{n}}{\sigma}\right) = 1-\beta$ pro hybridní přístup.

### 4.3.2. Průměrná klasická síla testu

Nyní budeme chtít získat vztah pro průměrnou sílu testu. Průměrná síla testu představuje vážený průměr pravděpodobnosti $p\left(S_\alpha^C|\theta\right)$ přes všechny možné hodnoty parametru $\theta$, přičemž váhy odpovídají jeho apriornímu rozdělení. Pokud je průměrná síla testu vysoká, znamená to, že test má s ohledem na apriorně očekávané hodnoty $\theta$ dobré schopnosti detekovat případný efekt. V opačném případě by to znamenalo, že má test malou schopnost zamítnout nesprávnou nulovou hypotézu, což by vedlo ke zvýšení velikosti výběru nebo

přehodnocení designu studie.

K získání průměrné síly můžeme přistupovat dvěma způsoby. První možností je integrace klasické síly $p\left(S_\alpha^C|\theta\right)$ s ohledem na apriorní rozdělení, tedy

$$p\left(S_\alpha^C\right) = \int_{-\infty}^{\infty} p\left(S_\alpha^C|\theta\right) p\left(\theta\right) d\theta, \tag{3}$$

kde $p\left(S_\alpha^C\right)$ značí průměrnou sílu testu. Výpočet tohoto integrálu může být mnohdy náročný, proto se častěji používá druhý způsob. Hlavní princip spočívá ve využití rozdělení veličiny $T_n \sim N\left(\mu_{prior}, \sigma^2\left(\frac{1}{n}+\frac{1}{n_0}\right)\right)$. Toto rozdělení získáme za pomoci vlastností podmíněné distribuce, viz [2] rovnice (1.8) a (1.9):

$$\begin{aligned}
\mathrm{E}\left(T_n\right) &= \mathrm{E}\left(\mathrm{E}\left(T_n|\theta\right)\right) = \mathrm{E}\left(\theta\right) = \mu_{prior}, \\
\mathrm{var}\left(T_n\right) &= \mathrm{E}\left(\mathrm{var}\left(T_n|\theta\right)\right) + \mathrm{var}\left(\mathrm{E}\left(T_n|\theta\right)\right) = \frac{\sigma^2}{n} + \frac{\sigma^2}{n_0}.
\end{aligned}$$

Normalitu rozdělení $T_n$ dokládá Gelman v knize [2], *Appendix A* část *Univariate normal*. Opět budeme vycházet z pravděpodobnosti $p\left(S_\alpha^C|\theta\right)$, tentokrát ale nebudeme provádět normalizaci podle veličiny $T_n|\theta$, ale podle veličiny $T_n$. Díky tomu nebude síla testu na $\theta$ již záviset:

$$\begin{aligned}
p\left(S_\alpha^C\right) &= p\left(\frac{T_n-\mu_{prior}}{\sigma\sqrt{\frac{1}{n_0}+\frac{1}{n}}} > \frac{-z_\alpha\frac{\sigma}{\sqrt{n}}-\mu_{prior}}{\sigma\sqrt{\frac{1}{n_0}+\frac{1}{n}}}\right) \\
&= 1 - p\left(\frac{T_n-\mu_{prior}}{\sigma\sqrt{\frac{1}{n_0}+\frac{1}{n}}} < \sqrt{\frac{n_0}{n+n_0}}\left(-z_\alpha - \frac{\sqrt{n}\mu_{prior}}{\sigma}\right)\right) \\
&= 1 - \Phi\left(\sqrt{\frac{n_0}{n+n_0}}\left(-z_\alpha - \frac{\sqrt{n}\mu_{prior}}{\sigma}\right)\right) \\
&= \Phi\left(\sqrt{\frac{n_0}{n+n_0}}\left(z_\alpha + \frac{\sqrt{n}\mu_{prior}}{\sigma}\right)\right).
\end{aligned}$$

Průměrná síla testu se od té klasické liší v zásadě členem $\sqrt{\frac{n_0}{n+n_0}}$. Pokud by se velikost vzorku apriorního rozdělení $n_0$ blížila nekonečnu, byla by nejistota kolem střední hodnoty $\mu_{prior}$ velmi malá, což by mělo za následek to, že by se průměrná síla testu blížila klasické síle testu počítané pro $\mu_{prior}$, jak popisuje kniha [4] v kapitole 6.5.2.

### 4.3.3. Průměrná klasická síla testu podmíněná alternativní hypotézou

Můžeme také počítat průměrnou sílu testu podmíněnou platností alternativní hypotézy. Jedná se o pravděpodobnost správného zjištění pozitivního efektu. Na rozdíl od nepodmíněné průměrné síly testu zkoumá sílu jen v bodech odpovídajících alternativní hypotéze. Stejně jako klasická silofunkce však nezohledňuje věrohodnost alternativní hypotézy. Získat ji můžeme následujícím vztahem:

$$\begin{aligned} p\left(S_\alpha^C|\theta>0\right) &= \int_0^\infty p\left(S_\alpha^C|\theta\right) p\left(\theta|\theta>0\right) d\theta \\ &= \int_0^\infty p\left(S_\alpha^C|\theta\right) \frac{p\left(\theta\right)}{\int_0^\infty p\left(t\right) dt} d\theta, \end{aligned} \tag{4}$$

kde $p\left(S_\alpha^C|\theta>0\right)$ značí průměrnou sílu testu podmíněnou platností alternativní hypotézy.

Nyní můžeme porovnat rovnice pro průměrnou klasickou sílu testu (3) a pro průměrnou sílu testu podmíněnou platností alternativní hypotézy (4). Rovnice se liší množinou, přes kterou je integrováno. V případě (3) integrujeme přes celé $\mathbb{R}$, kdežto v situaci (4), kde podmiňujeme $S_\alpha^C$ alternativní hypotézou $\theta > 0$, integrujeme pouze přes $\mathbb{R}^+$. Dále se vzorce liší v použité normovací konstantě. Pro (3) je normovací konstantou přímo hodnota 1, jelikož sílu

testu integrujeme přes celé apriorní rozdělení. Pro (4) je normovací konstantou $p(\theta > 0)$, jelikož sílu testu průměrujeme pouze přes hodnoty v množině alternativní hypotézy. V případě, kdy budeme mít velký vzorek výběru, a silofunkce tedy bude nabývat velmi rychle hodnoty rovné skoro 1, bude platit vztah $p\left(S_\alpha^C|\theta > 0\right) \approx \frac{p\left(S_\alpha^C\right)}{p(\theta>0)}$.

Na naší studii prakticky ukážeme, jak se od sebe budou průměrná síla testu a průměrná síla testu podmíněná platností alternativní hypotézy lišit pro různé velikosti výběru a různá apriorní rozdělení:

- apriorní rozdělení $\theta_d \sim N(0; 0,04)$, viz obr. 8

Jako první máme k dispozici apriorní rozdělení, jehož střední hodnota je rovna hodnotě, která od sebe odděluje nulovou a alternativní hypotézu. S rostoucí velikostí výběru rostou i obě průměrné síly testu, každá však jinak rychle. V případě velikosti vzorku 10 je průměrná síla rovna 0,07 a podmíněná průměrná síla rovna 0,11. V situaci s velikostí rozsahu 500 má průměrná síla testu velikost 0,31 a podmíněná průměrná síla má velikost 0,61. Zde můžeme názorně vidět vztah $p\left(S_\alpha^C|\theta > 0\right) \approx \frac{p\left(S_\alpha^C\right)}{p(\theta>0)}$ popsaný výše, kde je v tomto případě normující konstantou hodnota 0,5, jinak řečeno je podmíněná síla testu zhruba dvakrát větší než síla testu nepodmíněná, jelikož je hustota apriorního rozdělení symetrická kolem 0. Pokud můžeme předpokládat platnost alternativní hypotézy, jsme schopni pro počet 2 200 pacientů na skupinu dosáhnout podmíněné průměrné síly testu 80 %. Tento počet pacientů jsme pro účely srovnání s nepodmíněnou průměrnou silou získali experimentálně tak, že jsme postupně navyšovali počet pacientů při výpočtu podmíněné průměrné síly až do doby, než jsme dosáhli požadované síly 80 %. Jak jsme již zmínili výše, s rostoucí velikostí výběru poroste i nepodmíněná průměrná síla, ovšem jen do určitého momentu. I kdyby byla síla pro libovolnou hodnotu alternativy

rovna zhruba 1, byla by nepodmíněná průměrná síla přibližně rovna 0,5, protože pro záporné hodnoty, jejichž celková apriorní pravděpodobnost je 0,5, je síla testu menší než hladina významnosti. V našem příkladu tedy nejsme pro jakýkoliv počet pacientů schopni překročit hodnotu 50 %.

- apriorní rozdělení $\theta_d \sim N(0,8; 0,2)$, viz obr. 9

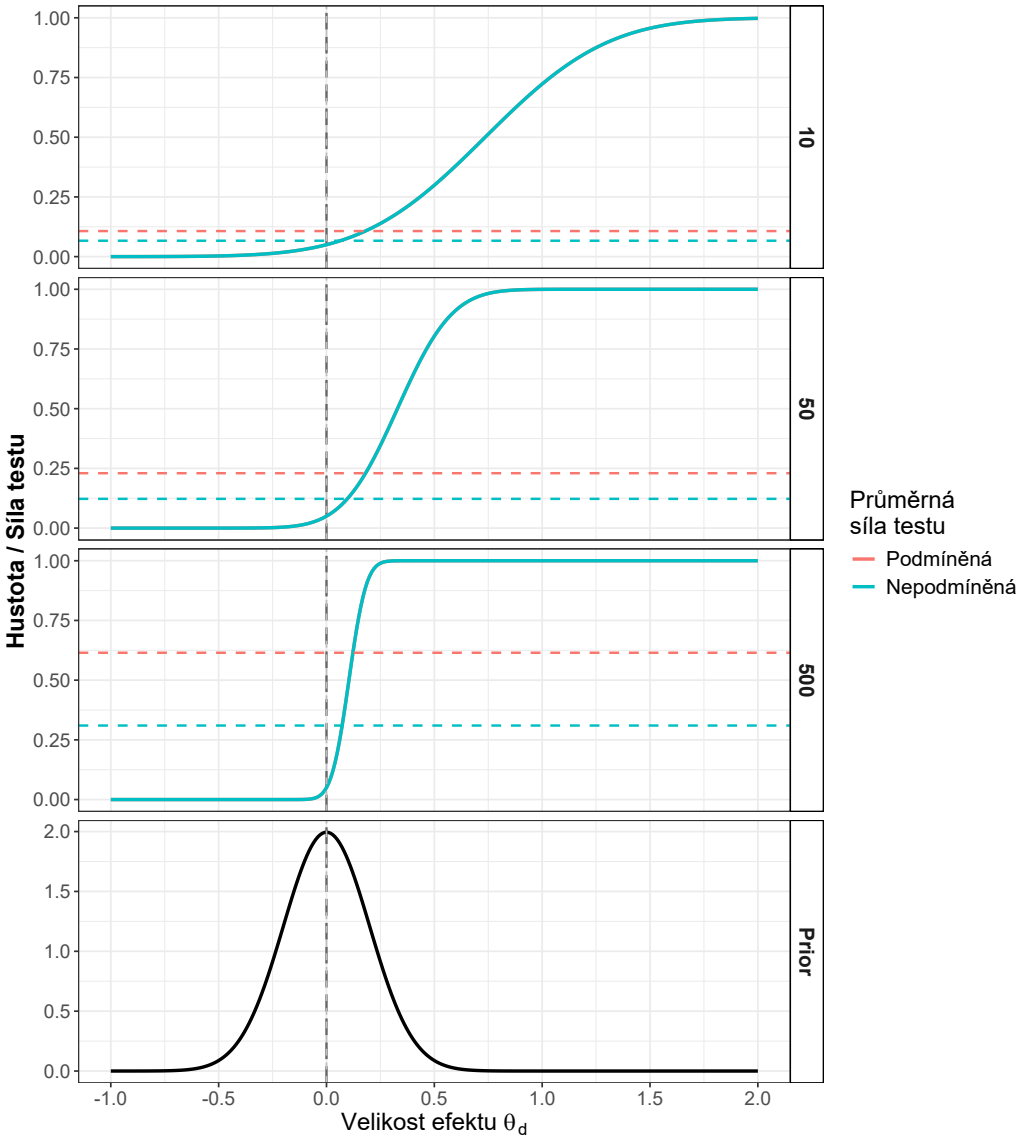
Dále máme apriorní rozdělení, které favorizuje kladný efekt. Z grafu je patrné, že se od sebe oba průměry příliš neliší ani při větším počtu pacientů. Bez ohledu na to, zda bychom mohli předpokládat platnost alternativní hypotézy či nikoliv, bude pro dosažení dostatečné průměrné síly testu (i té podmíněné) stačit rozsah výběru čítající 50 pacientů na skupinu, jelikož se zde síla pohybuje již okolo 80 %.

- apriorní rozdělení $\theta_d \sim N(-1,4; 1)$, viz obr. 10

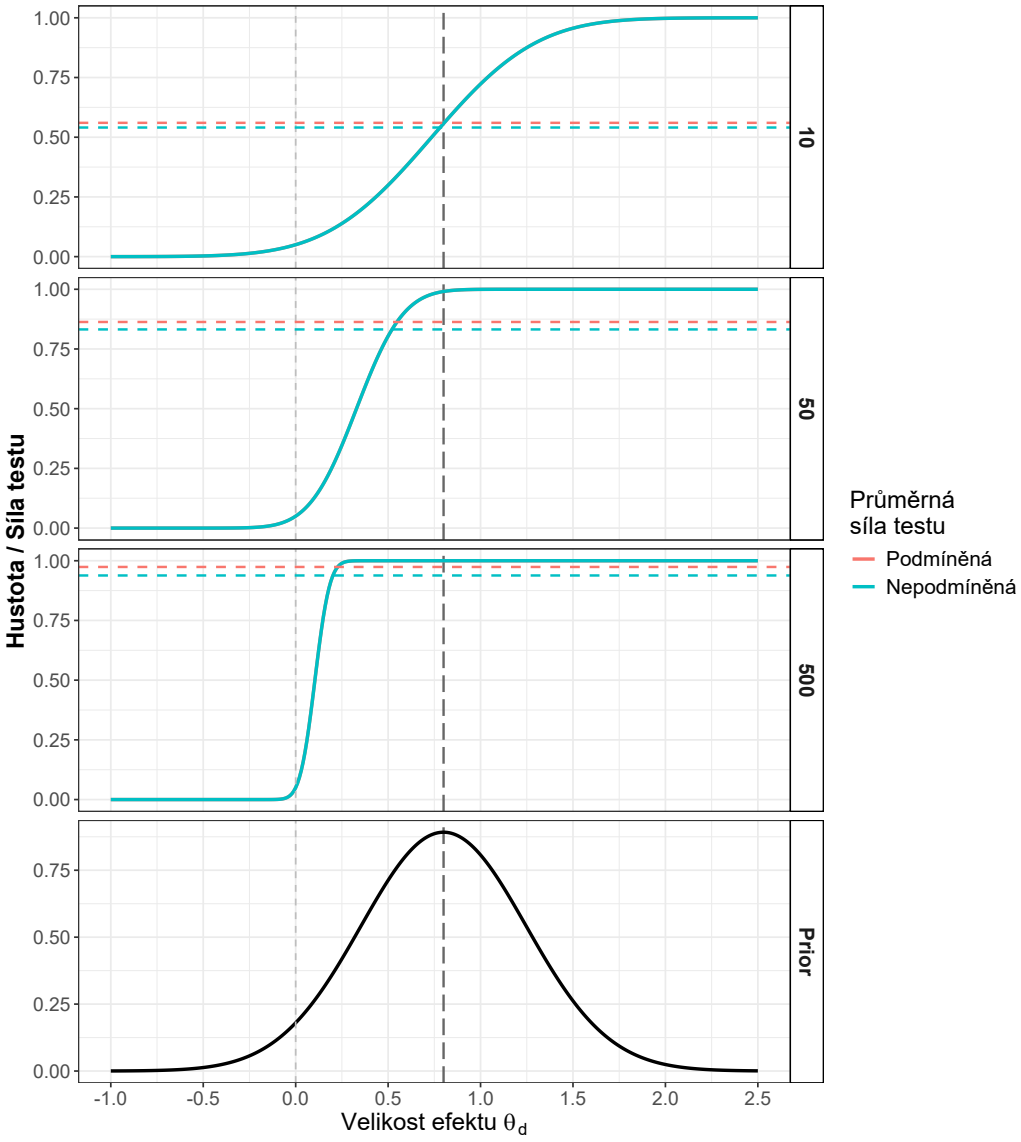
Nejvíce rozdílných hodnot dosahují průměry v případě apriorního rozdělení favorizujícího negativní efekt. Ve všech třech případech nabývá průměrná nepodmíněná síla testu velmi nízkých hodnot, a ani s případným navýšením pacientů nedosáhneme hodnot vyšších. Předpoklad platnosti alternativní hypotézy umožňuje průměrné podmíněné síle testu nabývat hodnot vyšších. Jak můžeme vidět na grafu, nejčastěji požadované síly testu 80 % se dosahuje začleněním zhruba 500 pacientů.

### 4.3.4. Bayesovská síla testu

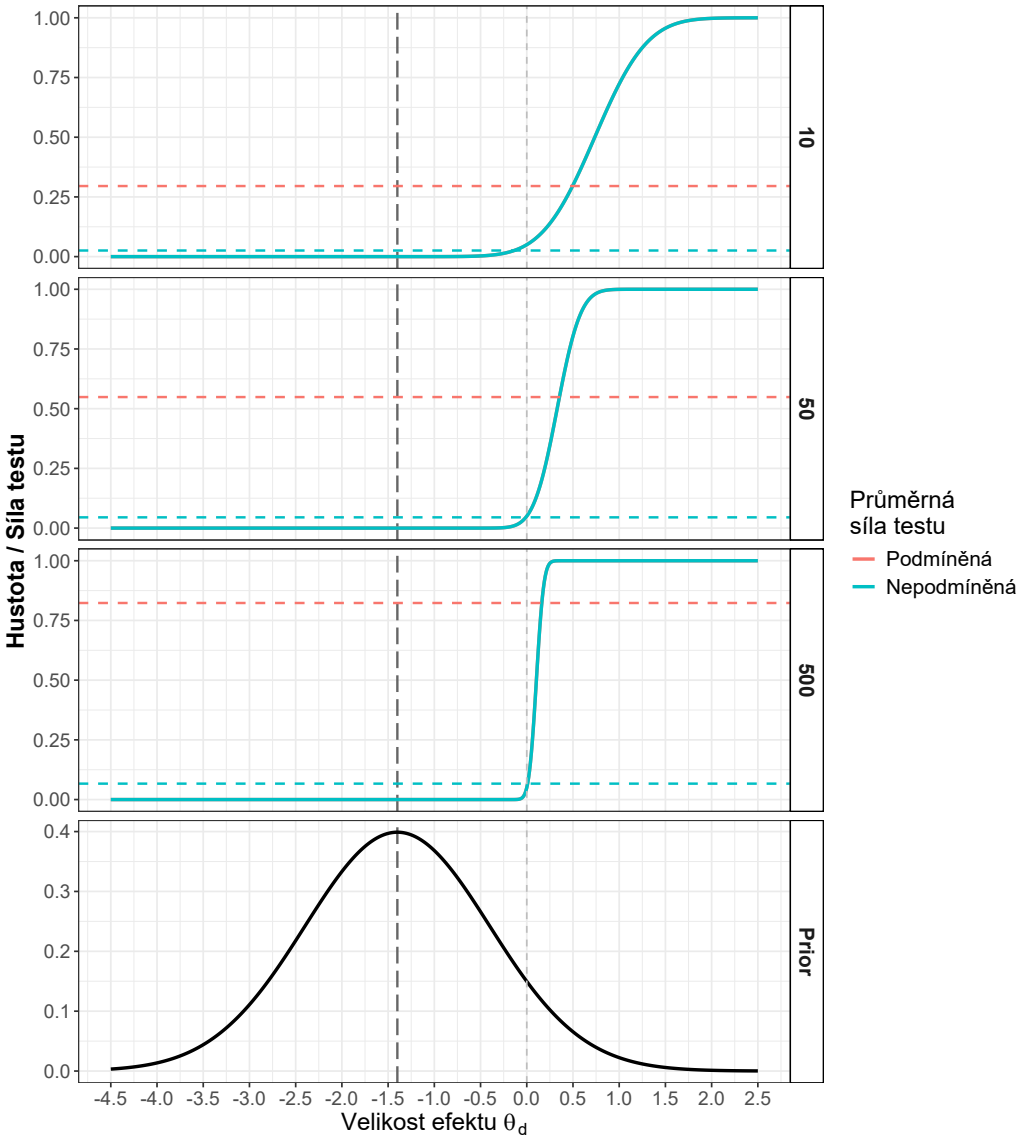
Nyní přejdeme k čistě bayesovské statistice. Vycházet budeme z kapitoly 6.5.3 knihy [4]. Oproti předchozí situaci budeme pracovat s aposteriorní pravděpodobností zamítnutí nulové hypotézy.




Obrázek 8: Graf podmíněné a nepodmíněné průměrné síly testu a klasické silofunkce v závislosti na velikosti vzorku a graf hustoty apriorního rozdělení $\theta_d$. Tyrkysová křivka znázorňuje silofunkci v prvním řádku pro rozsah vzorku 10, ve druhém 50 a ve třetím 500. Tyrkysová čárkovaná čára reprezentuje nepodmíněnou průměrnou sílu testu. Pro vzorek velikosti 10 to je 0,07, pro vzorek velikosti 50 to je 0,12 a pro vzorek velikosti 500 to je 0,31. Lososová čárkovaná čára reprezentuje podmíněnou průměrnou sílu testu. Pro vzorek velikosti 10 to je 0,11, pro vzorek velikosti 50 to je 0,23 a pro vzorek velikosti 500 to je 0,61. Ve čtvrtém řádku se nachází hustota apriorního rozdělení $\theta_d \sim N\,(0; 0,04)$.




Obrázek 9: Graf podmíněné a nepodmíněné průměrné síly testu a klasické silofunkce v závislosti na velikosti vzorku a graf hustoty apriorního rozdělení $\theta_d$. Tyrkysová křivka znázorňuje silofunkci v prvním řádku pro rozsah vzorku 10, ve druhém 50 a ve třetím 500. Tyrkysová čárkovaná čára reprezentuje nepodmíněnou průměrnou sílu testu. Pro vzorek velikosti 10 to je 0,54, pro vzorek velikosti 50 to je 0,83 a pro vzorek velikosti 500 to je 0,94. Lososová čárkovaná čára reprezentuje podmíněnou průměrnou sílu testu. Pro vzorek velikosti 10 to je 0,56, pro vzorek velikosti 50 to je 0,86 a pro vzorek velikosti 500 to je 0,97. Ve čtvrtém řádku se nachází hustota apriorního rozdělení $\theta_d \sim N\left(0,8; 0,2\right)$.




Obrázek 10: Graf podmíněné a nepodmíněné průměrné síly testu a klasické silofunkce v závislosti na velikosti vzorku a graf hustoty apriorního rozdělení $\theta_d$. Tyrkysová křivka znázorňuje silofunkci v prvním řádku pro rozsah vzorku 10, ve druhém 50 a ve třetím 500. Tyrkysová čárkovaná čára reprezentuje nepodmíněnou průměrnou sílu testu. Pro vzorek velikosti 10 to je 0,03, pro vzorek velikosti 50 to je 0,05 a pro vzorek velikosti 500 to je 0,07. Lososová čárkovaná čára reprezentuje podmíněnou průměrnou sílu testu. Pro vzorek velikosti 10 to je 0,30, pro vzorek velikosti 50 to je 0,55 a pro vzorek velikosti 500 to je 0,82. Ve čtvrtém řádku se nachází hustota apriorního rozdělení $\theta_d \sim N(-1,4;1)$.

Připomeňme, že aposteriorní rozdělení parametru $\theta$, které jsme odvodili v kapitole 3.2.3, je ve tvaru $\theta|T_n \sim N\left(\mu_{post}, \sigma^2_{post}\right)$, kde $\mu_{post} = \frac{nT_n+n_0\mu_{prior}}{n+n_0}$ a $\sigma^2_{post} = \frac{\sigma^2}{n+n_0}$. Stále se navíc pohybujeme v kontextu hypotéz $H_0\colon \theta \leq 0$ a $H_A\colon \theta > 0$.

Vycházet budeme z apriorní pravděpodobnosti, že platí nulová hypotéza, tedy $p\left(\theta \leq 0\right)$. Po tom, co získáme v rámci současné studie data $T_n$, dostaneme aposteriorní pravděpodobnost, že platí nulová hypotéza za podmínky pozorovaných dat $T_n$, tedy $p\left(\theta \leq 0|T_n\right)$. Pokud je tato aposteriorní pravděpodobnost malá, budeme se přiklánět k alternativní hypotéze. Proto tuto pravděpodobnost omezíme nějakou malou kladnou konstantou, jakou je například hladina významnosti $\alpha$, tedy $p\left(\theta \leq 0|T_n\right) < \alpha$. Tato podmínka je splněna v následujícím případě:

$$
\begin{aligned}
p\left(\theta \leq 0|T_n\right) = \Phi\left(\frac{0 - \frac{nT_n+n_0\mu_{prior}}{n+n_0}}{\frac{\sigma}{\sqrt{n+n_0}}}\right) &< \alpha \\
\frac{-\frac{nT_n+n_0\mu_{prior}}{n+n_0}}{\frac{\sigma}{\sqrt{n+n_0}}} &< \Phi^{-1}\left(\alpha\right) \\
-\frac{nT_n + n_0\mu_{prior}}{n + n_0} &< \frac{z_\alpha\sigma}{\sqrt{n + n_0}} \\
T_n &> \frac{-z_\alpha\sigma\sqrt{n + n_0} - n_0\mu_{prior}}{n},
\end{aligned}
$$

kde $z_\alpha = \Phi^{-1}\left(\alpha\right)$. Jev $\left(T_n > \frac{-z_\alpha\sigma\sqrt{n+n_0}-n_0\mu_{prior}}{n}\right)$ označíme jako $S^B_\alpha$.

Jev $S_\alpha^B$ nastane s pravděpodobností

$$\begin{aligned} p\left(S_\alpha^B|\theta\right) &= p\left(\frac{T_n-\theta}{\frac{\sigma}{\sqrt{n}}} > \frac{\frac{-z_\alpha\sigma\sqrt{n+n_0}-n_0\mu_{prior}}{n}-\theta}{\frac{\sigma}{\sqrt{n}}}\right) \\ &= p\left(\frac{T_n-\theta}{\frac{\sigma}{\sqrt{n}}} < -\frac{-\sqrt{n_0+n}z_\alpha\sigma-n_0\mu_{prior}-\theta n}{n}\frac{\sqrt{n}}{\sigma}\right) \\ &= \Phi\left(\left(\frac{\sqrt{n_0+n}z_\alpha\sigma}{n}+\frac{n_0\mu_{prior}}{n}+\theta\right)\frac{\sqrt{n}}{\sigma}\right) \\ &= \Phi\left(\frac{\theta\sqrt{n}}{\sigma}+\frac{\mu_{prior}n_0}{\sqrt{n}\sigma}+\frac{\sqrt{n_0+n}}{\sqrt{n}}z_\alpha\right). \end{aligned} \tag{5}$$

Získáváme tedy bayesovskou sílu testu $p\left(S_\alpha^B|\theta\right)$. Vzorec pro bayesovskou sílu (5) můžeme nyní porovnat se vzorcem pro sílu testu klasickou (2). Pokud bychom uvažovali $n_0 \to 0^+$, tedy počet pacientů apriorní studie blížící se 0, rovnala by se bayesovská síla síle testu klasické. Apriorní rozdělení by mělo velký rozptyl, a tak by nám do současné studie žádnou informaci navíc o parametru $\theta$ nepřinášelo.

### 4.3.5. Průměrná bayesovská síla testu

Dále se zaměříme na průměrnou bayesovskou sílu testu. Průměrná bayesovská síla testu bude opět představovat vážený průměr přes všechny možné hodnoty parametru $\theta$, tentokrát však pravděpodobnosti $p\left(S_\alpha^B|\theta\right)$. Interpretace výsledků bude ekvivalentní situaci s průměrnou klasickou silofunkcí. Průměrnou bayesovskou sílu testu můžeme získat buďto integrací součinu silofunkce $p\left(S_\alpha^B|\theta\right)$ a apriorní hustoty $p\left(\theta\right)$ přes všechny možné hodnoty $\theta$:

$$p\left(S_\alpha^B\right) = \int_{-\infty}^{\infty} p\left(S_\alpha^B|\theta\right) p\left(\theta\right) d\theta,$$

nebo přímým výpočtem pravděpodobnosti $p\left(S_\alpha^B\right)$ s využitím nepodmíněného rozdělení veličiny $T_n \sim N\left(\mu_{prior}, \sigma^2\left(\frac{1}{n}+\frac{1}{n_0}\right)\right)$:

$$\begin{aligned} p\left(S_\alpha^B\right) &= p\left(\frac{T_n-\mu_{prior}}{\sigma\sqrt{\frac{1}{n_0}+\frac{1}{n}}} > \frac{\frac{-z_\alpha\sigma\sqrt{n+n_0}-n_0\mu_{prior}}{n}-\mu_{prior}}{\sigma\sqrt{\frac{1}{n}+\frac{1}{n_0}}}\right) \\ &= \Phi\left(\frac{\frac{z_\alpha\sigma\sqrt{n+n_0}+n_0\mu_{prior}+n\mu_{prior}}{n}}{\sigma\sqrt{\frac{1}{n}+\frac{1}{n_0}}}\right) \\ &= \Phi\left(\frac{z_\alpha}{\frac{n}{\sqrt{nn_0}}}+\frac{\mu_{prior}\left(n+n_0\right)}{\sigma n\sqrt{\frac{n+n_0}{nn_0}}}\right) \\ &= \Phi\left(\sqrt{\frac{n_0}{n}}z_\alpha+\frac{\mu_{prior}\sqrt{n+n_0}\sqrt{n_0}}{\sigma\sqrt{n}}\right). \end{aligned}$$

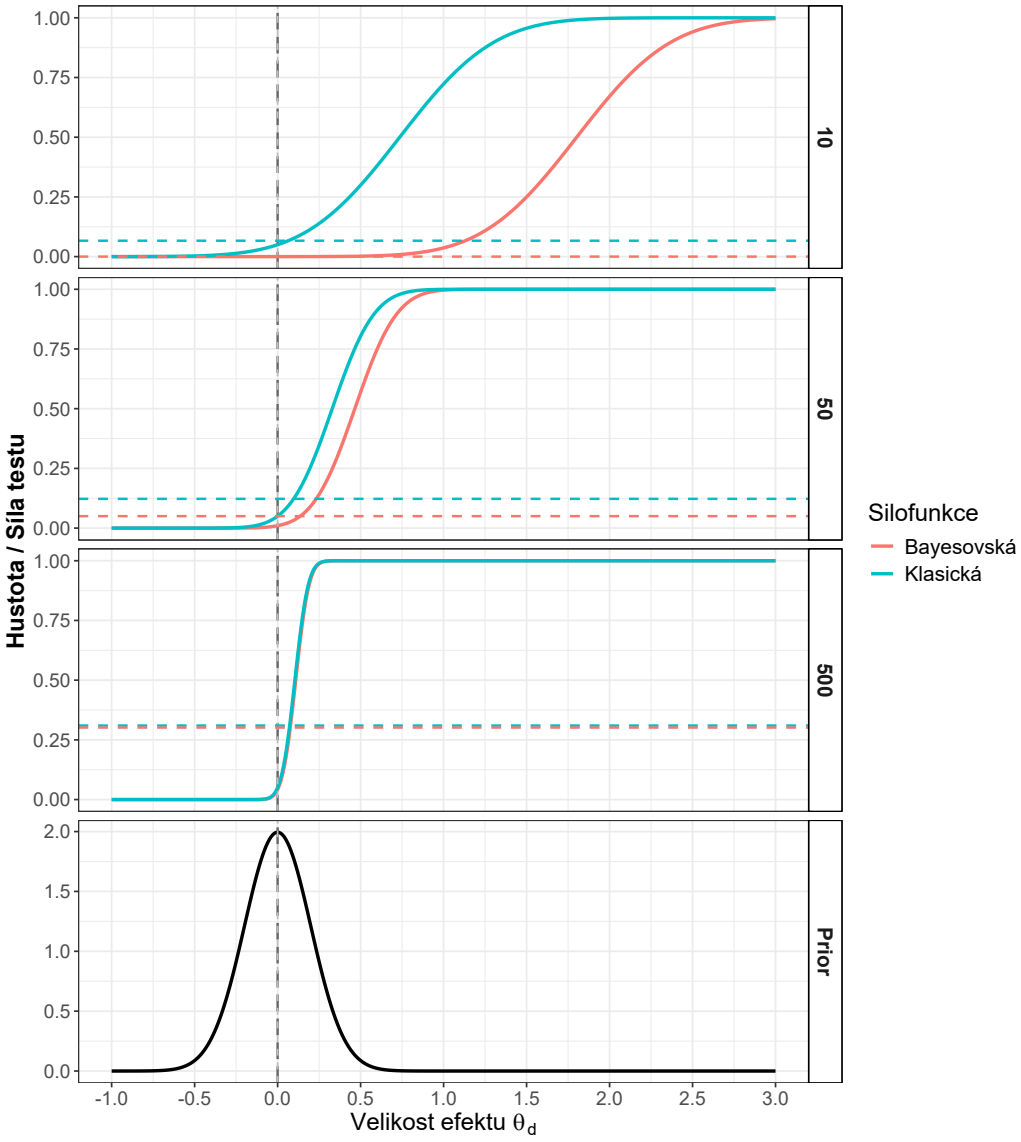
Stejně jako v předchozí situaci bychom mohli v případě bayesovské síly předpokládat platnost alternativní hypotézy a počítat tak průměrnou podmíněnou bayesovskou sílu testu. Postup by byl obdobný jako v předchozí situaci.

### 4.3.6. Srovnání klasické a bayesovské silofunkce a jejich průměrných hodnot

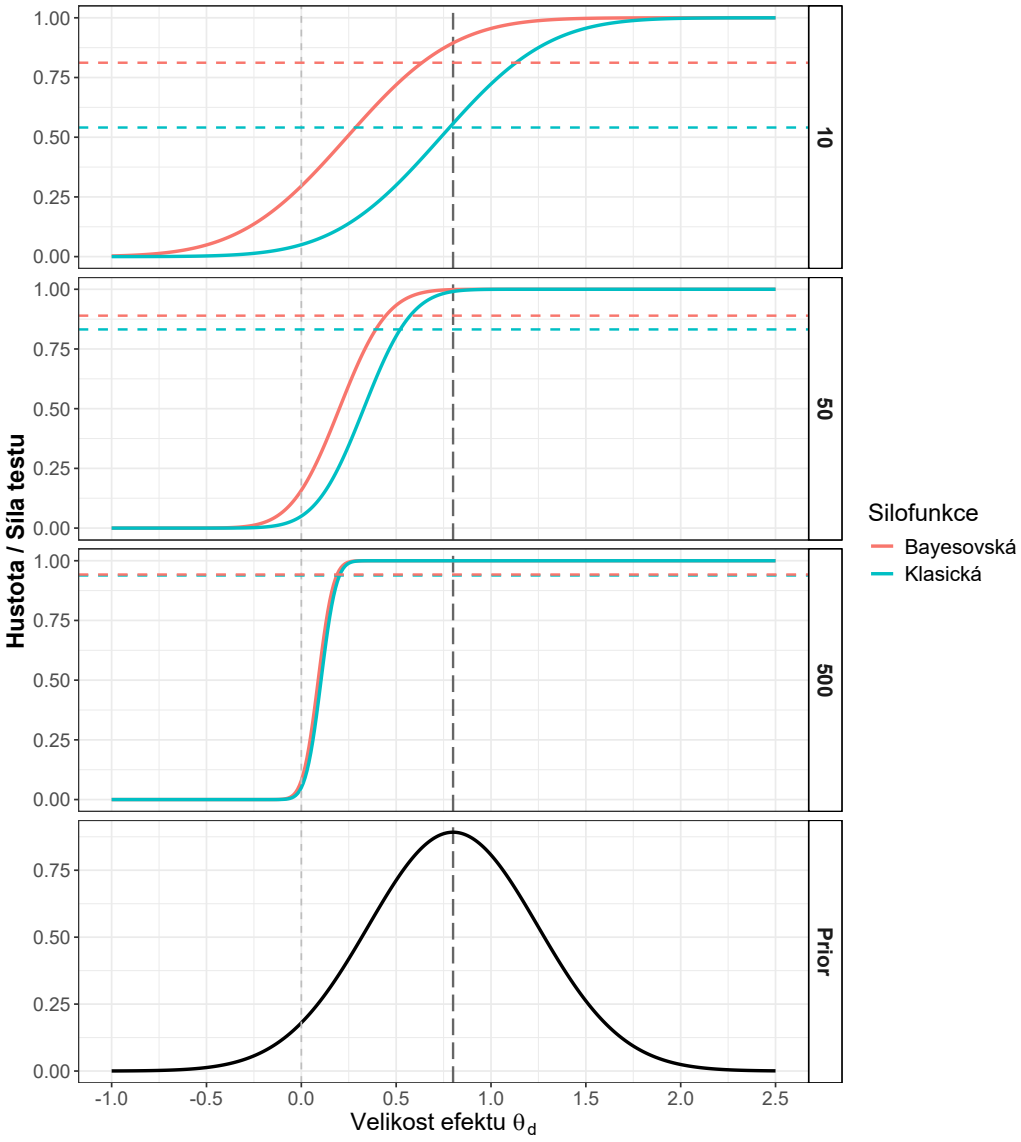
Nyní porovnáme klasickou silofunkci se silofunkcí bayesovskou a průměrnou sílu testu s průměrnou bayesovskou sílou testu na naší studii:

- apriorní rozdělení $\theta_d \sim N\left(0; 0,04\right)$, viz obr. 11

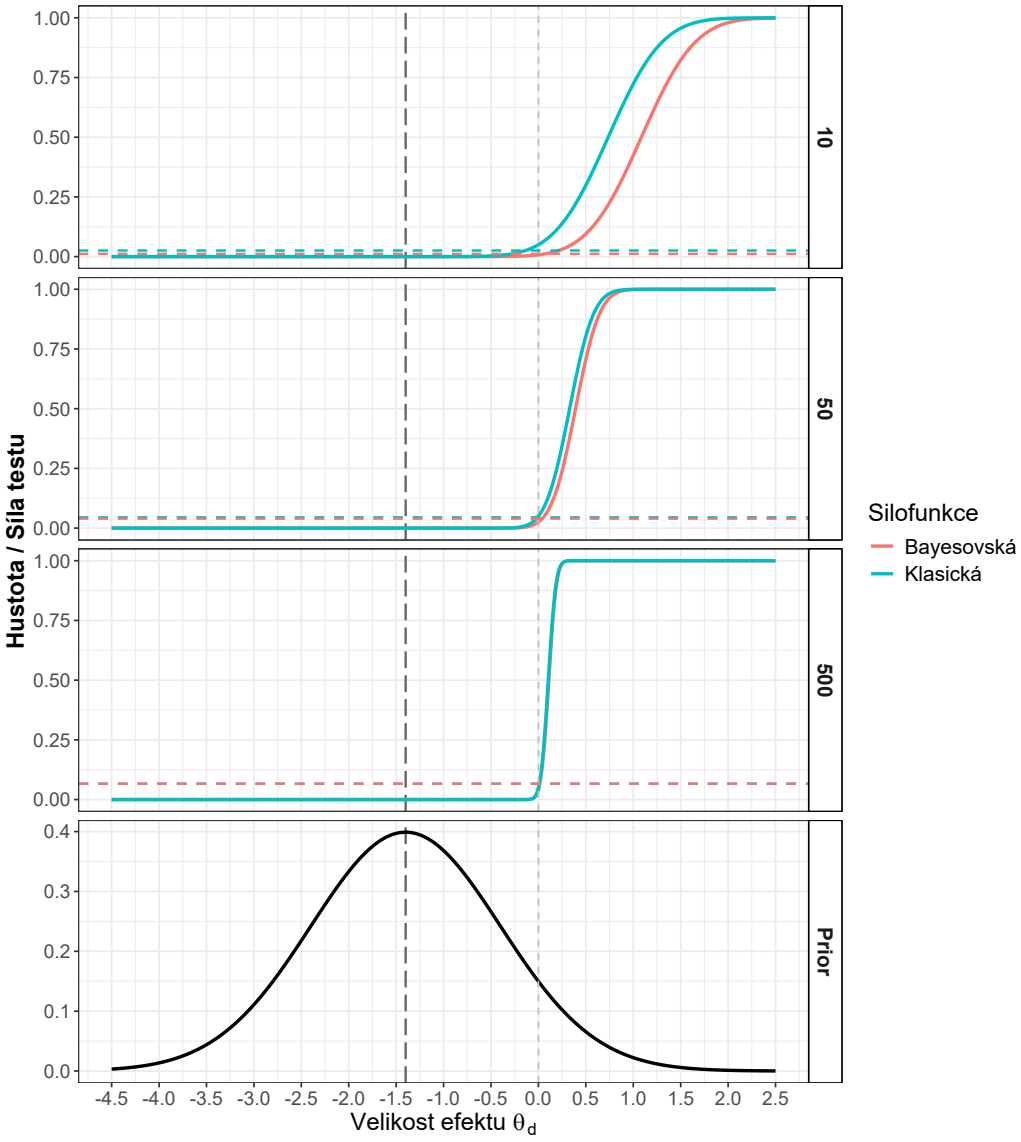
Na prvním příkladu můžeme vidět, že bayesovská silofunkce ukazuje pro malé velikosti efektu velmi nízkou až skoro nulovou pravděpodobnost odhalení efektu. Stejně je tomu i pro průměrnou bayesovskou sílu, jejíž hodnota se pohybuje blízko 0. Apriorní rozdělení s nemalou pravděpodobností favorizuje nulový až negativní efekt, a proto je i bayesovská síla k odhalení




Obrázek 11: Graf klasické a bayesovské silofunkce, průměrné klasické a bayesovské síly testu a hustoty apriorního rozdělení $\theta_d$. Tyrkysová, resp. lososová křivka znázorňuje klasickou, resp. bayesovskou silofunkci v prvním řádku pro rozsah vzorku 10, ve druhém 50 a ve třetím 500. Tyrkysová čárkovaná čára reprezentuje průměrnou sílu testu. Pro vzorek velikosti 10 to je 0,07, pro vzorek velikosti 50 to je 0,12 a pro vzorek velikosti 500 to je 0,31. Lososová čárkovaná čára reprezentuje průměrnou bayesovskou sílu testu. Pro vzorek velikosti 10 to je méně než 0,01, pro vzorek velikosti 50 to je 0,05 a pro vzorek velikosti 500 to je 0,30. Ve čtvrtém řádku se nachází hustota apriorního rozdělení $\theta_d \sim N\,(0; 0,04)$.




Obrázek 12: Graf klasické a bayesovské silofunkce, průměrné klasické a bayesovské síly testu a hustoty apriorního rozdělení $\theta_d$. Tyrkysová, resp. lososová křivka znázorňuje klasickou, resp. bayesovskou silofunkci v prvním řádku pro rozsah vzorku 10, ve druhém 50 a ve třetím 500. Tyrkysová čárkovaná čára reprezentuje průměrnou sílu testu. Pro vzorek velikosti 10 to je 0,54, pro vzorek velikosti 50 to je 0,83 a pro vzorek velikosti 500 to je 0,94. Lososová čárkovaná čára reprezentuje průměrnou bayesovskou sílu testu. Pro vzorek velikosti 10 to je 0,81, pro vzorek velikosti 50 to je 0,89 a pro vzorek velikosti 500 to je 0,94. Ve čtvrtém řádku se nachází hustota apriorního rozdělení $\theta_d \sim N\left(0,8; 0,2\right)$.




Obrázek 13: Graf klasické a bayesovské silofunkce, průměrné klasické a bayesovské síly testu a hustoty apriorního rozdělení $\theta_d$. Tyrkysová, resp. lososová křivka znázorňuje klasickou, resp. bayesovskou silofunkci v prvním řádku pro rozsah vzorku 10, ve druhém 50 a ve třetím 500. Tyrkysová čárkovaná čára reprezentuje průměrnou sílu testu. Pro vzorek velikosti 10 to je 0,03, pro vzorek velikosti 50 to je 0,05 a pro vzorek velikosti 500 to je 0,07. Lososová čárkovaná čára reprezentuje průměrnou bayesovskou sílu testu. Pro vzorek velikosti 10 to je přibližně 0,01, pro vzorek velikosti 50 to je 0,04 a pro vzorek velikosti 500 to je 0,07. Ve čtvrtém řádku se nachází hustota apriorního rozdělení $\theta_d \sim N\left(-1,4;1\right)$.

kladného efektu pro menší rozsahy výběru poměrně skeptická. Klasická silofunkce, která toto apriorní rozdělení nezohledňuje, ukazuje pravděpodobnost odhalení menších efektů o něco větší. Pro větší rozsah čítající 50 pacientů na skupinu udává již bayesovská silofunkce větší pravděpodobnost odhalení případného efektu i pro menší efekt, průměrná bayesovská síla zůstává však stále v řádu setin, stejně tak i klasická průměrná síla testu. S rostoucím počtem pacientů se bayesovská silofunkce, resp. bayesovská průměrná síla blíží klasické silofunkci, resp. průměrné síle.

- apriorní rozdělení $\theta_d \sim N\,(0,8; 0,2)$, viz obr. 12

V případě tohoto rozdělení je situace opačná. Vzhledem k tomu, že apriorní rozdělení předpokládá především kladný efekt, udává bayesovská síla testu větší pravděpodobnost odhalení i menšího efektu. Stejně tak je i průměrná bayesovská síla už pro velikost rozsahu 10 dostačující. Konkrétně je rovna 0,81, na rozdíl od té klasické, která je rovna 0,54. S rostoucím počtem pacientů se tentokrát klasická silofunkce, resp. klasická průměrná síla blíží bayesovské silofunkci, resp. průměrné bayesovské síle.

- apriorní rozdělení $\theta_d \sim N\,(-1,4; 1)$, viz obr. 13

Stejně jako v prvním případě popisuje bayesovská silofunkce pro malé rozsahy výběru menší schopnost testu detekovat menší efekt oproti silofunkci klasické. Je tomu tak proto, že apriorní rozdělení předpokládá především negativní efekt. Bayesovská silofunkce tuto skutečnost na rozdíl od klasické silofunkce zohledňuje. S větším počtem pacientů však opět schopnost detekce stoupá a blíží se klasické silofunkci. Průměrné síly testu jsou v obou případech velmi nízké.

### 4.3.7. Explicitní vztah pro výpočet velikosti výběru

V závěru kapitoly se ještě podíváme zpět na vztah pro průměrnou sílu testu $p\left(S_\alpha^C\right) = \Phi\left(\sqrt{\frac{n_0}{n+n_0}}\left(z_\alpha + \frac{\sqrt{n}\mu_{prior}}{\sigma}\right)\right)$ a pro průměrnou bayesovskou sílu testu $p\left(S_\alpha^B\right) = \Phi\left(\sqrt{\frac{n_0}{n}}z_\alpha + \frac{\mu_{prior}\sqrt{n+n_0}\sqrt{n_0}}{\sigma\sqrt{n}}\right)$. Z těchto vztahů jsme schopni analyticky vyjádřit velikost výběru $n$. Pokud budeme mít předem dané apriorní rozdělení, hladinu významnosti a požadovanou průměrnou sílu testu, ať už klasickou nebo bayesovskou, lze na základě těchto vztahů určit velikost výběru odpovídající uvedeným parametrům.

Začneme odvozením pro klasickou sílu testu:

$$\begin{aligned} p\left(S_\alpha^C\right) &= \Phi\left(\sqrt{\frac{n_0}{n+n_0}}\left(z_\alpha + \frac{\sqrt{n}\mu_{prior}}{\sigma}\right)\right) \\ \Phi^{-1}\left(p\left(S_\alpha^C\right)\right) &= \sqrt{\frac{n_0}{n+n_0}}\left(z_\alpha + \frac{\sqrt{n}\mu_{prior}}{\sigma}\right) \\ &= \frac{\mu_{prior}\sqrt{n_0}}{\sigma}\sqrt{\frac{k}{k+1}} + z_\alpha\sqrt{\frac{1}{k+1}}, \end{aligned}$$

kde jsme použili substituci $n = kn_0$. Dále si označíme konstanty $c_1 = z_\alpha$, $c_2 = \frac{\mu_{prior}\sqrt{n_0}}{\sigma}$ a $c_3 = \Phi^{-1}\left(p\left(S_\alpha^C\right)\right)$:

$$c_3 = \sqrt{\frac{k}{k+1}}c_2 + \sqrt{\frac{1}{k+1}}c_1 \tag{6}$$

$$(k+1)\,c_3^2 = kc_2^2 + 2c_1c_2\sqrt{k} + c_1^2 \tag{7}$$

$$\begin{aligned} 0 &= k^2\left(c_3^2 - c_2^2\right)^2 + k\left(2\left(c_3^2 - c_1^2\right)\left(c_3^2 - c_2^2\right) - 4c_1^2c_2^2\right) + \left(c_3^2 - c_1^2\right)^2 \\ k_{1,2} &= \frac{\pm 2\sqrt{c_1^2c_2^2c_3^2\left(c_1^2 + c_2^2 - c_3^2\right)} + c_1^2c_2^2 + c_1^2c_3^2 + c_2^2c_3^2 - c_3^4}{\left(c_2^2 - c_3^2\right)^2} \\ n_{1,2} &= k_{1,2}n_0. \end{aligned}$$

Při přechodu od rovnice (6) k (7) dochází k neekvivalentní úpravě, proto je

potřeba existenci kořenů $n_{1,2}$ dosazením do rovnice (6) ověřit. Tím dostaneme pouze jeden kořen $n$.

Nyní se podíváme na sílu testu bayesovskou:

$$\begin{aligned} p\left(S_\alpha^B\right) &= \Phi\left(\sqrt{\frac{n_0}{n}}z_\alpha + \frac{\mu_{prior}\sqrt{n+n_0}\sqrt{n_0}}{\sigma\sqrt{n}}\right) \\ \Phi^{-1}\left(p\left(S_\alpha^B\right)\right) &= \sqrt{\frac{n_0}{n}}z_\alpha + \frac{\mu_{prior}\sqrt{n+n_0}\sqrt{n_0}}{\sigma\sqrt{n}} \\ &= \sqrt{j}z_\alpha + \sqrt{j+1}\frac{\mu_{prior}\sqrt{n_0}}{\sigma}, \end{aligned}$$

kde jsme použili substituci $j = \frac{n_0}{n}$. Dále si opět označíme konstanty $c_1 = z_\alpha$, $c_2 = \frac{\mu_{prior}\sqrt{n_0}}{\sigma}$ a nově označíme $c_4 = \Phi^{-1}\left(p\left(S_\alpha^B\right)\right)$:

$$\begin{aligned} c_4 &= \sqrt{j+1}c_2 + \sqrt{j}c_1 \\ \sqrt{j+1}c_2 &= c_4 - \sqrt{j}c_1 \qquad (8) \\ (j+1)\,c_2^2 &= c_4^2 - 2\sqrt{j}c_1c_4 + jc_1^2 \qquad (9) \\ 0 &= j^2\left(c_1^2 - c_2^2\right)^2 + j\left(2\left(c_1^2 - c_2^2\right)\left(c_4^2 - c_2^2\right) - 4c_1^2c_4^2\right) + \left(c_3^2 - c_2^2\right)^2 \\ j_{1,2} &= \frac{\pm 2\sqrt{c_1^2c_2^2c_4^2\left(c_1^2 - c_2^2 + c_4^2\right)} + c_1^2c_2^2 + c_1^2c_4^2 + c_2^2c_4^2 - c_2^4}{\left(c_1^2 - c_2^2\right)^2} \\ n_{1,2} &= \frac{n_0}{j_{1,2}}. \end{aligned}$$

Při přechodu od rovnice (8) k (9) dochází opět k neekvivalentní úpravě, proto musíme stejně jako v předchozí situaci ověřit existenci kořenů $n_{1,2}$ dosazením do rovnice (8). Konečně dostaneme opět pouze jeden kořen $n$.

Je třeba si uvědomit, že v případě zadání průměrné síly testu (frekventistické či bayesovské), které nelze dosáhnout – viz příklady v kapitole 4.3.3 a 4.3.6– není možné výpočet provést. Dále je nutné uvažovat pouze $n \in \mathbb{N}$, neboť jde o počty pacientů. Výslednou hodnotu je tedy třeba zaokrouhlit na celé číslo.

## 4.4. Rozdělení síly testu

V této kapitole si představíme, jakým způsobem můžeme díky apriornímu rozdělení $\theta$ získat pravděpodobností rozdělení síly testu pro konkrétní volbu velikosti výběru $n$. Díky tomuto rozdělení jsme schopni posoudit, zda je konkrétní velikost výběru $n$ pro naši studii dostačující, nebo je třeba ji navýšit.V případě, kdy bude rozdělení favorizovat zbytečně vysoké hodnoty síly testu, je vhodné velikost výběru naopak snížit.

Vycházet budeme opět z apriorního rozdělení parametru $\theta \sim N\left(\mu_{prior}, \frac{\sigma^2}{n_0}\right)$. Každé hodnotě $\theta$ odpovídá určitá síla testu, tedy $\varphi(\theta) \in [0;1]$. Jelikož $\theta$ je náhodná veličina, je i $\varphi(\theta)$ náhodná veličina, kterou budeme označovat symbolem $W$.

Ke konstrukci pravděpodobnostního rozdělení síly testu budeme přistupovat dvojím způsobem. Jako první si představíme metodu založenou na simulaci. Princip této metody spočívá v tom, že z distribuce apriorního rozdělení parametru $\theta$, jehož hustotu označíme jako $f_\theta(\theta)$, budeme nejprve simulovat námi zvolený počet hodnot efektu $\theta$, např. 10 000. Každé vygenerované hodnotě efektu $\theta$ přiřadíme sílu testu pro odhalení efektu této velikosti. Rozdělení takto získaných hodnot $W$ můžeme následně vizualizovat histogramem, čímž získáme odhad hustoty veličiny $W$, kterou označíme $g_W(w)$. Tato hustota charakterizuje rozdělení síly testu. Podle tohoto odhadu jsme následně schopni rozhodovat o adekvátnosti volby rozsahu výběru.

Druhou metodou můžeme získat hustotu veličiny $W$ prostřednictvím následujícího vztahu (viz [3], věta 2.4):

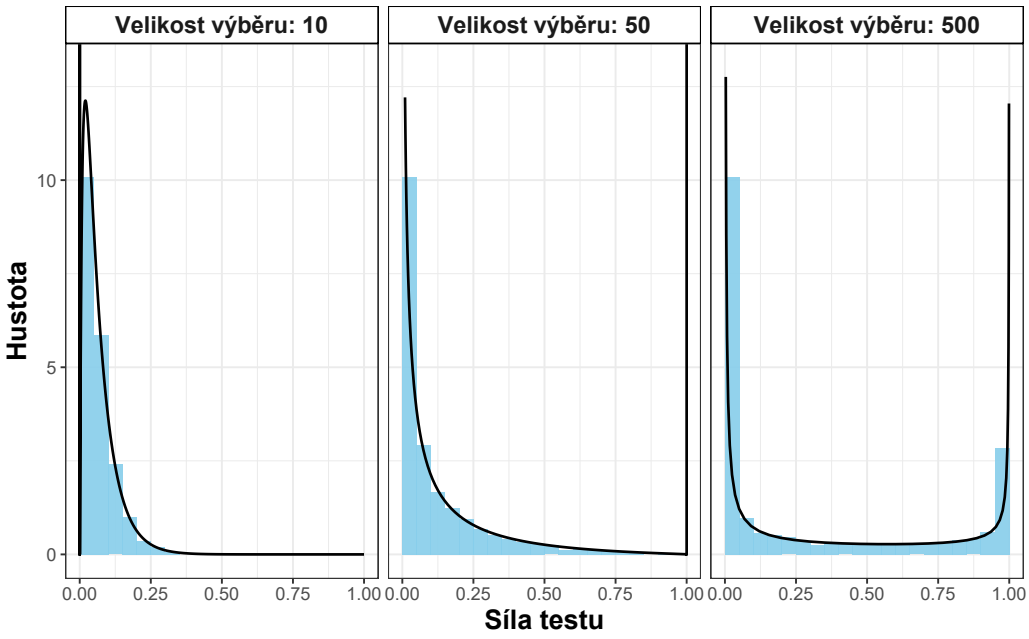
$$g_W(w) = f_\theta\left[\varphi^{-1}(w)\right] \mid \left[\varphi^{-1}(w)\right]' \mid .$$

Nevýhodou tohoto výpočtu je, že neznáme explicitní vztah pro funkci $\varphi(\theta)$. Neznáme tedy ani vztah pro inverzní funkci, a tím ani derivaci této inverze. Tento problém má však řešení. Jelikož jsme schopni každé hodnotě efektu $\theta$ přiřadit konkrétní sílu testu, jsme i naopak schopni zpětně přiřadit síle testu odpovídající efekt. Tímto postupem tedy získáme inverzi $\varphi^{-1}(w)$. Derivaci této funkce jsme následně schopni získat prostřednictvím numerické derivace.

Nyní představíme použití obou metod na naší ukázkové studii, kde jsme pro výpočet derivace použili formuli pro centrální diferenční derivaci (viz [1], rovnice 4.5):

- apriorní rozdělení $\theta_d \sim N(0; 0,04)$, viz obr. 14

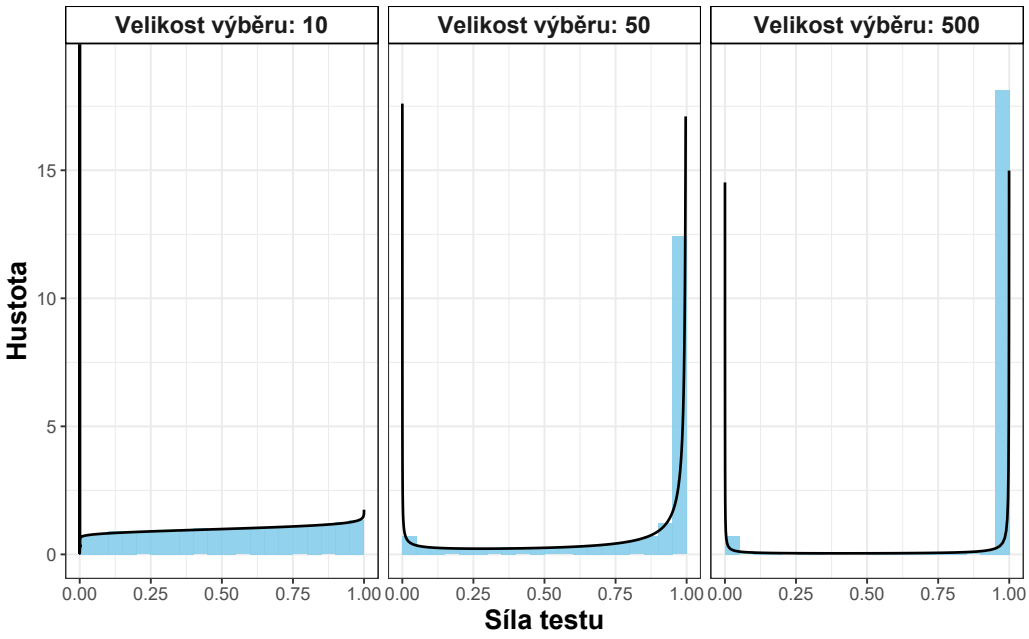
První případ vychází z apriorní hustoty symetrické kolem 0. Pro velikost výběru 10 se hustota síly testu koncentruje v oblasti velmi malých hodnot a s rostoucí silou testu rychle klesá. S rostoucím počtem pacientů poroste i síla testu, jak bylo patrné již na vizualizaci hustoty apriorního rozdělení a silofunkcí, viz obr. 5. Stejný trend můžeme pozorovat i zde. Hustota pro velikost výběru 50 je ve srovnání s hustotou pro velikost výběru 10 více zešikmená doprava, a zahrnuje tedy i vyšší hodnoty síly testu, ač stále v menší míře. Pro větší rozsahy výběru můžeme pozorovat podobnou situaci, jako jsme popisovali v kapitole 4.3.3 pro nepodmíněnou sílu testu. Pokud bychom dále navyšovali rozsah výběru a silofunkce by tedy velmi rychle nabývala hodnot blízkých 1, bude přibližně 50 % generovaných hodnot $\theta$ odpovídat síle testu blízké 1, zatímco zbývajících 50 % bude mít sílu testu blízkou 0. Průměrná síla testu tedy bude činit přibližně 50 %.




Obrázek 14: Graf hustoty síly testu a histogram znázorňující odhad hustoty síly testu pro hustotu apriorního rozdělení $\theta_d \sim N\,(0; 0,04)$.

- apriorní rozdělení $\theta_d \sim N\,(0,8; 0,2)$, viz obr. 15

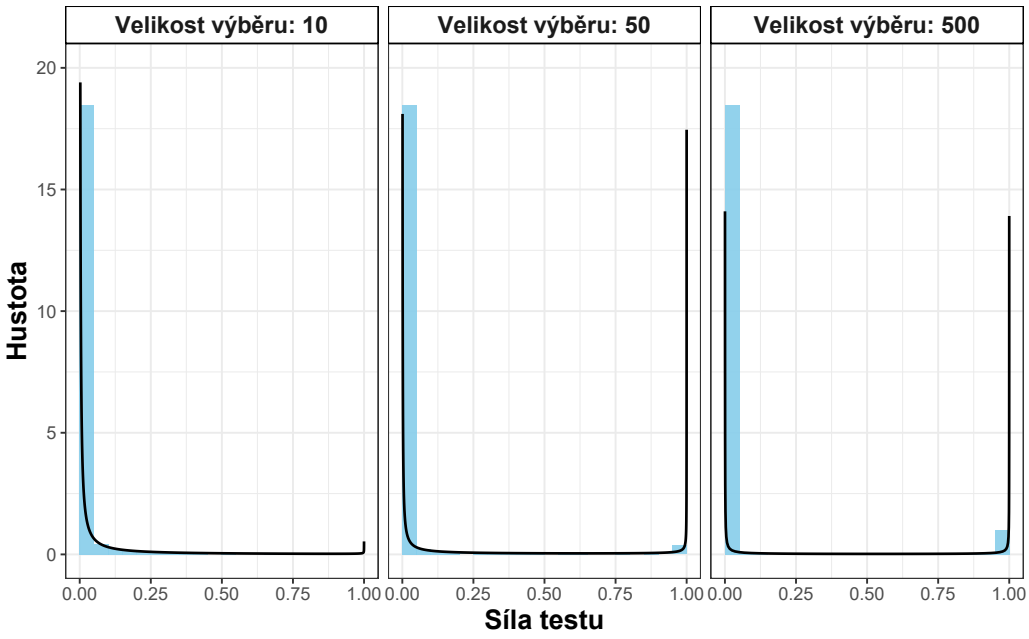
V tomto případě favorizuje apriorní hustota pozitivní efekt. Stejně jako jsme uváděli u předchozích metod v kontextu tohoto apriorního rozdělení, i zde můžeme pozorovat, že rozsah výběru čítající 500 pacientů na skupinu je už zbytečně velký – většině generovaných hodnot je přiřazena síla testu blízká 1. U rozsahu výběru 50 se hustota síly testu stále koncentruje u relativně vysokých hodnot. Naopak v případě velikosti výběru 10 mohou hodnoty síly testu podobně nabývat nízkých, středních i vysokých hodnot, což ukazuje na větší variabilitu možných výstupů testu. Z toho vyplývá, že by bylo vhodné volit rozsah velikosti výběru mezi 10 a 50.




Obrázek 15: Graf hustoty síly testu a histogram znázorňující odhad hustoty síly testu pro hustotu apriorního rozdělení $\theta_d \sim N\,(0,8; 0,2)$.

- apriorní rozdělení $\theta_d \sim N\,(-1,4; 1)$, viz obr. 16

V rámci posledního případu vycházíme z rozdělení $\theta$, které favorizuje negativní efekt. Tato skutečnost se výrazně odráží ve výsledné hustotě síly testu. Pro velikost vzorku 10, a obdobně i pro vzorek velikosti 50, se hustota síly testu koncentruje převážně u 0. U velikosti vzorku 500 sice stále dominuje vysoká pravděpodobnost nízkých hodnot síly testu, avšak objevuje se i určitá pravděpodobnost výskytu vysokých hodnot blízkých jedné. K této situaci dochází proto, že klasická silofunkce při této velikosti výběru nabývá velmi rychle hodnot blízkých jedné, a tudíž je většina hodnot $\theta > 0$ spojena s vysokou silou testu. Stále je však síla testu ve většině případů velmi malá. Vhodné tedy bude přehodnotit návrh klinické studie a především volbu hypotéz.




Obrázek 16: Graf hustoty síly testu a histogram znázorňující odhad hustoty síly testu pro hustotu apriorního rozdělení $\theta_d \sim N(-1, 4; 1)$.

# Závěr

Cílem práce bylo čtenáři představit metody bayesovské statistiky pro stanovení rozsahu výběru v klinických studiích a tyto metody dále demonstrovat pomocí statistického softwaru R.

Před samotným představením metod bylo třeba vytvořit teoretický rámec. V první kapitole byl vysvětlen pojem klinická studie, kapitola druhá se poté zaměřovala na způsob formulování hypotéz v kontextu klinické studie a také na možné závěry testování těchto hypotéz. Třetí kapitola byla věnována bayesovské inferenci. Bylo představeno, jak při bayesovské inferenci postupovat. Dále byla vysvětlena role apriorního rozdělení, věrohodnosti a následně odvozeného aposteriorního rozdělení.

V hlavní části práce se podařilo představit čtyři základní metody pro zohlednění apriorní představy při designu klinické studie, zejména při stanovení rozsahu výběru. První z nich je vizualizace silofunkce a apriorního rozdělení. Tato metoda slouží především pro rychlé a jednoduché posouzení, zda byl rozsah výběru v kontextu dané studie zvolen adekvátně. Pro případ, kdy má apriorní rozdělení malý rozptyl, je vhodná metoda využívající střední hodnotu apriorního rozdělení jako bodovou alternativu. Tuto metodu jsme jako jedinou do prostředí softwaru R neimplementovali, protože se jedná pouze o úpravu hypotéz.

Třetí studovaná metoda se zaměřuje na sílu testu, průměrnou sílu testu a jejich protějšky v bayesovské statistice. Průměrné síly testu můžeme v praxi využít jako ukazatele schopnosti testu zamítnout nulovou hypotézu. Uvedli jsme také situaci, kdy lze průměrnou sílu testu podmínit platností alternativní hypotézy. Jednotlivé síly a silofunkce byly následně porovnány na ilustrativní studii. V rámci této části se navíc podařilo odvodit explicitní

vztahy pro výpočet velikosti výběru pro námi zvolené vstupní parametry. Pokud je nám známo, žádný z uvedených ani jiných dostupných zdrojů tyto vztahy neuvádí. Jsou tedy naším vlastním přínosem. V poslední kapitole byla popsána možnost využití rozdělení síly testu.

Metody byly implementovány a aplikovány na ilustrativní studii v prostředí softwaru R. Získané výsledky byly dále analyzovány. Příslušné kódy jsou k dispozici v příloze této bakalářské práce.

# Literatura


[1] BURDEN, Richard L. a FAIRES, J. Douglas. Numerical analysis. 9th ed. Boston: Brooks/Cole, c2011. ISBN 0-538-73351-9.

[2] GELMAN, Andrew. Bayesian data analysis. 3rd ed. Chapman & Hall/CRC texts in statistical science. Boca Raton: CRC Press, 2014. ISBN 978-1-4398-4095-5.

[3] HRON, Karel; KUNDEROVÁ, Pavla a VENCÁLEK, Ondřej. Základy počtu pravděpodobnosti a metod matematické statistiky. 4. doplněné vydání. Olomouc: Univerzita Palackého v Olomouci, Přírodovědecká fakulta, 2021. ISBN 978-80-244-5990-5.

[4] SPIEGEHALTER, David; ABRAMS, Keith R. a MYLES, Jonathan P. Bayesian approaches to clinical trials and health-care evaluation. Statistics in practice. Hoboken: John Wiley, c2004. ISBN 0-471-49975-7.

[5] SVOBODNÍK, Adam; DEMLOVÁ, Regina a PECEN, Ladislav. Klinické studie v praxi. Brno: Facta Medica, 2014. ISBN 978-80-904731-8-8.

[6] Zákon č. 378/2007 Sb., o léčivech: ve znění účinném od 1.1.2025.