A Bayes-becslés és a bayesi hipotézisvizsgálat

Szakdolgozat

Eötvös Loránd Tudományegyetem

Természettudományi Kar

Valószínűségelméleti és Statisztika Tanszék

Dr. Csiszár Villő

adjunktus

Kiss-Fülöp Zsófia

Alkalmazott matematikus BSc

Budapest, 2020

Tartalomjegyzék

Köszönetnyilvánítás

Szeretném megköszönni témavezetőmnek, Dr. Csiszár Villőnek, aki támogatott a témaválasztásban, ösztönzött, valamint odafigyelésével és türelmével egyengette a szakdolgozatom létrejöttét az elmúlt félévben.

Rengeteg hálával tartozom férjemnek és szüleimnek, akik biztosították számomra, hogy legyen elegendő időm tanulni.

Végezetül köszönöm kislányomnak, aki a világ legjobb babája.

Bevezetés

Az egyetemen folytatott tanulmányaim során találkoztam először a Bayes-becsléssel. Már akkor megtetszett a módszer szubjektivitása. Kimondottan érdekesnek találtam azt, hogy az eredményt nagy mértékben befolyásolni tudja az adott dologgal kapcsolatos előzetes ismeretünk. Így szakdolgozatomban kiemelten foglalkoztam a Bayes-becsléshez szükséges a priori eloszlások megválasztásával. Szintén fontosnak tartottam, hogy egyszerű példákon keresztül szemléltessem a módszert, valamint azt is, hogy megmutassam, hogy a mintától függően bizonyos a priori eloszlások megválasztása még könnyebbé teheti a számolást.

A szakdolgozatom második felében a bayesi hipotézisvizsgálattal foglalkoztam. Néhány nevezetes próbát mutattam be részletesen. Az R program segítségével egyszerű példákon keresztül végeztem el a számolásokat, majd az eredményt összevetettem a klasszikus esettel is.

1. Általános fogalmak, jelölések

1.1. Definíció (Feltételes valószínűség). *Ha A, B események és tudjuk, hogy B esemény már bekövetkezett, akkor az A esemény bekövetkezésének valószínűsége $P(A|B)$.*

Ha $P(B) > 0$, akkor

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

1.2. Definíció (Teljes eseményrendszer). *A $B_1, B_2, \dots, B_n$ események teljes eseményrendszert alkotnak, ha $B_i \cap B_j = \emptyset \quad (\forall \ i \neq j) \quad$ és $\quad \bigcup_i B_i = \Omega$.*

1.1. Tétel (Teljes valószínűség tétel). *Ha a $B_1, B_2, \dots, B_n$ események teljes eseményrendszert alkotnak, valamint ismerjük egy A esemény B_i eseményekre vonatkozó feltételes valószínűségeit, akkor az A esemény valószínűsége a következőképpen számolható:*

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

A könnyebb értelmezhetőség érdekében a szakdolgozatomban a következő egységes jelöléseket fogom használni. Legyen

x (vagy y) a minta,

K az x (vagy y) mintán kívüli tudásunk a világról,

θ minden, amit nem tudunk biztosan.

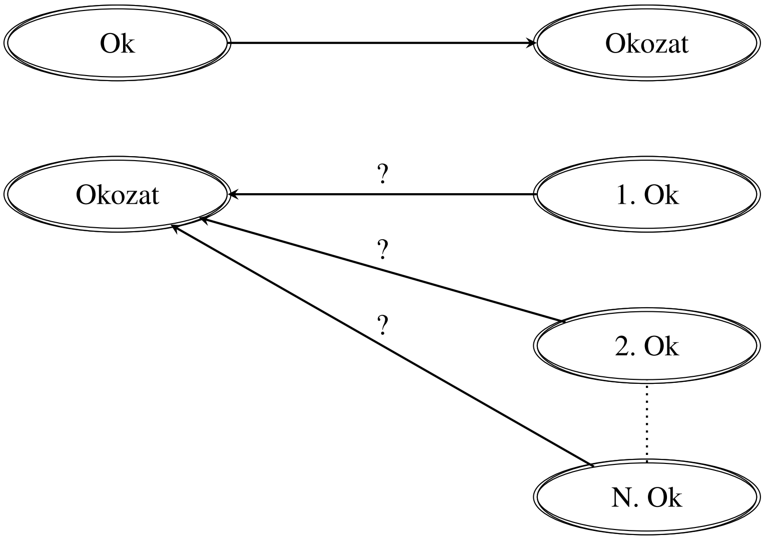
A bayesi statisztikus célja döntést hozni egy számára ismeretlen dologról úgy, hogy közben felhasználja, beépíti az általa ismert tudást. Tehát a $p(\theta|y, K)$ valószínűséget vizsgálja és ezen érték alapján dönt.

2. A Bayes-tétel és változatai

Ebben a fejezetben a Bayes-tétellel fogunk megismerkedni, ugyanis a Bayes-statisztika gyakorlatilag erre az egy tételre, illetve ennek változataira épül. Az olvasóban felmerülhet a kérdés, hogy a Bayes-statisztika mitől másabb a klasszikus statisztikánál? A következőkben erre próbálunk választ adni.

A klasszikus értelemben vett statisztika a "direkt" gondolkodásmódot követi. Legtöbbször megpróbáljuk megjósolni valaminek a következményét. Például milyen irányba mozdul el egy mágnes, ha mellé teszek egy másikat?

Ezzel ellentétben a bayesi statisztika inverz gondolkodásmódú. Az eredményt látva próbálja kikövetkeztetni, hogy mi okozhatta azt. Például mi mozdíthatta el a mágnesemet? Egy másik mágnes? Vagy a szél?



1. ábra. A direkt és inverz gondolkodásmód

2.1. Tétel (Bayes-tétel). *Ha ismert A és B esemény valószínűsége, és $P(A) \neq 0$, $P(B) \neq 0$, illetve $P(A|B)$ feltételes valószínűség, akkor*

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

A $P(B)$ valószínűséget *a priori*, a $P(B|A)$ valószínűséget *a poszteriori* valószínűségnek hívják.

A Bayes-tétel voltaképpen egy képlet, amely segítségével következtetést tudunk levonni a B esemény valószínűségéről, amennyiben az A esemény bekövetkezik. Ha az A esemény a megfigyelt eseményünk és a B eseményre egyfajta hipotézisként tekintünk, akkor Bayes tétele azt mondja meg, hogy amennyiben az A esemény bekövetkezik, az milyen mértékben támasztja alá a hipotézisünk helyességét. A tétel szerint ehhez szükséges az a priori valószínűségek pontos tudása. Azonban a mindennapi életben általában ezek a valószínűségek nem ismertek. Ezért ha a statisztikus például különböző eloszlásokból vesz értékeket, akkor más-más eredményre juthat a végén.

Ha a B nem egyetlen esemény, hanem egy $B_1, B_2, \ldots, B_n$ eseményekből, álló teljes eseményrendszer, akkor a Bayes-tétel a következőképpen fogalmazható meg.

2.2. Tétel (Bayes-tétel teljes eseményrendszerre). *Ha $B_1, B_2, \ldots, B_n$ események egy teljes eseményrendszert alkotnak és $P(B_i) > 0 \quad \forall\, i$-re, akkor*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum\limits_{j=1}^{n} P(A|B_j)P(B_j)}, i = 1, 2, \ldots, n \ \textit{esetén.}$$

2.1. Példa. Tegyük fel, hogy van két érménk. Az egyik szabályos, a másik pedig "súlyozott", azaz 0.75 annak a valószínűsége, hogy fejet dobunk vele, és 0.25 annak a valószínűsége, hogy írást. Tegyük fel, hogy nem tudjuk ránézésre megkülönböztetni egymástól a két érmét. Az egyik érmét találomra kiválasztjuk, és 20-szor dobunk vele. A 20 dobásból 15 alkalommal fejet dobtunk. Mi annak a valószínűsége, hogy a kiválasztott érménk a súlyozott érme volt? Jelölje A, B_1, B_2, a következő eseményeket!

$$\begin{aligned} A &: \text{a kiválasztott érmével 20 dobásból 15-ször fejet dobunk} \\ B_1 &: \text{a súlyozott érmével dobunk} \\ B_2 &: \text{a szabályos érmével dobunk} \end{aligned}$$

Ekkor a B_1 és a B_2 események teljes eseményrendszert alkotnak. A megoldást a Bayes-tételt felhasználva kapjuk (2.2. Tétel).

$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} \\ &= \frac{\binom{20}{15}\, 0.75^{15}\, 0.25^5\, 0.5}{\binom{20}{15}\, 0.75^{15}\, 0.25^5\, 0.5 + \binom{20}{15}\, 0.5^{15}\, 0.5^5\, 0.5} \approx 0.932 \end{aligned}$$

2.1. Definíció (Likelihood függvény). *Legyen $Y_1, Y_2, \ldots, Y_n$ minta egy diszkrét eloszlásból, $y_1, y_2, \ldots, y_n$ pedig a minta realizáció, valamint legyen a mintához tartozó ismeretlen paraméter a θ. Ekkor a minta likelihood-függvénye*

$$p(y|\theta) = P_\theta(Y_1 = y_1, \ldots, Y_n = y_n) = \prod_{i=1}^{n} P_\theta(Y_i = y_i)$$

Ha a mintánk abszolút folytonos, akkor a likelihood-függvény

$$p(y|\theta) = f_\theta(y) = \prod_{i=1}^{n} f_\theta(y_i),$$

ahol $f_\theta(y_i)$ az Y_i mintaelem sűrűségfüggvényét jelöli az y_i helyen.

2.3. Tétel (Teljes valószínűség tétel folytonos esetben). *Tegyük fel, hogy θ egy folytonos paraméter, mely az $[a,\ b]$ intervallumból veszi fel értékeit és sűrűségfüggvénye $p(\theta)$. Az y minta likelihoodja a θ paraméterre $p(y|\theta)$. Ekkor:*

$$p(y) = \int_a^b p(y|\theta)p(\theta)\, d\theta$$

2.4. Tétel (Bayes-tétel folytonos esetben). *Ha θ folytonos paraméter, melynek sűrűségfüggvénye $p(\theta)$ az $[a,\ b]$ intervallumon, az y mintánk likelihoodja $p(y|\theta)$, akkor*

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int_a^b p(y|\theta)p(\theta)\, d\theta}$$

3. A gamma és az inverz gamma eloszlások

A későbbiekben a statisztikusok által leggyakrabban használt nevezetes eloszlások (binomiális, exponenciális, Poisson, ... stb.) mellett, az inverz gamma eloszlást is szeretnénk felhasználni a számításainkban. Egy ritkábban használt eloszlás révén érdemes áttekinteni a tulajdonságait. Ehhez először a gamma eloszlás definícióját nézzük meg, mivel az inverz gamma eloszlás ebből az eloszlásból eredeztethető.

3.1. A gamma eloszlás

A gamma eloszlásnak többféle definíciója is ismert. Nézzük ezek közül a leggyakrabban használtat.

3.1. Definíció. *Az X valószínűségi változó α rendű, β paraméterű gamma eloszlású valószínűségi változó (jelölés: $X \sim Gamma(\alpha, \beta)$), ha sűrűségfüggvénye a következő:*

$$f(x) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha}\, x^{\alpha-1} e^{-\beta x},$$

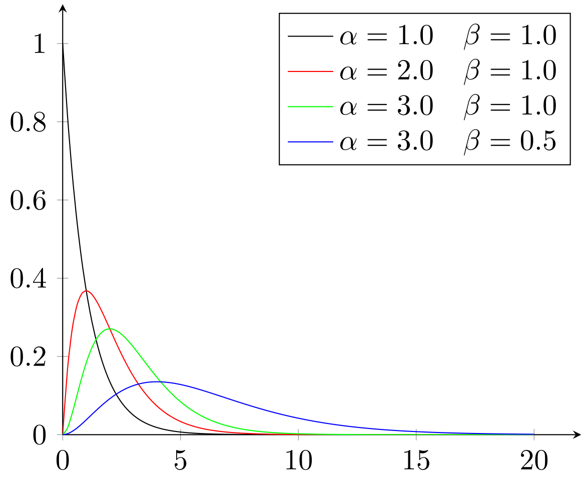
ahol $x > 0$ és $\alpha, \beta > 0$, valamint

$$\Gamma(\alpha) = \int_{0}^{\infty} t^{\alpha-1} e^{-t}\, dt \text{ a gamma függvény.}$$

Ekkor:

$$E(X) = \frac{\alpha}{\beta} \text{ és } D(X) = \frac{\sqrt{\alpha}}{\beta}.$$

A 2. ábra a gamma eloszlás sűrűségfüggvényét szemlélteti néhány α és β érték esetén.



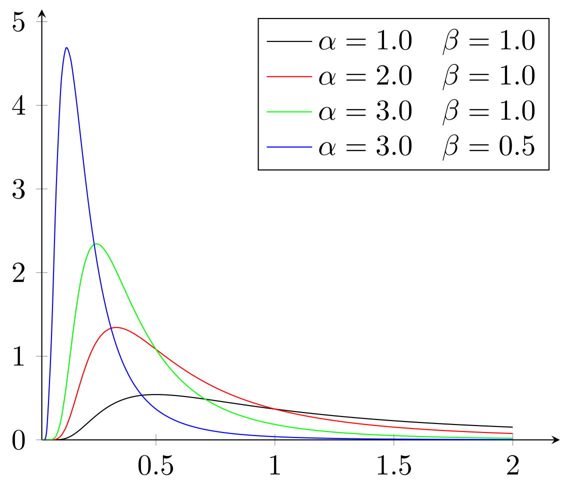
2. ábra. A gamma eloszlás sűrűségfüggvénye

3.2. Az inverz gamma eloszlás

3.2. Definíció. *Az X valószínűségi változó α rendű, β paraméterű inverz gamma eloszlású valószínűségi változó (jelölés: $X \sim IG(\alpha, \beta)$), ha sűrűségfüggvénye a következő:*

$$h(x) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} x^{-\alpha-1} e^{-\frac{\beta}{x}}, \ x > 0 \text{ esetén.}$$

A 3. ábra az inverz gamma eloszlás sűrűségfüggvényét szemlélteti néhány α és β érték esetén.



3. ábra. Az inverz gamma eloszlás sűrűségfüggvénye

Az állítjuk, hogy ha $X \sim Gamma(\alpha, \beta)$, akkor $Y = \frac{1}{X} \sim IG(\alpha, \beta)$. Ugyanis a sűrűségfüggvény transzformációs képletet felhasználva:

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y))\Big|\frac{d}{dy}g^{-1}(y)\Big| = f_X(y^{-1})\Big|\frac{d}{dy}y^{-1}\Big| \\ &= f_X(y^{-1})\Big|-y^{-2}\Big| = \frac{1}{\Gamma(\alpha)}\beta^\alpha y^{-\alpha+1}e^{-\frac{\beta}{y}}y^{-2} \\ &= \frac{1}{\Gamma(\alpha)}\beta^\alpha y^{-\alpha-1}e^{-\frac{\beta}{y}} \end{aligned}$$

ahol $g(y) = \frac{1}{y}$.

3.2.1. Az inverz gamma eloszlás momentumai és szórásnégyzete

A következőkben az $X \sim IG(\alpha, \beta)$ momentumait szeretnénk kiszámolni.
Ha $\alpha > n$, akkor:

$$\begin{aligned} E(X^n) &= \int_0^\infty x^n h(x)\,dx = \int_0^\infty x^n \frac{1}{\Gamma(\alpha)}\beta^\alpha x^{-\alpha-1}e^{-\frac{\beta}{x}}\,dx = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{n-\alpha-1}e^{-\frac{\beta}{x}}\,dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha-n)}{\beta^{\alpha-n}} = \frac{\beta^n\Gamma(\alpha-n)}{(\alpha-1)\ldots(\alpha-n)\Gamma(\alpha-n)} = \frac{\beta^n}{(\alpha-1)\ldots(\alpha-n)} \end{aligned}$$

Az utolsó előtti egyenlőségnél felhasználtuk, hogy

$$ha\ \alpha > 1, akkor\ \Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1).$$

A következőkben a fenti képletet alkalmazva kiszámítjuk az első és a második momentumot, mert ezek ismerete szükséges az eloszlás szórásnégyzetének meghatározásához.

- Az első momentum $(n = 1)$ tehát a következő:

$$E(X) = \frac{\beta}{\alpha-1}, ha\ \alpha > 1.$$

- A második momentum $(n = 2)$ pedig:

$$E(X^2) = \frac{\beta^2}{(\alpha-1)(\alpha-2)},\ ha\ \alpha > 2.$$

Mivel a definíció szerint a szórásnégyzetről tudjuk, hogy:

$$D^2(X) = E(X^2) - E^2(X),$$

ezért ebbe a képletbe behelyettesítve a fenti momentum értékeket, az inverz gamma eloszlás szórásnégyzetét kapjuk α és β paraméterek esetén:

$$D^2(X) = \frac{\beta^2}{(\alpha-1)(\alpha-2)} - \left(\frac{\beta}{\alpha-1}\right)^2 = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)},\ ha\ \alpha > 2.$$

4. A Bayes-becslés[1]

A Bayes-becslés az ismeretlen θ paraméterünk becsléséről szól. Legyen y a mintánk valamilyen θ paraméterű M modellből, ahol θ az ismeretlen paraméter. Ekkor a Bayes-tételt és a teljes valószínűség tételét felhasználva:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)\,d\theta},$$

ahol

$p(\theta|y)$-t a poszteriori eloszlásnak nevezzük,

$p(\theta)$-t a priori eloszlásnak nevezzük,

$p(y|\theta)$ a minta feltételes eloszlása (likelihood), és

$p(y)$ a minta marginális eloszlása.

A poszterior eloszlás összegzi a tudásunkat a nem megfigyelhető paraméterekről, miután a megfigyelhető adatokat kielemeztük.

Ha θ folytonos paraméter, akkor $p(\theta)$ és $p(\theta|y)$ sűrűségfüggvények, illetve ha y folytonos, akkor $p(y)$ és $p(y|\theta)$ is sűrűségfüggvények.

Hasznos rövidítés:

Ha $p(y) = \int p(y|\theta)p(\theta)\,d\theta < \infty$, akkor használhatjuk a

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta) \text{ jelölést,}$$

ahol a $\propto$ azt jelenti, hogy minden olyan tényező, amelyik nem függ θ-tól elhagyható (a normaló konstans elhagyható).

A bayesi megközelítés szerint bármely olyan mennyiség, amelynél bizonytalanak vagyunk a valódi értékét illetően, megfogalmazható eloszlással. A klasszikus esetben azonban nem lehet paraméterek eloszlását megadni, hiszen a paraméterek rögzített mennyiségek. Egyedül a minta véletlenszerű. Így eloszlásokkal csak a mintát lehet kifejezni.

4.1. Az a priori eloszlás megválasztása

A Bayes-becslés lényeges pontja az a priori eloszlás. Ettől a tényezőtől válik szubjektívvé a módszer. Az objektív szemléletet kedvelők főleg emiatt támadják a bayesianusok szemléletét.

A következőkben néhány nevezetes a priori eloszlást fogok bemutatni.

[1] A fejezet a [8] alapján készült.

4.1.1. Konjugált priorok[2]

Az előző fejezetben láttuk, hogy az a poszterior eloszlás az a prior eloszlás és a likelihood szorzataként írható fel:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{1}$$

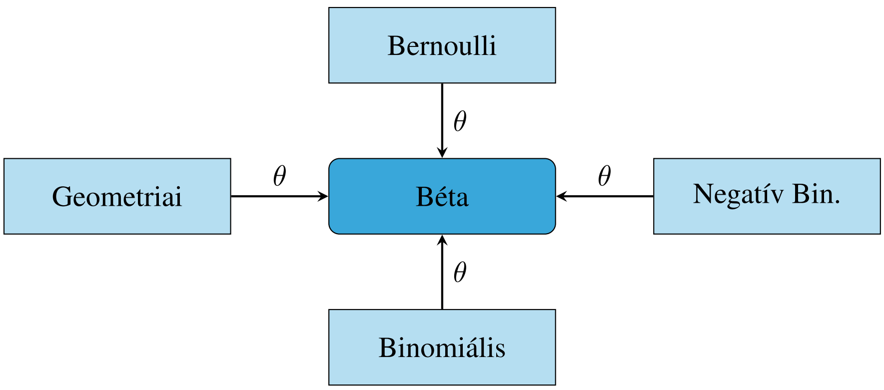
Mivel az a priori eloszlást a statisztikus adja meg, ezért érdemes lehet olyan eloszlást választani, amely formájában megegyezik a likelihood függvénnyel, így az a poszteriori eloszlás függvényformája is ezekkel megegyező lesz. Ez azért érdekes, mert így ugyanabban az eloszláscsaládban maradunk. Azokat az a priori eloszlásokat, amelyek ugyanabba az eloszláscsaládba tartoznak, mint a belőlük számított a poszteriori eloszlások, **konjugált prioroknak** nevezzük. Ilyenkor azt mondjuk, hogy az a prior és a likelihood konjugáltak.

Mivel az a poszteriori és az a priori eloszlásunk is ugyanabból az eloszláscsaládból származnak, ezért csak ezen eloszláscsalád paramétereit kell figyelembe vennünk, ezek hordoznak információt magukban.

A konjugált priorokra jellemző, hogy:

- egy a priori több likelihoodnak is lehet konjugáltja. Például a Béta prior a binomiális és negatív binomiális likelihood konjugáltja is, illetve
- nem mindegyik likelihood függvénynek létezik ismert konjugáltja.

A 4. és 5. ábra néhány nevezetes eloszlást és annak konjugáltjának az a priori eloszlását ismerteti. A nyilak a minta eloszlásából mutatnak a konjugált prior eloszlásra. A nyilakra írt értékek arra a paraméterre vonatkoznak, amelyhez az a priori eloszlást választottuk.



4. ábra. Konjugált likelihoodok Béta prior esetén

A szokásos módon az n elemű minták legyen $x = (x_1, \ldots, x_n)$.

[2] A fejezet a [7] alapján készült.

4.1. Példa. Ha a mintánk *Binomiális*(m, θ) eloszlású, ahol m ismert és θ ismeretlen paraméter, és az a priori eloszlásunk θ-ra a *Béta*(α, β) eloszlás, akkor egy mintaelem sűrűségfüggvénye:

$$p(x|\theta) = \binom{m}{x} \theta^x (1-\theta)^{m-x}.$$

Ekkor az a poszteriori eloszlás:

$$p(\theta|x) \propto p(x|\theta)p(\theta) = \left(\prod_{i=1}^{n} \binom{m}{x_i} \theta^{x_i} (1-\theta)^{m-x_i} \right) \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\propto \theta^{\sum\limits_{i=1}^{n} x_i} (1-\theta)^{nm-\sum\limits_{i=1}^{n} x_i} \theta^{\alpha-1}(1-\theta)^{\beta-1} = \theta^{\sum\limits_{i=1}^{n} x_i+\alpha-1} (1-\theta)^{nm-\sum\limits_{i=1}^{n} x_i+\beta-1}.$$

A fentiekből leolvasható, hogy az a poszteriori eloszlásunk:

$$p(\theta|x) \sim \textit{Béta}\big(\sum_{i=1}^{n} x_i + \alpha, nm - \sum_{i=1}^{n} x_i + \beta\big).$$

4.2. Példa. Ha a mintánk *Bernoulli*(θ) eloszlású, ahol θ ismeretlen paraméter, és a θ paraméterre az a priori eloszlás a *Béta*(α, β) eloszlás, akkor az a poszteriori eloszlás:

$$p(\theta|x) \sim \textit{Béta}\big(\sum_{i=1}^{n} x_i + \alpha, n - \sum_{i=1}^{n} x_i + \beta\big),$$

mert a Bernoulli eloszlás $m = 1$ esetén megegyezik a Binomiális eloszlással.

4.3. Példa. Ha a mintánk *NegatívBinomiális*(r, θ) eloszlású, ahol r ismert és θ ismeretlen paraméter, és a θ paraméterre az a priori eloszlás a *Béta*(α, β) eloszlás, akkor egy mintaelem sűrűségfüggvénye:

$$p(x|\theta) = \binom{r+x-1}{x} \theta^r (1-\theta)^x.$$

Ekkor az a poszteriori eloszlás a következőképpen számolható:

$$p(\theta|x) \propto p(x|\theta)p(\theta) = \left(\prod_{i=1}^{n} \binom{r+x_i-1}{x_i} \theta^r (1-\theta)^{x_i} \right) \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\propto \theta^{nr}(1-\theta)^{\sum\limits_{i=1}^{n} x_i} \theta^{\alpha-1}(1-\theta)^{\beta-1} = \theta^{nr+\alpha-1}(1-\theta)^{\sum\limits_{i=1}^{n} x_i+\beta-1}.$$

Így az a poszteriori eloszlás:

$$p(\theta|x) \sim \textit{Béta}\big(\alpha + nr, \beta + \sum_{i=1}^{n} x_i\big).$$

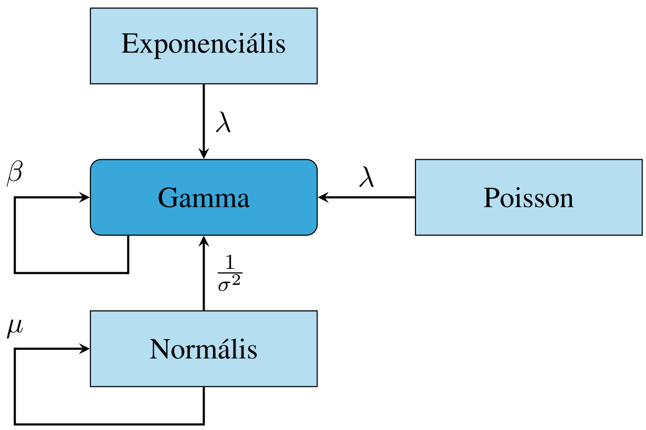
4.4. Példa. Ha a mintánk *Geometriai*(θ) eloszlású, ahol θ ismeretlen paraméter, és az a priori eloszlás a θ paraméterre a *Béta*(α, β) eloszlás, akkor egy mintaelem sűrűségfüggvénye:

$$p(x|\theta) = \theta(1-\theta)^x.$$

Ekkor az a poszteriori eloszlás:

$$p(\theta|x) \sim \textit{Béta}\big(\alpha + n, \beta + \sum_{i=1}^{n} x_i\big).$$

A fentiekben felhasználtuk, hogy a Geometriai eloszlás θ paraméterrel éppen a Negatív Binomiális eloszlás $r = 1$ és θ paraméter esetén.



5. ábra. Konjugált likelihoodok Gamma prior esetén

4.5. Példa. Ha a mintánk gamma eloszlású, ahol a α ismert és β ismeretlen paraméter, és az a priori eloszlásunk a β paraméterre a *Gamma*(α_0, β_0) eloszlás, akkor egy mintaelem sűrűségfüggvénye:

$$p(x|\beta) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} x^{\alpha-1} e^{-\beta x}.$$

Ekkor az a poszteriori eloszlás az alábbiak szerint számolható:

$$p(\beta|x) \propto p(x|\beta)p(\beta) = \left(\prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)} \beta^{\alpha} {x_i}^{\alpha-1} e^{-\beta x_i} \right) \frac{1}{\Gamma(\alpha_0)} {\beta_0}^{\alpha_0} \beta^{\alpha_0 - 1} e^{-\beta_0 \beta}$$

$$\propto \beta^{n\alpha} e^{-\beta \sum\limits_{i=1}^{n} x_i} \beta^{\alpha_0 - 1} e^{-\beta_0 \beta} = \beta^{n\alpha + \alpha_0 - 1} e^{-\beta(\beta_0 + \sum\limits_{i=1}^{n} x_i)}.$$

Így az a poszteriori eloszlás:

$$p(\beta|x) \sim Gamma(\alpha_0 + n\alpha, \beta_0 + \sum_{i=1}^{n} x_i).$$

4.6. Példa. Ha a mintánk *Exponenciális*(λ), ahol λ ismeretlen paraméter, és az a priori eloszlásunk a λ paraméterre a *Gamma*(α_0, β_0) eloszlás, akkor egy mintaelem sűrűségfüggvénye:

$$p(x|\lambda) = \lambda e^{-\lambda x}.$$

Ekkor az a poszteriori eloszlás:

$$p(\lambda|x) \sim Gamma(\alpha_0 + n, \beta_0 + \sum_{i=1}^{n} x_i),$$

mivel az Exponenciális eloszlás éppen a Gamma eloszlás $\alpha = 1, \beta = \lambda$ helyettesítéssel.

4.7. Példa. Ha a mintánk *Poisson*(λ) eloszlású, ahol λ ismeretlen paraméter, és az a priori eloszlásunk a λ paraméterre a *Gamma*(α_0, β_0) eloszlás, akkor egy mintaelem sűrűségfüggvénye:

$$p(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}.$$

Ekkor az a poszteriori eloszlás:

$$p(\lambda|x) \propto p(x|\lambda)p(\lambda) = \left(\prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) \frac{1}{\Gamma(\alpha_0)} {\beta_0}^{\alpha_0} \lambda^{\alpha_0 - 1} e^{-\beta_0 \lambda}$$

$$\propto \lambda^{\sum\limits_{i=1}^{n} x_i} e^{-n\lambda} \lambda^{\alpha_0 - 1} e^{-\beta_0 \lambda} = \lambda^{\alpha_0 - 1 + \sum\limits_{i=1}^{n} x_i} e^{-\lambda(\beta_0 + n)}.$$

A fentiekből leolvasható, hogy:

$$p(\lambda|x) \sim \mathit{Gamma}(\alpha_0 + \sum_{i=1}^{n} x_i, \beta_0 + n).$$

4.8. Példa. Ha a mintánk *Normális*$(\mu, \tau = \frac{1}{\sigma^2})$ eloszlású, ahol μ ismert és τ ismeretlen paraméter, és az a priori eloszlásunk a τ paraméterre *Gamma*(α, β) akkor egy mintaelem sűrűségfüggvénye:

$$p(x|\tau) = \left(\frac{\tau}{2\pi} \right)^{\frac{1}{2}} exp\left(- \frac{\tau(x-\mu)^2}{2} \right).$$

Ekkor az a poszteriori eloszlás:

$$p(\tau|x) \propto p(x|\tau)p(\tau) = \left(\prod_{i=1}^{n} \left(\frac{\tau}{2\pi} \right)^{\frac{1}{2}} exp\left(- \frac{\tau(x_i-\mu)^2}{2} \right) \right) \frac{1}{\Gamma(\alpha)} \beta^{\alpha} \tau^{\alpha-1} e^{-\beta\tau}$$

$$\propto \tau^{\frac{n}{2}} exp\left(- \frac{\tau}{2} \left(\sum_{i=1}^{n} (x_i - \mu)^2 \right) - \beta\tau \right) \tau^{\alpha-1}$$

$$\propto \tau^{\frac{n}{2} + \alpha - 1} exp\left(-\tau \left(\frac{1}{2} S^2 (n-1) + \beta \right) \right).$$

Így:

$$p(\tau|x) \sim Gamma\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}(n-1)S^2 \right),$$

ahol

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2.$$

4.9. Példa. Ha a mintánk *Normális*$(\mu, \tau = \frac{1}{\sigma^2})$ eloszlású, ahol τ ismert és μ ismeretlen paraméter, valamint az a priori eloszlásunk μ-re a *Normális*(μ_0, τ_0) eloszlás, akkor egy mintaelem sűrűségfüggvénye:

$$p(x|\mu) = \Big(\frac{\tau}{2\pi}\Big)^{\frac{1}{2}} exp\Big(-\frac{\tau(x-\mu)^2}{2}\Big).$$

Ekkor az a poszteriori eloszlást a következőképpen tudjuk számolni:

$$p(\mu|x) \propto p(x|\mu)p(\mu) =$$

$$= \Big(\prod_{i=1}^{n}\Big(\frac{\tau}{2\pi}\Big)^{\frac{1}{2}} exp\Big(-\frac{1}{2}\tau(x_i-\mu)^2\Big)\Big)\Big(\frac{\tau_0}{2\pi}\Big)^{\frac{1}{2}} exp\Big(-\frac{1}{2}\tau_0(\mu-\mu_0)^2\Big)$$

$$= \Big(\frac{\tau}{2\pi}\Big)^{\frac{n}{2}}\Big(\frac{\tau_0}{2\pi}\Big)^{\frac{1}{2}} exp\Big(-\frac{1}{2}(\sum_{i=1}^{n}\tau(\mu-x_i)^2 + \tau_0(\mu-\mu_0)^2)\Big)$$

$$= A\ exp\Big(-\frac{1}{2}\Big(n\tau\mu^2 - 2\mu\tau\sum x_i + \tau\sum x_i^2 + \tau_0\mu^2 - 2\mu\mu_0\tau_0 + \tau_0\mu_0^2\Big)\Big)$$

$$= A\ exp\Big(-\frac{1}{2}\Big(\mu^2(n\tau+\tau_0) - 2\mu\Big(\tau\sum x_i + \mu_0\tau_0\Big) + \Big(\tau\sum x_i^2 + \tau_0\mu_0^2\Big)\Big)$$

$$= A\ exp\Big(-\frac{1}{2}\Big[(\tau_0+n\tau)\Big(\mu^2 - 2\mu\frac{\tau\sum x_i + \mu_0\tau_0}{\tau_0+n\tau}\Big)\Big] + \Big(\tau\sum x_i^2 + \tau_0\mu_0^2\Big)\Big)$$

$$= A\ exp\Big(-\frac{1}{2}\Big((\tau_0+n\tau)\Big((\mu - \frac{\tau\sum x_i + \mu_0\tau_0}{\tau_0+n\tau})^2 - (\frac{\tau\sum x_i + \mu_0\tau_0}{\tau_0+n\tau})^2\Big)$$

$$+ (\tau\sum x_i^2 + \tau_0\mu_0^2)\Big)\Big)$$

$$= A\ exp\Big(-\frac{1}{2}\Big((\tau_0+n\tau)\Big(\mu - \frac{\tau\sum x_i + \mu_0\tau_0}{\tau_0+n\tau}\Big)^2\Big)\Big)$$

$$* exp\Big(-\frac{1}{2}\Big(-\frac{(\tau\sum x_i + \mu_0\tau_0)^2}{\tau_0+n\tau} + (\tau\sum x_i^2 + \tau_0\mu_0^2)\Big)\Big)$$

$$= A\ B\ exp\Big(-\frac{1}{2}\Big((\tau_0+n\tau)\Big(\mu - \frac{\tau\sum x_i + \mu_0\tau_0}{\tau_0+n\tau}\Big)^2\Big)\Big).$$

A kapott eredményből leolvasható, hogy

$$p(\mu|x) \sim Normális\Big(\frac{\mu_0\tau_0 + \tau\sum_{i=1}^{n} x_i}{\tau_0+n\tau}, \tau_0 + n\tau\Big).$$

4.1.2. Improprius priorok[3]

A sűrűségfüggvényekről tudjuk, hogy nemnegatívak és integráljuk a $-\infty$ és $+\infty$ között 1. **Improprius prioroknak** nevezzük azokat az a priorikat, amelyek ugyan nemnegatívak, de integ-

[3] A fejezet az [5] alapján készült.

ráljuk ∞. Vegyük például az $y = (Y_1, \dots, Y_n)$ mintát, melyre $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$, függetlenek, és tegyük fel, hogy σ^2 ismert, μ ismeretlen paraméter.

Ha nem rendelkezünk előzetes ismeretekkel a μ paraméterről, akkor kézenfekvőnek tűnhet a $p(\mu) = c > 0$ választás. Ekkor:

$$\int_{-\infty}^{\infty} p(\mu)\, d\mu = \int_{-\infty}^{\infty} c\, d\mu = \infty.$$

Tehát ez az a priori improprius. Előfordulhat olyan eset, hogy habár az a priorink improprius, mégis ezzel továbbszámolva, az a poszteriori eloszlásnak egy valódi eloszlást kapnánk.

4.10. Példa. Legyen

$$Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2),$$

ahol μ ismeretlen σ^2 ismert paraméter, és legyen az a priori eloszlásunk μ-re a konstans c. Az előbbiekben láttuk, hogy ez egy improprius a priori. Ekkor az a poszteriori eloszlást az alábbiak szerint tudjuk számolni.

$$\begin{aligned}
p(\mu|y) &\propto p(y|\mu)p(\mu) \\
&= \Bigg(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp\Big(-\frac{1}{2}\sigma^{-2}(y_i-\mu)^2\Big)\Bigg)c \\
&= c\left(\frac{1}{\sqrt{2\pi}}\right)^n \sigma^{-n} exp\Big(-\frac{1}{2}\sum \sigma^{-2}(\mu-y_i)^2\Big) \\
&= A\ exp\Big(-\frac{1}{2}\Big(\frac{n}{\sigma^2}(\mu^2-2\mu\bar{y}) + \frac{1}{\sigma^2}\sum y_i^2\Big)\Big) \\
&= A\ exp\Big(-\frac{1}{2}\Big(\frac{n}{\sigma^2}(\mu-\bar{y})^2 - \frac{n}{\sigma^2}\bar{y}^2 + \frac{1}{\sigma^2}\sum y_i^2\Big)\Big) \\
&= A\ exp\Big(-\frac{1}{2}\Big(\frac{n}{\sigma^2}(\mu-\bar{y})^2\Big)\ exp\Big(-\frac{1}{2}\Big(-\frac{n}{\sigma^2}\bar{y}^2 + \frac{1}{\sigma^2}\sum y_i^2\Big)\Big) \\
&= A\ B\ exp\Big(-\frac{1}{2}\Big(\frac{n}{\sigma^2}(\mu-\bar{y})^2\Big)\Big)
\end{aligned}$$

Az eredményből már leolvasható, hogy

$$\mu|y \sim N\Big(\bar{y}, \frac{\sigma^2}{n}\Big).$$

Azaz az a poszterior eloszlásunk egy valódi eloszlás.

Egy nevezetes improprius a priori **Haldane a priorija**:

$$p(\theta) \propto \frac{1}{\theta(1-\theta)} = \theta^{-1}(1-\theta)^{-1} \sim \textit{"Béta(0,0)"},$$

ahol θ a [0,1] intervallumból vett paraméter. Ez valóban egy improprius a priori, mert a sűrűségfüggvényének integrálja a $[0,\ 1]$ intervallumon végtelen.

$$\int_0^1 \frac{1}{\theta(1-\theta)}\, d\theta = \infty$$

4.1.3. Neminformatív priorok[4]

Előfordulhat olyan eset, amikor a statisztikus nem tud vagy szándékosan nem akar megválasztani bármilyen információt hordozó a priori eloszlást, hogy az semmilyen formában ne tudja befolyásolni az a poszteriori eloszlást (például kis elemszámú minta esetén).

Jeffreys feltevése szerint egy "igazi" a priorinak invariánsnak kell lennie az átparaméterezésre. Ez a szükséges feltétele annak, hogy a priorink neminformatív legyen. Azaz, ha $p(\theta)$ neminformatív, akkor az a priori bármilyen átparaméterezésének is neminformatívnak kell lennie. Például, ha egy statisztikus a σ paraméterre választ egy $p_1(\sigma)$ a priori eloszlást, és egy másik statisztikus σ^2 paraméterre ugyanazt az a priori eloszlást $\big(p_2(\sigma^2)\big)$, és a két statisztikus a számolás után különböző eredményre jut, akkor ez az eljárás szubjektívnek tekinthető, mert függ az átparaméterezéstől.

Jefferys priorja egy dimenzióban a következő:

$$p(\theta) \propto \sqrt{I(\theta)},$$

ahol $I(\theta)$ az egy elemű minta Fisher-információja, azaz

$$\mathbb{E}_\theta\Bigg(\bigg(\frac{dlog(p(y|\theta))}{d\theta}\bigg)^2\Bigg).$$

Ismeretes, hogy bizonyos regularitási feltételek mellett

$$-\mathbb{E}_\theta\bigg(\frac{d^2log(p(y|\theta))}{d\theta^2}\bigg) = \mathbb{E}_\theta\Bigg(\bigg(\frac{dlog(p(y|\theta))}{d\theta}\bigg)^2\Bigg).$$

Ez az állítás nem tartozik szorosan a szakdolgozat témájához, ezért bizonyítás nélkül fogunk a továbbiakban számolni vele.

A következőkben lássuk be, hogy Jeffreys priorja invariáns az átparaméterezésre. Tegyük fel, hogy az első statisztikus Jeffreys priorját használja a θ paraméterre, az a priorit jelöljük $p_1(\theta)$-val. A második statisztikus szintén Jeffreys priorját használja $\tau = h(\theta)$ esetén (θ egy átparaméterezése), jelöljük az a priorit $p_2(\tau)$-val. A sűrűségfüggvény transzformációs képletet felhasználva:

$$p_3(\tau) = p_1(h^{-1}(\tau))\bigg|\frac{dh^{-1}(\tau)}{d\tau}\bigg|.$$

[4] A fejezet a [11] alapján készült.

Belátjuk, hogy $p_2(\tau) = p_3(\tau)$ vagyis, hogy Jeffreys priorja a τ paraméterre megegyezik azzal, mintha Jeffreys priorját a θ paraméterre alkalmaztuk volna:

$$p_3(\tau) = p_1(\theta)\left|\frac{d\theta}{d\tau}\right| \propto \sqrt{I(\theta)\left(\frac{d\theta}{d\tau}\right)^2} = \sqrt{\mathbb{E}\left[\left(\frac{dlog(p(y|\theta))}{d\theta}\right)^2\right]\left(\frac{d\theta}{d\tau}\right)^2}$$

$$= \sqrt{\mathbb{E}\left[\left(\frac{dlog(p(y|\theta))}{d\theta}\frac{d\theta}{d\tau}\right)^2\right]} = \sqrt{\mathbb{E}\left[\left(\frac{dlog(p(y|\tau))}{d\tau}\right)^2\right]} = \sqrt{I(\tau)} = p_2(\tau).$$

Ebben az értelemben az a priorink neminformatív lesz, mivel mindkét statisztikus ugyanazzal a likelihoodal számol, ugyanazt az a priorit veszik, így megegyező a poszteriorira jutnak. Jeffreys priorjának egyik fontos tulajdonsága, hogy nem függ magától a mintától, mivel már a Fisher információ sem függött attól. (Vannak olyan esetek, amikor az a priori eloszlást a mintától függően érdemes megválasztanunk, a minta eloszlásának mutatóit felhasználva, amennyiben van sejtésünk az értékükről. A későbbiekben látni fogunk erre egy példát is.) Jeffreys priorja gyakran impropríus, ezért érdemes meggyőződni arról, hogy az a poszteriori valóban létező eloszlás, amely sűrűségfüggvényének integrálja mínusz végtelen és plusz végtelen között 1.

Jeffreys priorja többdimenzióban

Jefferys priorja a $\theta = (\theta_1, \ldots, \theta_n)$ paraméterekre:

$$p(\theta) \propto \sqrt{det(I(\theta))},$$

ahol $I(\theta)$ egy $n \times n$-s Fisher-információs mátrix.
A mátrix i-edik sorának j-edik eleme:

$$-\mathbb{E}\left(\frac{d^2 logp(y|\theta)}{d\theta_i\, d\theta_j}\right).$$

4.11. Példa. Jeffreys priorjának kiszámítása a Bernoulli eloszlás esetén

Legyen Y a Bernoulli eloszlásból vett egyelemű mintánk θ paraméterrel. Ekkor $\mathbb{E}(Y) = \theta$ és

$$p(Y|\theta) = \theta^Y(1-\theta)^{1-Y}$$

$$logp(Y|\theta) = log(\theta^Y(1-\theta)^{1-Y}) = Ylog\theta + (1-Y)log(1-\theta)$$

$$\frac{d}{d\theta}logp(Y|\theta) = Y\frac{1}{\theta} + (1-Y)\frac{1}{1-\theta}(-1) = \frac{Y}{\theta} - \frac{1-Y}{1-\theta}$$

$$-\frac{d^2}{d\theta^2}logp(Y|\theta) = \frac{Y}{\theta^2} + \frac{1-Y}{(1-\theta)^2}$$

Mivel $\mathbb{E}(Y) = \theta$, így a Fisher-információ:

$$I(\theta) = \mathbb{E}\Big[-\frac{d^2}{d\theta^2} logp(Y|\theta)\Big] = \mathbb{E}\Big[\frac{Y}{\theta^2} + \frac{1-Y}{(1-\theta)^2}\Big] = \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta(1-\theta)}.$$

Így Jeffreys priorja:

$$p(\theta) \propto \sqrt{I(\theta)} = \sqrt{\frac{1}{\theta(1-\theta)}},$$

azaz a θ a priorija a Béta eloszlás $\alpha = \frac{1}{2}$ $\beta = \frac{1}{2}$ paraméterekkel.

4.2. A Bayes-becslés

Láttuk, hogy amennyiben meghatároztuk az a priori eloszlást, az (1)-es képlet segítségével az a poszteriori eloszlást is ki tudjuk számolni. A kapott eredmény egy "majdnem" sűrűségfüggvény (a konstansoktól eltekintve). Általában végeredményül nem a paraméter eloszlását szokás megadni, hanem az abból számított valamilyen értéket.

Egy úgynevezett veszteségfüggvényt határozunk meg, melynek várható értékét szeretnénk minimalizálni. Jelölje a veszteségfüggvényt $L(\theta, \hat{\theta})$, ahol θ a paraméterünk és $\hat{\theta}$ a paraméterünk becslése. Az eltérést többféleképpen is értelmezhetjük. Például, ha a veszteségfüggvényünk négyzetes ($L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$), akkor az a poszteriori eloszlás várható értékét szokás a becslés eredményéül adni. Ha a veszteségfüggvény abszolút értékű ($L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$), akkor a medián az eredmény. Egy másik esetben úgy is gondolkodhatunk, hogy a veszteségfüggvényünk 0-t vesz fel, ha a paraméterünk valódi értéke megegyezik a kapott értékkel, és egy konstans számot (például 1-et), ha nem. Ekkor a becslésünk eredménye az a poszteriori eloszlásunk módusza.

Veszteségfüggvény	A poszteriori mutató
Lineáris	Medián
Négyzetes	Várható érték
0/1	Módusz

4.12. Példa. Legyen $y = (y_1, \dots, y_n)$, $y|\theta \sim Exponencilis(\theta)$, és az a priori eloszlás θ-ra Gamma(α, β). Korábban láttuk, hogy ekkor az a poszteriori eloszlás:

$$\theta|y \sim Gamma(\alpha + n, \beta + \sum_{i=1}^{n} y_i).$$

Így ha a veszteségfüggvényünk négyzetes, akkor a Bayes-becslésünk:

$$\hat{\theta} = E(\theta|y) = \frac{\alpha + n}{\beta + \sum_{i=1}^{n} y_i}.$$

Vegyük észre, hogy $\frac{n}{\sum_{i=1}^{n} y_i}$ éppen az exponenciális eloszlású minta θ ismeretlen paraméterének maximum likelihood becslése, míg $\frac{\alpha}{\beta}$ az α és β paraméterű Gamma eloszlás várható értéke. Ha n értéke kicsi, akkor a Bayes-becslés inkább az a priori várható érték, ha pedig n értéke nagy, akkor a Bayes-becslés a maximum likelihood becsléshez van közel.

4.13. Példa. Legyen a mintánk $y = (y_1, \ldots, y_n)$, $y_i|\theta \sim Geometriai(\theta)$, és az a priori eloszlás θ-ra Béta(α, β). Ekkor az a poszteriori eloszlás:

$$\theta|y \sim \textit{\textbf{Béta}}(\alpha + n, \beta - n + \sum_{i=1}^{n} y_i).$$

Így ha a veszteségfüggvényünk négyzetes, akkor a Bayes-becslésünk:

$$\hat{\theta} = E(\theta|y) = \frac{\alpha + n}{\alpha + n + \beta - n + \sum_{i=1}^{n} y_i} = \frac{\alpha + n}{\alpha + \beta + \sum_{i=1}^{n} y_i}.$$

Hasonlóan megjegyezhető, hogy $\frac{n}{\sum_{i=1}^{n} y_i}$ éppen a geometriai eloszlású minta θ ismeretlen paraméterének maximum likelihood becslése, míg $\frac{\alpha}{\alpha+\beta}$ az α és β paraméterű Béta eloszlás várható értéke.

4.3. A Bayes-becslés előnye és hátránya[5]

A bayesi gondolkodás egyik legfontosabb ismérve, hogy ellentétben az objektív klasszikus módszerekkel, a Bayes-becslés erősen függ a már ismert tudásunktól. Az a priori eloszlást ugyanis a statisztikus választja. Ez pedig attól függően, hogy az elemzőnek milyen korábbi ismeretei vannak az adott dologról, akár lényegesen is megváltoztathatja az a poszteriori eloszlást.

Amennyiben újabb minták elérhetők, úgy a már meglévő ismereteinket folyamatosan tudjuk bővíteni a Bayes-becslés segítségével. Ezt a következőképpen tehetjük meg.

Legyen $p(\theta)$ az a priori eloszlásunk (ezt mi választjuk meg), és legyen x_1 egy tetszőleges minta. Ekkor az a poszteriori eloszlás:

$$p(\theta|x_1) = \frac{p(x_1|\theta)p(\theta)}{p(x_1)} \propto p(x_1|\theta)p(\theta).$$

Most vegyük az új, x_2 mintát, amely az x_1 mintánktól független. Legyen az előzőekben kiszámolt a poszteriori eloszlásunk az új a priori eloszlásunk. Így egy új a poszteriori valószínűséget készíthetünk:

$$p(\theta|x_1, x_2) \propto p(x_2|\theta)p(\theta|x_1) \propto p(x_2|\theta)p(x_1|\theta)p(\theta) = p(\theta)p(x_1, x_2|\theta).$$

A fentiekben felhasználtuk, hogy mivel az x_1 és x_2 mintáink függetlenek ezért az együttes sűrűségfüggvényük szorzattá bomlik, azaz: $p(x_1, x_2|\theta) = p(x_1|\theta)p(x_2|\theta)$.

[5] A fejezet a [2] alapján készült.

Ha x_1 és x_2 mintáink már a kezdetben, egyszerre rendelkezésre álltak volna, akkor ugyanezt az a poszteriori eloszlást kaptuk volna meg.

A folyamatot újabb és újabb minta érkezése esetén a végtelenségig lehet ismételni. Tehát folyamatosan tudjuk frissíteni az a priori eloszlásunkat:

$$p(\theta) \rightarrow p(\theta|x_1) \rightarrow p(\theta|x_1, x_2) \rightarrow \ldots \rightarrow p(\theta|x_1, x_2, \ldots, x_n)$$

A szakirodalom ezt a módszert *bayesian update*-nek nevezi.

4.4. A normális eloszlás bayesi elemzése

Attól függően, hogy a normális eloszlás mely paraméterei ismertek és ismeretlenek, három különböző esetet tudunk megkülönböztetni:

1. Ha a várható érték ismeretlen, a szórás ismert
2. Ha a várható érték ismert, a szórás ismeretlen
3. Ha a várható érték és a szórás is ismeretlen

4.4.1. A várható érték ismeretlen, a szórás ismert[6]

A 4.9 Példánál láttuk, hogy ha a várható érték ismeretlen és a szórás ismert paraméterek, valamint az a priori eloszlás normális eloszlás, akkor az a poszteriori eloszlás is normális eloszlás lesz.

4.4.2. A várható érték ismert, a szórás ismeretlen

Legyenek $Y_1, Y_2, \ldots, Y_n \sim N(\mu, \sigma^2)$ független valószínűségi változók, ahol a μ ismert, a σ^2 ismeretlen paraméter. Ekkor a valószínűségi változók együttes likelihoodja a következő:

$$p(y|\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right).$$

Válasszuk σ^2 konjugált priorját az a priori eloszlásnak, azaz:

$$\sigma^2 \sim IG(\alpha, \beta).$$

így

$$p(\sigma^2) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} (\sigma^2)^{-(\alpha+1)} exp\left(-\frac{\beta}{\sigma^2}\right),$$

[6]A fejezet a [2] alapján készült.

ahol $\sigma^2 > 0$ és $\alpha, \beta > 0$. Az a poszteriori eloszlás a következőképpen számolható:

$$\begin{aligned}
p(\sigma^2|y) &\propto p(y|\sigma^2)p(\sigma^2) \\
&= \left[\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} exp\bigg(-\frac{(y_i-\mu)^2}{2\sigma^2}\bigg)\right] \frac{1}{\Gamma(\alpha)} \beta^{\alpha} (\sigma^2)^{-(\alpha+1)} exp\bigg(-\frac{\beta}{\sigma^2}\bigg) \\
&\propto (\sigma^2)^{-\frac{n}{2}} exp\bigg(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu)^2\right]\bigg) \frac{1}{\Gamma(\alpha)} \beta^{\alpha} (\sigma^2)^{-(\alpha+1)} exp\bigg(-\frac{\beta}{\sigma^2}\bigg) \\
&\propto (\sigma^2)^{-\frac{n}{2}} exp\bigg(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\bigg) (\sigma^2)^{-(\alpha+1)} exp\bigg(-\frac{\beta}{\sigma^2}\bigg) \\
&= (\sigma^2)^{-(\alpha+\frac{n}{2}+1)} exp\bigg(-\frac{\beta+\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2}{\sigma^2}\bigg).
\end{aligned}$$

A fentiből látszik, hogy:

$$\sigma^2|y \sim IG\bigg(\alpha+\frac{n}{2}, \beta+\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2\bigg).$$

Megjegyzés

Hogy érdemes megválasztanunk az a priori eloszlás α és β paramétereit?

Az inverz gamma eloszlásnál láttuk, hogy a várható érték és a szórásnégyzet:

$$E(\sigma^2) = \tfrac{\beta}{\alpha-1}, \text{ ha } \alpha > 1.$$

$$D^2(\sigma^2) = \tfrac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \text{ ha } \alpha > 2.$$

Fejezzük ki α-t és β-t a fenti két egyenletből!

$$(E(\sigma^2))^2 = (\alpha-2)D^2(\sigma^2) = (\alpha-2)\tfrac{\beta^2}{(\alpha-1)^2(\alpha-2)} = \tfrac{\beta^2}{(\alpha-1)^2}$$

Ebből:

$$\alpha = \tfrac{(E(\sigma^2))^2}{D^2(\sigma^2)} + 2.$$

$$\beta = E(\sigma^2)(\alpha-1) = E(\sigma^2)\bigg(\tfrac{(E(\sigma^2))^2}{D^2(\sigma^2)} + 2 - 1\bigg) = E(\sigma^2)\bigg(\tfrac{(E(\sigma^2))^2}{D^2(\sigma^2)} + 1\bigg).$$

Ez azt jelenti, hogy ha a σ^2 várható értékére és szórására van valamilyen megérzésünk akár a mintából adódóan, akkor azokból kifejezhetjük α-t és β-t.

4.4.3. A várható érték és a szórás is ismeretlen[7]

Tegyük fel, hogy a mintánk normális eloszlású, ahol μ és σ^2 ismeretlen paraméterek.

$$x = (x_1, \dots, x_n) \sim N(\mu, \sigma^2 = \tau)$$

Így a likelihood a következőképpen számolható:

$$\begin{aligned}
p(x|\mu,\tau) &= \prod_{i=1}^{n} (2\pi\tau)^{-\frac{1}{2}} exp\Big(- \frac{(x_i-\mu)^2}{2\tau}\Big) \\
&= \big(2\pi\big)^{-\frac{n}{2}} \tau^{-\frac{n}{2}} exp\Bigg(- \frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{2\tau}\Bigg) \\
&= \big(2\pi\big)^{-\frac{n}{2}} \tau^{-\frac{n}{2}} exp\Big(- \frac{1}{2\tau}(\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2)\Big) \\
&= \big(2\pi\big)^{-\frac{n}{2}} \tau^{-\frac{n}{2}} exp\Big(- \frac{1}{2\tau}(S^2 + n(\bar{x}-\mu)^2)\Big),
\end{aligned}$$

ahol $S^2 := \sum\limits_{i=1}^{n} (x_i - \bar{x})^2$.

Válasszuk az a priori eloszlásoknak Jeffreys a priorjait:

$$\begin{aligned}
p(\mu) &\propto 1, \\
p(\tau) &\propto \frac{1}{\tau}, \\
p(\mu,\tau) &\propto \frac{1}{\tau}.
\end{aligned}$$

Így az a poszteriori:

$$\begin{aligned}
p(\mu,\tau|x) &\propto p(\mu,\tau)p(x|\mu,\tau) \\
&= \tau^{-1}\tau^{-\frac{n}{2}} exp\Big(- \frac{1}{2\tau}(S^2 + n(\bar{x}-\mu)^2)\Big) \\
&= \tau^{-\frac{n}{2}-1} exp\Big(- \frac{1}{2\tau}(S^2 + n(\bar{x}-\mu)^2)\Big).
\end{aligned}$$

A későbbi elemzés miatt hasznos lehet, ha n helyett a szabadsági fokra, v-re fejezzük ki az a poszteriorit. Mivel $v = n - 1$, ezért:

$$p(\mu,\tau|x) \propto \tau^{-\frac{v+1}{2}-1} exp\Big(- \frac{1}{2\tau}(S^2 + n(\bar{x}-\mu)^2)\Big).$$

A várható érték a poszteriori eloszlása

A következőkben μ a poszteriori eloszlását szeretnénk kiszámolni. Ismeretes, hogy az együttes

[7] A fejezet a [4] alapján készült.

sűrűségfüggvény kiintegrálva az egyik változó szerint, a másik változó sűrűségfüggvényét kapjuk eredményül. Vagyis:

$$p(\mu|x) = \int p(\mu, \tau|x)\, d\tau \propto \int_0^\infty \tau^{-\frac{n}{2}-1} exp\bigg(-\frac{1}{2\tau}(S^2 + n(\bar{x}-\mu)^2)\bigg)\, d\tau. \tag{2}$$

Alkalmazzuk az $u = \frac{1}{2\tau}(S^2 + n(\bar{x}-\mu)^2)$ helyettesítést.

$$u = \frac{1}{2\tau}(S^2 + n(\bar{x}-\mu)^2) = \frac{1}{2\tau}B$$

$$\tau = \frac{1}{2u}B = (2u)^{-1}B$$

$$\frac{d\tau}{du} = -2^{-1}u^{-2}B$$

Ekkor a (2)-es egyenlet a következőképpen alakítható át.

$$\begin{aligned} &\int_\infty^0 ((2u)^{-1}B)^{-\frac{n}{2}-1} exp(-u)(-2^{-1}u^{-2}B)\, du \\ &= 2^{\frac{n}{2}} B^{-\frac{n}{2}} \int_0^\infty u^{\frac{n}{2}-1} exp(-u)\, du \\ &= 2^{\frac{n}{2}} B^{-\frac{n}{2}}\, \Gamma(\frac{n}{2}) \\ &\propto B^{-\frac{n}{2}} = (S^2 + n(\bar{x}-\mu)^2)^{-\frac{n}{2}}. \end{aligned}$$

Tehát a μ a poszteriori eloszlása:

$$p(\mu|x) \propto (S^2 + n(\bar{x}-\mu)^2)^{-\frac{n}{2}} = (S^2 + n(\bar{x}-\mu)^2)^{-\frac{v+1}{2}}.$$

Az eredményünk nem egy ismert eloszlás. Azonban, ha megfelelő helyettesítéseket alkalmazunk, akkor $\mu|x$ eloszlását kifejezhetjük ismert eloszlással is. Legyen

$$t = \frac{\bar{x}-\mu}{s}\sqrt{n},$$
$$s^2 = \frac{S^2}{n-1} = \frac{S^2}{v},$$

ahol s^2 a minta korrigált tapasztalati szórásnégyzete.

Helyettesítsük a μ-t t-vel. Ekkor a t a poszteriori sűrűségfüggvényét a következő levezetésből kapjuk:

$$\begin{aligned} p(t|x) &\propto (vs^2 + (ts)^2)^{-\frac{v+1}{2}} = \bigg[\frac{v(vs^2+(ts)^2)}{v}\bigg]^{-\frac{v+1}{2}} = \bigg[\frac{vs^2(v+t^2)}{v}\bigg]^{-\frac{v+1}{2}} \\ &\propto \bigg(1+\frac{t^2}{v}\bigg)^{-\frac{v+1}{2}}. \end{aligned}$$

Vegyük észre, hogy $t|x$ eloszlása éppen a t-eloszlás v szabadsági fokkal, azaz $t|x \sim t_v$.

Mivel $\mu = \bar{x} - \frac{ts}{\sqrt{n}}$, így a μ Bayes-becslése:

$$\hat{\mu} = \bar{x}$$

A szórásnégyzet a poszteriori eloszlása

A szórásnégyzet a poszteriori eloszlását az előbbiekhez hasonlóan tudjuk számolni. Most a τ helyett a μ változó szerint kell kiintegráljuk a két paraméter a poszteriori eloszlását.

$$p(\tau|x) = \int p(\mu, \tau|x)\, d\mu \propto \int_{-\infty}^{\infty} \tau^{-\frac{n}{2}-1} exp\Big(-\frac{1}{2\tau}(S^2 + n(\bar{x}-\mu)^2)\Big)\, d\mu$$

$$= \tau^{-\frac{n}{2}-1} \int_{-\infty}^{\infty} exp\Big(-\frac{1}{2\tau}S^2 - \frac{1}{2\tau}(n(\bar{x}-\mu)^2)\Big)\, d\mu$$

$$= \tau^{-\frac{n}{2}-1} exp\Big(-\frac{1}{2\tau}S^2\Big) \int_{-\infty}^{\infty} exp\Big(-\frac{1}{2\tau}(n(\bar{x}-\mu)^2)\Big)\, d\mu$$

$$= \tau^{-\frac{n}{2}-1} exp\Big(-\frac{1}{2\tau}S^2\Big) \Big(\frac{2\tau\pi}{n}\Big)^{\frac{1}{2}} \int_{-\infty}^{\infty} \Big(\frac{2\tau\pi}{n}\Big)^{-\frac{1}{2}} exp\Big(-\frac{1}{2\tau}(n(\bar{x}-\mu)^2)\Big)\, d\mu$$

$$= \tau^{-\frac{v+1}{2}-1} exp\Big(-\frac{1}{2\tau}S^2\Big) \Big(\frac{2\tau\pi}{n}\Big)^{\frac{1}{2}} \; 1$$

$$\propto \tau^{-\frac{v}{2}-1} exp\Big(-\frac{1}{2\tau}S^2\Big)$$

Innen leolvasható, hogy az a poszteriori eloszlás inverz gamma eloszlású lesz $\frac{v}{2}$ és $\frac{S^2}{2}$ paraméterekkel.

$$\tau|x \sim IG\Big(\frac{v}{2}, \frac{S^2}{2}\Big)$$

Ekkor τ Bayes-becslése:

$$\hat{\tau} = \frac{\frac{S^2}{2}}{\frac{v}{2}-1} = \frac{S^2}{v-2} = \frac{S^2}{n-3}.$$

Mivel a τ ML-becslése $\frac{S^2}{n}$, ha n értéke nagy, akkor a Bayes-becslése lényegében a τ ML-becslése.

A fentiekben felhasználtuk, hogy

$$\int_{-\infty}^{\infty} \Big(\frac{2\tau\pi}{n}\Big)^{-\frac{1}{2}} exp\Big(-\frac{1}{2\tau}(n(\bar{x}-\mu)^2)\Big)\, d\mu = 1.$$

Az integrandus éppen a $N(\bar{x}, \frac{\tau}{n})$ sűrűségfüggvénye.

5. A klasszikus hipotéziselmélet

A gyakorlatban sokszor fordul elő, hogy egy rendelkezésre álló mintából nyert információk alapján szeretnénk következtetni egy nagyobb sokaságra. Például arra, hogy egy adathalmaz milyen eloszlású, vagy két adathalmaz eloszlásban megegyezik-e. Ez a "módszer" természetesen hordoz magában hibát, hiszen egy nagy adathalmazt nehéz/költséges előállítani, így a paraméterét nem tudjuk pontosan lemérni, annak csak egy kis részéről tudunk nyilatkozni. Viszont az éppen kiragadott mintánk csak nagyon ritkán adhat valós képet az adathalmaz egészéről.

Az adott paraméterről alkotott feltételezéseinket úgynevezett **hipotézisekben** fogalmazzuk meg. Ezek a hipotézisek lehetnek igazak és hamisak is. A mi feladatunk az igazságtartalmuk eldöntése. A mintából számított értékek alapján tudunk dönteni arról, hogy egy hipotézist elfogadunk-e vagy sem. **Nullhipotézisnek** nevezzük azt a hipotézist, amelyben megfogalmazzuk, hogy valamilyen adatunk vagy eloszlásunk nem tér el, változatlan egy másik adathoz vagy eloszláshoz képest. Az **ellenhipotézis** az előbbi hipotézisünk ellentéte. Így a két hipotézisünk egymást kizáró feltételezéseket fogalmaz meg, azaz egyszerre nem teljesülhet mindkettő. Gyakran a feladat inkább arra irányul, hogy az ellenhipotézisben feltett állításunk teljesül-e. Ha a nullhipotézist elvetjük, akkor ebből az következik, hogy az ellenhipotézist tartjuk valószínűbbnek. Láthatjuk, hogy célszerű a hipotéziseket olyan módon megfogalmazni, hogy a figyelem a nullhipotézis elvetésére irányuljon, hiszen ez az informatívabb döntés.

A hipotézisvizsgálat során mindig azt nézzük, hogy a nullhipotézisünk helyes-e, ezért a két hipotézis nem cserélhető fel egymással, nem szimmetrikusak. Mivel a mintánk a véletlen függvénye, ezért sosem lehetünk biztosak a hipotézisünk helyességében, illetve nem helyességében. Így viszont a célunk minél kisebb hibával dönteni. A hipotéziselméletben kétféle hibát különböztetünk meg:

- **Elsőfajú hibának** nevezzük azt a hibát, amikor a nullhipotézisünk igaz, mégis elutasítjuk azt.
- **Másodfajú hibának** pedig azt, ha elfogadjuk a nullhipotézist úgy, hogy az nem teljesül.

A kettő hiba közül az elsőfajú hibát tartjuk súlyosabbnak, ezért célunk, hogy ennek a hibának a valószínűséget egy bizonyos érték alatt tartsuk.

Ezek után felállítunk egy **próbastatisztikát**, amely voltaképpen egy valószínűségi változó. Egyéb adatok segítségével kiszámoljuk a kritikus értéket, a kritikust tartományt, majd az így kapott információkat összevetve a próbastatisztikával döntést hozunk a nullhipotézis helyességével kapcsolatban.

A hipotézisvizsgálat egyéb fontos értékei:

- Konfidencia szint $(1-\alpha)$

  Ha a nullhipotézisünk igaz, akkor annak a valószínűsége, hogy elfogadjuk a nullhipotézist.

- Szignifikancia szint, terjedelem (α)

  Ha a nullhipotézis igaz, akkor annak a valószínűsége, hogy elutasítjuk a nullhipotézist. (Ezt az értéket már hipotézisvizsgálat előtt eldöntjük).

- Empirikus szignifikancia szint (**p-érték**)

  A p-érték az a legnagyobb terjedelem, ami mellett nem utasítjuk el a nullhipotézist.

6. A bayesi hipotéziselmélet[8]

A bayesi hipotézisvizsgálat habár alapjaiban hasonló, mégis sokban különbözik a klasszikus hipotézisvizsgálattól. Fontos különbség, hogy míg a klasszikus statisztikában a nullhipotézis és az ellenhipotézis egymás ellentettjei, addig a bayesi hipotézisvizsgálat esetén nem feltétlenül, így akár azok szerepet is cserélhetnek. Vagyis a bayesi hipotézisvizsgálatban nincs kitüntetett nullhipotézis.

Másrészt, a klasszikus hipotézisvizsgálat során vizsgált szignifikanciaszint vagy a p-érték a bayesi hipotézisvizsgálatban nem szerepel.

Legyen H egy tetszőleges hipotézis, y pedig a H-hoz tartozó tapasztalaton nyugvó bizonyíték, a mintánk. Legyen $p(H)$ a H hipotézis mintavétel előtti valószínűsége, más néven az a priori valószínűség, $p(y)$ pedig a mintánk bekövetkezésének valószínűsége. $p(y|H)$ jelentse azt, hogy ha a H hipotézisünk fennáll, akkor mennyi annak a valószínűsége, hogy az y mintát kapjuk. $p(H|y)$ pedig jelentse azt, hogy amennyiben adott a mintánk, akkor mennyi annak a valószínűsége, hogy a hipotézisünk teljesül. Ekkor a következőképpen írhatjuk fel a Bayes-tételt:

$$p(H|y) = \frac{p(y|H)p(H)}{p(y)}.$$

Ha több hipotézisünk van (H_i), akkor hasonlóan az előbbiekhez, felhasználva a teljes valószínűség tételét, a nevezőt az alábbi módon írhatjuk át:

$$p(y) = \sum_i p(y|H_i)p(H_i).$$

Így ekkor a következőt kapjuk:

$$p(H_j|y) = \frac{p(y|H_j)p(H_j)}{\sum\limits_i p(y|H_i)p(H_i)}.$$

[8] A fejezet az [1], [2], [3] alapján készült.

A kapott formula kiemelkedően fontos a Bayes-statisztikában. Ugyanis a képlet szerint kiindulunk valamilyen tudásból, majd tapasztalatokat szerzünk, azokat kiértékeljük és összehasonlítjuk a kezdeti feltételezéseinkkel, így a kezdeti tudásunkat bővíteni tudtuk.

A bayesi hipotézisvizsgálat két fontos mennyisége két hipotézis esetén:

$\frac{p(H_1)}{p(H_0)}$, amelyet *a priori esélyhányadosnak*, illetve a

$\frac{p(H_1|y)}{p(H_0|y)}$, amelyet *a poszteriori esélyhányadosnak* hívunk.

A két érték között felírható egy összefüggés is:

$$\frac{p(H_1|y)}{p(H_0|y)} = \frac{p(H_1)}{p(H_0)}\frac{p(y|H_1)}{p(y|H_0)}$$

A jobboldali szorzótényező voltaképpen a Bayes-faktor legegyszerűbb formájának felel meg (lásd. későbbi fejezet).

A klasszikus hipotézisvizsgálattal szemben a bayesi hipotézisvizsgálat döntéshozatala lényegesen egyszerűbb. Az úgynevezett *MAP teszt (Maximum A Posteriori Probability Test)* segítségével tudunk dönteni. A teszt szerint a nagyobb értékű $p(H_j|y)$-hoz tartozó H_j hipotézist tartjuk a legvalószínűbbnek. Ez két hipotézis esetén a következőképpen fogalmazható át:

- Ha az a poszteriori esélyhányados > 1, akkor elvetjük a H_0-t, mert a mintavétel és az információ feldolgozás után az ellenhipotézis esélyesebbnek tűnik.
- Ha az a poszteriori esélyhányados < 1, akkor elfogadjuk a H_0-t.
- Ha az a poszteriori esélyhányados $= 1$, akkor a két hipotézis egyformán valószínű.

6.1. A minta eloszlásának kiszámítása adott hipotézis mellett

A $p(y|H_j)$ kiszámítása egyszerű hipotézis esetén

Ha $H_0 : \theta = \theta_0$, akkor $\theta|H_0 \sim \delta_{\theta_0}$ eloszlásúnak tekinthető, ahol δ_{θ_0} a θ_0-ra koncentrált Dirac delta függvény.

$$\delta_{\theta_0}(x) = \begin{cases} 1, & ha \quad \theta_0 = x \\ 0, & ha \quad \theta_0 \neq x \end{cases}$$

Ekkor:

$$p(y|H_0) = \int p(y|\theta)p(\theta|H_0)\, d\theta = \int p(y|\theta)p(\theta|\theta = \theta_0)\, d\theta = p(y|\theta_0).$$

A $p(y|H_j)$ kiszámítása összetett hipotézis esetén

Legyen a j-edik hipotézisünk a következő

$$H_j : \theta \in \Theta_j$$

Ebben az esetben θ nem egy konkrét érték, hanem a Θ_j paramétertérből vehet fel értéket. Ekkor a Θ_j paramétertéren meg kell adnunk egy $p(\theta|H_j)$ sűrűségfüggvényt. Így

$$p(y|H_j) = \int\limits_{\Theta_j} p(y|\theta)p(\theta|H_j)\, d\theta.$$

A következő két alfejezetben a bayesi hipotézisvizsgálat azon eseteit fogjuk alaposabban is megvizsgálni, amikor a hipotéziseink ugyanolyan típusúak, azaz a hipotéziseink vagy mind egyszerűek, vagy mind összetettek.

6.2. Ha minden hipotézisünk egyszerű[9]

Legyen $Y \sim p(y|\theta)$, ahol $p(y|\theta)$ a modell eloszlása. A hipotéziseink pedig a következő egyszerű hipotézisek $j = 1, 2 \ldots, J$-ig:

$$H_j : \theta = d_j.$$

A H_j hipotézis mintavétel előtti valószínűséget (a priori) jelöljük p_j-vel:

$$p(H_j) = p(\theta = d_j) = p_j.$$

Ekkor a posteriori valószínűség a Bayes-tétel felhasználásával következőképpen számolható:

$$\begin{aligned} p(H_j|y) = p(\theta = d_j|y) &= \frac{p(y|H_j)p(H_j)}{\sum\limits_{k=1}^{J} p(y|H_k)p(H_k)} = \frac{p(y|\theta = d_j)p(\theta = d_j)}{\sum\limits_{k=1}^{J} p(y|\theta = d_k)p(\theta = d_k)} \\ &= \frac{p(y|\theta = d_j)p_j}{\sum\limits_{k=1}^{J} p(y|\theta = d_k)p_k} \propto p(y|\theta = d_j)p_j. \end{aligned}$$

Az így kapott lehetséges értékekből ($j = 1, \ldots, J$ esetén) a legnagyobb értékűnek megfelelő hipotézisnek a legvalószínűbb a bekövetkezése. Ha több ilyen is van, akkor azok a hipotézisek egyformán valószínűek.

6.1. Példa. *Egyszerű hipotézisre*

Legyenek $Y_i \sim Bernoulli(\theta)$ független, azonos eloszlású valószínűségi változók $i = 1, 2, \ldots, n$ esetén. Legyen az a priori eloszlás:

$$p(\theta = d_j) = \frac{1}{11}, \text{ ahol } d_j = \frac{j}{10}, \text{ ha } j = 0, 1, \ldots, 10.$$

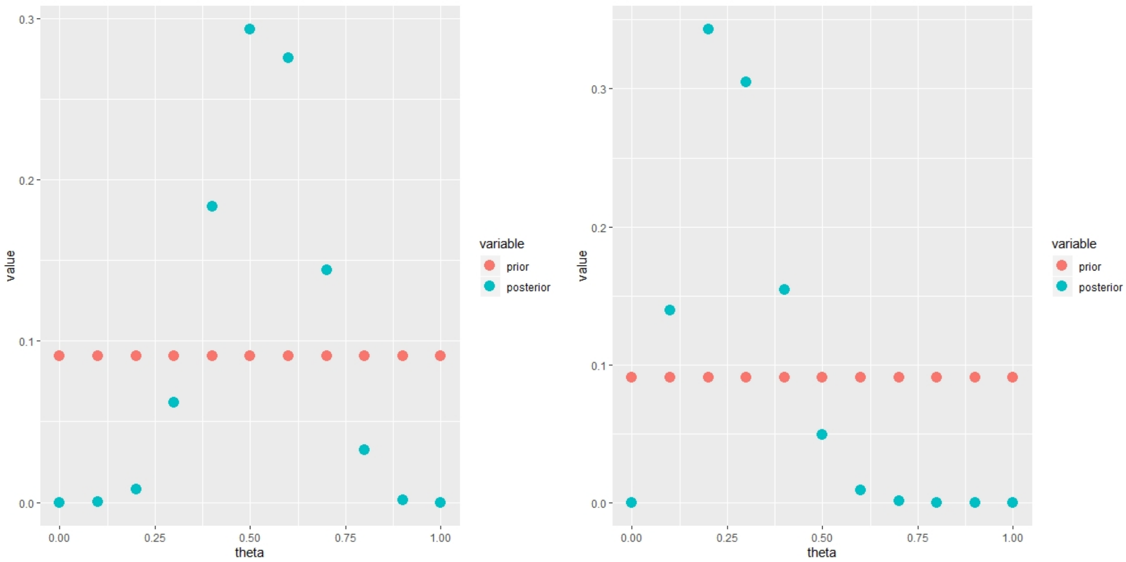
[9] A fejezet a [2] alapján készült.

Ez azt jelenti, hogy kezdetben minden hipotézisünket egyformán valószínűnek tekintjük. Ekkor annak a valószínűsége, hogy a H_j hipotézis teljesül, amennyiben az y minta adott:

$$p(H_j|y) = p(\theta = d_j|y) \propto p(y|\theta = d_j)p_j = \frac{1}{11}\prod_{i=1}^{n} {d_j}^{y_i}(1-d_j)^{1-y_i}$$
$$= \frac{1}{11}{d_j}^{\sum\limits_{i=1}^{n} y_i}(1-d_j)^{n-\sum\limits_{i=1}^{n} y_i} \propto {d_j}^{n\bar{y}}(1-d_j)^{n(1-\bar{y})},$$

ahol $\bar{y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$ a mintaátlagot jelöli.

A következő két ábra két különböző mintából számított a poszteriori eloszlást ábrázol. $n = 13$ és $\theta = 0.4$.



(a) ábra

(b) ábra

6. ábra. Példa diszkrét a priori esetén

Az (a) ábrához tartozó mintánk:

$$y = (1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0).$$

A (b) ábrához tartozó mintánk pedig:

$$y = (0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0).$$

Az (a) ábrán a H_5 hipotézishez tartozó a poszteriori érték a legnagyobb, ezért a legvalószínűbbnek a H_5 hipotézis bekövetkezését tekintjük, azaz $\theta = d_5$. A H_6 hipotézis a poszteriori értéke a második legmagasabb. A többi hipotézisé jóval alacsonyabb. Ezek bekövetkezésének valószínűsége egyre kisebb.

A (b) ábrán jól látható, hogy a H_2 és a H_3 hipotézisekhez tartozó a poszteriori értékek a legmagasabbak. A H_7, H_8, H_9, H_{10} hipotézisek a poszteriori értéke nagyon alacsony, majdnem 0 vagy éppen 0. Ezért annak a valószínűsége, hogy ezek a hipotézisek teljesülnek, azaz θ a d_7, d_8, d_9 vagy d_{10} értékek valamelyikét veszi fel szinte biztosan nem igaz.

6.3. Ha minden hipotézisünk összetett[10]

Legyen $Y \sim p(y|\theta)$ a modell eloszlása. A hipotéziseink pedig a következő összetett hipotézisek $j = 1, 2 \ldots, J$-ig:

$$H_j : \theta \in (E_{j-1}, E_j],$$

ahol az $(E_{j-1}, E_j]$ intervallumok diszjunktak. Ekkor a H_j hipotézis mintavétel előtti (a priori) valószínűségét az alábbiak szerint tudjuk számolni:

$$p(H_j) = p(\theta \in (E_{j-1}, E_j]) = p(E_{j-1} < \theta \leq E_j) = p(E_{j-1} < \theta < E_j) = \int\limits_{E_{j-1}}^{E_j} p(\theta)\, d\theta.$$

ahol $p(\theta)$ a θ paraméterünk a priori eloszlásának sűrűségfüggvénye. A fenti egyenlőségnél felhasználtuk, hogy $p(\theta = E_j) = 0$, mert θ folytonos valószínűségi változó.
A posteriori eloszlást ekkor a következőből kapjuk:

$$p(H_j|y) = p(E_{j-1} < \theta < E_j|y) = \int\limits_{E_{j-1}}^{E_j} p(\theta|y)\, d\theta.$$

Hasonlóan az egyszerű hipotézis esetéhez a legnagyobb $p(H_j|y)$ értékhez tartozó H_j hipotézis a legvalószínűbb.

6.2. Példa. *Összetett hipotézisre*
Legyenek $Y_i \sim Bernoulli(\theta)$ független, azonos eloszlású valószínűségi változók $i = 1, 2, \ldots, n$ esetén. Legyen $E_j = \frac{j}{10}$, ha $j = 0, 1, \ldots, 10$. Tételezzük fel, hogy $\theta \sim$ *Béta*$(1, 1)$, vagyis θ egyenletes a $[0, 1]$ zárt intervallumon.
Ekkor az a poszteriori eloszlás:

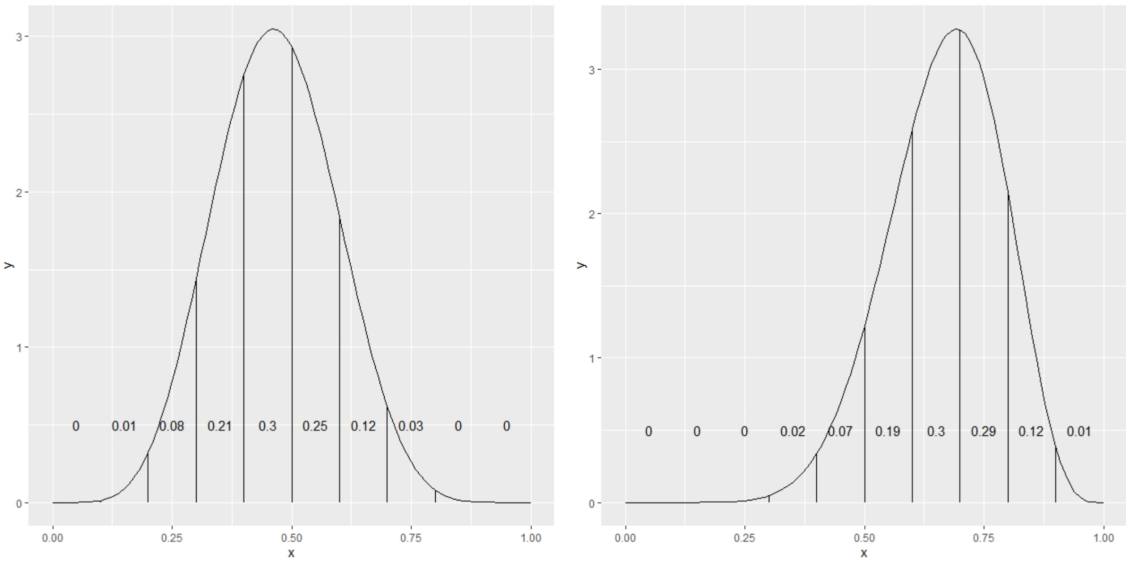
$$p(\theta|y) \propto p(y|\theta)p(\theta) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} = \theta^{n\bar{y}}(1-\theta)^{n(1-\bar{y})},$$

$$\theta|y \sim \textit{Béta}(1 + n\bar{y}, 1 + n(1-\bar{y})).$$

A hipotézisek a poszteriori valószínűségeit pedig a következő képletből számolhatjuk.

$$p(H_j|y) = \int\limits_{E_{j-1}}^{E_j} p(\theta|y)\, d\theta,$$

ha behelyettesítjük a $\theta|y$ sűrűségfüggvényét, majd integráljuk a megfelelő intervallumon.



(a) ábra (b) ábra

7. ábra. Példa Béta a priori esetén

Az (a) ábrához tartozó mintánk a következő:

$$y = (0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1).$$

Az ábrán az egyes $(E_{j-1}, E_j]$ intervallumokhoz tartozó a poszteriori értékek is fel vannak tüntetve. Az ábrából leolvasható, hogy a az $(E_4, E_5]$ intervallumon a legnagyobb a poszteriori eloszlás értéke, így a MAP teszt szerint a H_5 hipotézis a legvalószínűbb, azaz a θ ismeretlen paraméterünk ebbe az intervallumba kell essen. Az is leolvasható, hogy annak a valószínűsége, hogy a θ paraméter a H_1, H_9, H_{10} hipotézisekhez tartozó intervallumba esik, szinte biztosan nem igaz.

A (b) ábra az előbbi gondolatmenethez hasonlón elemezhető. Ebben az esetben a mintánk a következő volt:

$$y = (1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1).$$

A legnagyobb a poszteriori érték a a H_7 hipotézishez tartozik, ez az érték 0.3. Tehát a legvalószínűbb, hogy a θ ismeretlen paraméterünk az ehhez a hipotézishez tartozó intervallumba, azaz $(E_6, E_7]$ intervallumba esik.

[10] A fejezet a [2] alapján készült.

7. A Bayes-faktor[11]

Két hipotézis esetén lehetőségünk van úgynevezett Bayes-faktorral számolni, amely értéke segíthet eldönteni, hogy melyik hipotézist érdemesebb támogatnunk. Legyen a két hipotézisünk a H_1 és H_2, a mintánk pedig y. A $\frac{p(H_1)}{p(H_2)}$ hányadost *a priori esélyhányadosnak* nevezzük. Egyes szakirodalmak a következőképpen jelölik:

$$O[H_1 : H_2] = \frac{p(H_1)}{p(H_2)},$$

ahol $p(H_i)$ a H_i hipotézis valószínűségét jelöli $i = 1, 2$ esetén. Ez a tört azt mondja meg, hogy kezdetben, a mintavétel előtt a H_1 és a H_2 hipotézisek közül melyiket tartjuk esélyesebbnek.

- Ha $\frac{p(H_1)}{p(H_2)} > 1$, akkor a H_1 hipotézist tartjuk esélyesebbnek.
- Ha $\frac{p(H_1)}{p(H_2)} < 1$, akkor a H_2 hipotézist tartjuk esélyesebbnek.
- Ha $\frac{p(H_1)}{p(H_2)} = 1$, akkor a két hipotézist egyformán esélyesnek tartjuk.

Ha $p(H_1) = 1 - p(H_2)$, akkor az a prior esélyhányados a következőképpen is írható:

$$O[H_1 : H_2] = \frac{p(H_1)}{p(H_2)} = \frac{p(H_1)}{1 - p(H_1)}.$$

A $\frac{p(H_1|y)}{p(H_2|y)}$ hányadost *a poszteriori esélyhányadosnak* nevezzük, és a következőképpen jelöljük:

$$PO[H_1 : H_2] = \frac{p(H_1|y)}{p(H_2|y)}.$$

Felhasználva a Bayes-tételt, az a poszteriori esélyhányadost átírhatjuk egy másik alakra.

$$PO[H_1 : H_2] = \frac{p(H_1|y)}{p(H_2|y)} = \frac{p(y|H_1)p(H_1)}{p(y)} \; \frac{p(y)}{p(y|H_2)p(H_2)} \tag{3}$$

$$= \frac{p(y|H_1)}{p(y|H_2)} \times \frac{p(H_1)}{p(H_2)} = \text{Bayes-faktor} \times \text{prior esélyhányados} \tag{4}$$

A *Bayes-faktor* tehát a következő hányados:

$$BF[H_1 : H_2] = \frac{p(y|H_1)}{p(y|H_2)}.$$

Ha a Bayes-faktort fordított esetben szeretnénk kiszámolni, akkor

$$BF[H_2 : H_1] = \frac{p(y|H_2)}{p(y|H_1)}.$$

A kettő felírás között megfogalmazhatunk egy kapcsolatot:

$$BF[H_2 : H_1] = \frac{p(y|H_2)}{p(y|H_1)} = \frac{1}{BF[H_1 : H_2]}.$$

[11] A fejezet a [6] és [10] alapján készült.

A (4)-es egyenlőség a jelölések felhasználásával az alábbiak szerint írható:

$$PO[H_1 : H_2] = BF[H_1 : H_2] \times O[H_1 : H_2].$$

Érdemes megjegyezni, hogy ha $P(H_1) = P(H_2) = 0.5$, akkor

$$PO[H_1 : H_2] = BF[H_1 : H_2].$$

A Bayes-faktor kiszámítása kétféleképpen történhet attól függően, hogy egyszerű vagy összetett esetben vagyunk. Egyszerű esetben a Bayes-faktor a likelihoodok hányadosa a megfelelő hipotézisek mellett:

$$BF[H_1 : H_2] = \frac{p(y|H_1)}{p(y|H_2)}.$$

Összetett esetben pedig a Bayes-faktor a marginális likelihoodok hányadosa a megfelelő hipotézisek mellett:

$$BF[H_1 : H_2] = \frac{\int p(y|\theta, H_1)\, d\theta}{\int p(y|\theta, H_2)\, d\theta}.$$

Az a poszteriori esélyhányados vizsgálatán kívül egy másik módszer arra, hogy a hipotéziseink közül válasszunk az, ha a Bayes-faktort vizsgáljuk.

Harold Jeffreys, angol matematikus 1961-ben megalkotott egy táblázatot, amely azt szemlélteti, hogy a Bayes-faktor ($BF[H_1 : H_2]$) ismeretében a két hipotézist mekkora meggyőződéssel érdemes elutasítanunk, illetve elfogadnunk.

Bayes-faktor		Érték	Értelmezés
	>	100	Extrém bizonyíték H_1 mellett
30	-	100	Nagyon erős bizonyíték H_1 mellett
10	-	30	Erős bizonyíték H_1 mellett
3	-	10	Közepes bizonyíték H_1 mellett
1	-	3	Anekdotikus bizonyíték H_1 mellett
	1		Nincs bizonyíték H_1 mellett
1/3	-	1	Anekdotikus bizonyíték H_2 mellett
1/10	-	1/3	Közepes bizonyíték H_2 mellett
1/30	-	1/10	Erős bizonyíték H_2 mellett
1/100	-	1/30	Nagyon erős bizonyíték H_2 mellett
	<	1/100	Extrém bizonyíték H_2 mellett

Kass és Raftery (1995) táblázata a Bayes-faktor természetes alapú logaritmusának értékeivel számol.

$2log(BF[H_2 : H_1])$		Érték	Értelmezés
0	-	2	Nincs bizonyíték H_1 ellen
2	-	6	Létező bizonyíték H_1 ellen
6	-	10	Erős bizonyíték H_1 ellen
	>	10	Nagyon erős bizonyíték H_1 ellen

7.1. Példa Bayes-faktor számítására

7.1. Példa. Legyen $Y_i \sim Bernoulli(\theta)$ (pl.: a pénzérme feldobás) $i = 1, 2, \ldots, n$. A két hipotézisünk pedig a következő:

$$H_1 : \theta = \frac{1}{2}\ (= \theta_0)$$

$$H_2 : \theta \neq \frac{1}{2}$$

Tegyük fel, hogy az a priori eloszlásunk $\theta|H_2 \sim$ *Béta*(α, β). Ekkor a Bayes-faktor a következőképpen számolható:

$$BF[H_1 : H_2] = \frac{p(y|H_1)}{p(y|H_2)} = \frac{\int p(y|\theta, H_1)\, d\theta}{\int p(y|\theta, H_2)\, d\theta} = \frac{\int p(y|\theta)p(\theta|H_1)\, d\theta}{\int p(y|\theta)p(\theta|H_2)\, d\theta}.$$

Korábban megállapítottuk, hogy egyszerű hipotézis esetén $p(y|H_1) = \int p(y|\theta)p(\theta|H_1)\, d\theta = p(y|\theta_0)$, ezért a számlálót a következőképpen tudjuk számolni:

$$p(y|\theta_0) = (\theta_0)^{n\bar{y}}(1-\theta_0)^{n(1-\bar{y})} = \left(\frac{1}{2}\right)^{n\bar{y}}\left(1-\frac{1}{2}\right)^{n(1-\bar{y})} = \left(\frac{1}{2}\right)^{n}.$$

A nevezőben:

$$p(y|\theta) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} = \theta^{\sum_{i=1}^{n} y_i}(1-\theta)^{n-\sum_{i=1}^{n} y_i} = \theta^{n\bar{y}}(1-\theta)^{n(1-\bar{y})},$$

illetve

$$p(\theta|H_2) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \text{ha } 0 \leq \theta \leq 1.$$

Ekkor a Bayes-faktor:

$$\begin{aligned} BF[H_1 : H_2] &= \frac{(\frac{1}{2})^n}{\int_0^1 \theta^{n\bar{y}}(1-\theta)^{n(1-\bar{y})}\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta} \\ &= \frac{(\frac{1}{2})^n}{\frac{1}{B(\alpha,\beta)}\int_0^1 \theta^{n\bar{y}+\alpha-1}(1-\theta)^{n(1-\bar{y})+\beta-1}\, d\theta} \\ &= \frac{(\frac{1}{2})^n}{\frac{1}{B(\alpha,\beta)}B(\alpha+n\bar{y}, n(1-\bar{y})+\beta)} = \frac{(\frac{1}{2})^n B(\alpha,\beta)}{B(\alpha+n\bar{y}, n(1-\bar{y})+\beta)}, \end{aligned}$$

ahol $B(\alpha, \beta) = \int\limits_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx$ a béta függvény.

8. Pédák bayesi hipotézisvizsgálatra

8.1. Bayesi u-próba (z-próba) egymintás esetben[12]

Legyen az y mintánk normális eloszlású, melyre σ^2 ismert, μ ismeretlen paraméter.

$$Y_1, Y_2, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$$

A továbbiakban azt szeretnénk eldönteni, hogy jogos-e az az állításunk, hogy a minta várható értéke megegyezik egy előre meghatározott értékkel. A feladat szövege szerint a hipotéziseink így a következőképpen fogalmazhatóak meg:

$$H_1 : \mu = \mu_0$$
$$H_2 : \mu \neq \mu_0,$$

ahol μ_0 az előre megadott érték.

A Bayes-faktor kiszámításához meg kell határoznunk a paraméterek a priori eloszlását. Mivel σ^2 ismert, ezért ez nem számít paraméternek, ennek az a priori eloszlásával nem kell foglalkoznunk. A H_1 hipotézis mellett μ pontosan μ_0, így ez 1 valószínűséggel teljesül. A H_2 hipotézis esetén már nem vagyunk biztosak μ értékében, a μ nincs specifikálva. A bizonytalanságunkat az a priori eloszlásban fogalmazzuk meg. Tegyük fel, hogy a paraméterünk a priori eloszlása a minta konjugált priorja, azaz normális eloszlású az alábbi paraméterekkel:

$$\mu|H_2 \sim N\left(\mu_0, \frac{\sigma^2}{n_0}\right).$$

Így feltettük, hogy a μ a hipotézisben szereplő μ_0 várható értékű normális eloszlást követi, valamint a normális eloszlás sűrűségfüggvényének szimmetriája miatt a μ ugyanannyira lehet kisebb és nagyobb is μ_0-nál. Az a tényező, amely eldönti, hogy μ értéke mennyivel tér el μ_0-tól, a szórásnégyzetben fogalmazódik meg. Legyen ez n_0.

A Bayes-faktorunkat szokásos módon a definíció szerint tudjuk számolni. A H_1 hipotézis egyszerű hipotézis, ezért a számlálóban nem kell integrálnunk.

$$\begin{aligned} BF[H_1 : H_2] &= \frac{p(y|H_1)}{p(y|H_2)} = \frac{p(y\big|\mu = \mu_0, \sigma^2)}{\int p(y\big|\mu, \sigma^2) p(\mu\big|\mu_0, n_0, \sigma^2)\, d\mu} \\ &= \left(\frac{n + n_0}{n_0}\right)^{1/2} exp\left(-\frac{1}{2}\frac{n}{n + n_0} Z^2\right), \end{aligned}$$

ahol

$$Z = \frac{\bar{y} - \mu_0}{\sigma}\sqrt{n} \text{ az u-próba próbastatisztikája, és } \bar{y} \text{ a minta átlagát jelöli.}$$

[12]A fejezet a [6] alapján készült.

Ezt a képletet a következő levezetésből kaphatjuk meg.

$$B[H_1 : H_2] = \frac{p(y|H_1)}{p(y|H_2)} = \frac{p(y|\mu = \mu_0, \sigma^2)}{\int p(y|\mu, \sigma^2)p(\mu|\mu_0, n_0, \sigma^2)\, d\mu} \tag{5}$$

$$= \frac{\prod\limits_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp\big(- \frac{(y_i-\mu_0)^2}{2\sigma^2}\big)}{\int\limits_{-\infty}^{\infty} \left(\prod\limits_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp\big(- \frac{(y_i-\mu)^2}{2\sigma^2}\big) \right) \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n_0}}} exp\big(- \frac{(\mu-\mu_0)^2}{2\frac{\sigma^2}{n_0}}\big)\, d\mu} \tag{6}$$

$$= \frac{exp\big(- \frac{\sum\limits_{i=1}^{n}(y_i-\mu_0)^2}{2\sigma^2}\big)}{\frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n_0}}} \int\limits_{-\infty}^{\infty} exp\Big(- \frac{1}{2\sigma^2}(\sum(\mu - y_i)^2 + n_0(\mu - \mu_0)^2)\Big)\, d\mu}. \tag{7}$$

Szeretnénk a (8) egyenletet egy szebb alakra kifejezni. Ezért tovább számolunk. Ehhez először az egyenlet nevezőjét alakítjuk át. Célunk, hogy az exponenciális függvény kitevőjében egy teljes négyzetet alakítsunk ki, amelyben a teljes négyzet egyik tényezője μ.

$$\frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n_0}}} \int\limits_{-\infty}^{\infty} exp\Big(- \frac{1}{2\sigma^2}(\sum(\mu - y_i)^2 + n_0(\mu - \mu_0)^2)\Big)\, d\mu$$

$$= \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n_0}}} \int\limits_{-\infty}^{\infty} exp\Big(- \frac{1}{2\sigma^2}(n\mu^2 - 2\mu \sum y_i + \sum y_i^2 + n_0\mu^2 - 2\mu n_0\mu_0 + n_0\mu_0^2)\Big)\, d\mu$$

$$= \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n_0}}} \int\limits_{-\infty}^{\infty} exp\Big(- \frac{1}{2\sigma^2}(\mu^2(n + n_0) - 2\mu(\sum y_i + n_0\mu_0) + (\sum y_i^2 + n_0\mu_0^2)\Big)\, d\mu$$

$$= \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n_0}}} \int\limits_{-\infty}^{\infty} exp\Big(- \frac{1}{2\sigma^2}((n + n_0)(\mu - \frac{\sum y_i + n_0\mu_0}{n + n_0})^2 - \frac{(\sum y_i + n_0\mu_0)^2}{n + n_0}$$

$$+ (\sum y_i^2 + n_0\mu_0^2))\Big)\, d\mu$$

$$= exp\Big(\frac{1}{2\sigma^2}(\frac{(\sum y_i + n_0\mu_0)^2}{n + n_0} - \sum y_i^2 - n_0\mu_0^2)\Big)$$

$$* \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n_0}}} \int\limits_{-\infty}^{\infty} exp\Big(- \frac{(\mu - \frac{\sum y_i + n_0\mu_0}{n+n_0})^2}{2\frac{\sigma^2}{n+n_0}}\Big)\, d\mu$$

$$= exp\Big(\frac{1}{2\sigma^2}(\frac{(\sum y_i + n_0\mu_0)^2}{n + n_0} - \sum y_i^2 - n_0\mu_0^2)\Big)$$

$$* \sqrt{\frac{n_0}{n + n_0}} \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n_0+n}}} \int\limits_{-\infty}^{\infty} exp\Big(- \frac{(\mu - \frac{\sum y_i + n_0\mu_0}{n+n_0})^2}{2\frac{\sigma^2}{n+n_0}}\Big)\, d\mu$$

$$= exp\Bigg(\frac{1}{2\sigma^2}(\frac{(\sum y_i + n_0\mu_0)^2}{n+n_0} - \sum y_i^2 - n_0\mu_0^2)\Bigg)\sqrt{\frac{n_0}{n+n_0}}\ 1$$

A fentiekben felhasználtuk, hogy

$$N(\frac{\sum y_i + n_0\mu_0}{n+n_0}, \frac{\sigma^2}{n+n_0})$$

eloszlás sűrűségfüggvényének integrálja mínusz végtelentől végtelenig 1.

Folytassuk a (8) egyenlet kiszámítását, felhasználva az előbbi számolásunkat.

$$= \frac{exp\big(-\frac{\sum(y_i-\mu_0)^2}{2\sigma^2}\big)}{exp\Bigg(\frac{1}{2\sigma^2}(\frac{(\sum y_i+n_0\mu_0)^2}{n+n_0} - \sum y_i^2 - n_0\mu_0^2)\Bigg)\sqrt{\frac{n_0}{n+n_0}}}$$

$$= \sqrt{\frac{n_0+n}{n_0}}\ exp\Bigg(-\frac{1}{2\sigma^2}(\sum(y_i-\mu_0)^2 + \frac{(\sum y_i + n_0\mu_0)^2}{n+n_0} - \sum y_i^2 - n_0\mu_0^2)\Bigg)$$

$$= \sqrt{\frac{n_0+n}{n_0}}\ exp\Bigg(-\frac{1}{2\sigma^2}(\sum y_i^2 - 2\mu_0\sum y_i + n\mu_0^2 + \frac{(\sum y_i + n_0\mu_0)^2}{n+n_0} - \sum y_i^2 - n_0\mu_0^2)\Bigg)$$

$$= \sqrt{\frac{n_0+n}{n_0}}\ exp\Bigg(-\frac{1}{2\sigma^2}((n-n_0)\mu_0^2 - 2\mu_0\sum y_i + \frac{(\sum y_i + n_0\mu_0)^2}{n+n_0})\Bigg)$$

$$= \sqrt{\frac{n_0+n}{n_0}}\ exp\Bigg(-\frac{1}{2\sigma^2(n+n_0)}(n^2\mu_0^2 - n_0^2\mu_0^2 - 2(n+n_0)\mu_0\sum y_i + (\sum y_i)^2$$

$$+ 2n_0\mu_0\sum y_i + n_0^2\mu_0^2)\Bigg)$$

$$= \sqrt{\frac{n_0+n}{n_0}}\ exp\Bigg(-\frac{1}{2\sigma^2(n+n_0)}(\sum y_i - n\mu_0)^2\Bigg)$$

$$= \sqrt{\frac{n_0+n}{n_0}}\ exp\Bigg(-\frac{1}{2\sigma^2(n+n_0)}n^2(\bar{y}-\mu_0)^2\Bigg)$$

$$= \sqrt{\frac{n_0+n}{n_0}}\ exp\Bigg(-\frac{1}{2}\frac{n}{n+n_0}(\frac{\bar{y}-\mu_0}{\sigma}\sqrt{n})^2\Bigg)$$

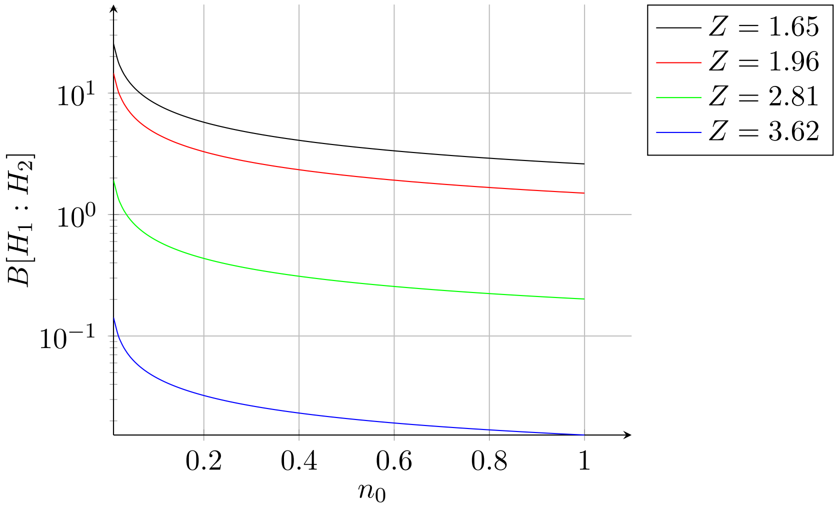
$$= \Bigg(\frac{n+n_0}{n_0}\Bigg)^{1/2} exp\Bigg(-\frac{1}{2}\frac{n}{n+n_0}Z^2\Bigg)$$

ahol

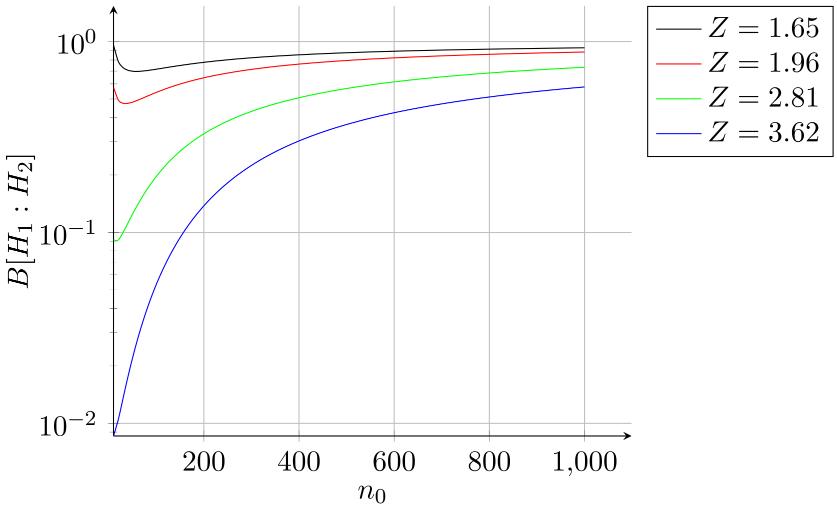
$$Z = \frac{\bar{y}-\mu_0}{\sigma}\sqrt{n} \text{ az u-próba próbastatisztikája, és } \bar{y} \text{ a minta átlagát jelöli.}$$

Ekkor a Bayes-faktorra úgy is tekinthetünk, mint az n, n_0 és a Z paraméterek függvényére, ahol az n darab y_i megfigyelésünk mellett még n_0 darab μ_0 megfigyelésünk is volt.

A 8. és 9. ábra a Bayes-faktor értékét ábrázolja n_0 függvényében $n = 100$ és adott Z értékek esetén. A 8. ábrán látható, hogy minél nagyobb a Z értéke, a Bayes-faktor annál inkább a H_2 hipotézist támogatja.



8. ábra. A Bayes-faktor értéke n_0 megváltoztatása esetén



9. ábra. A Bayes-faktor értéke n_0 megváltoztatása esetén

A 8. és 9. ábrán jól látható, hogy ha n_0-val 0-hoz tartunk, akkor a Bayes-faktor értéke ∞-hez tart, illetve ha n_0-val ∞-hez tartunk, akkor a Bayes-faktor értéke 1-hez tart.

A Bayes-faktor szemléltetéséhez alkalmazott Z értékeket az ábrákon a leggyakrabban használt szignifikancia szint (terjedelem) alapján választottuk ki.

α értéke	Z értéke
0.1	1.64
0.05	1.96
0.005	2.81
0.0004	3.60

8.1.1. Prior választás a standardizált hatás nagyságra

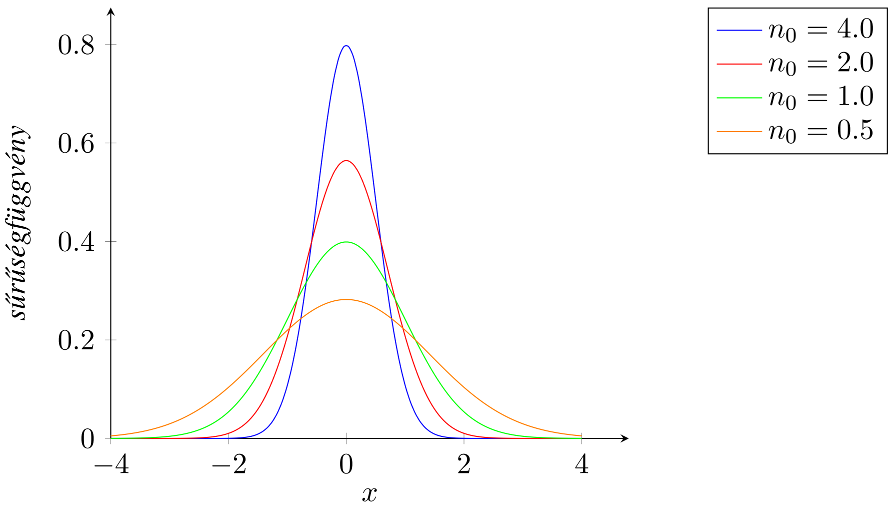
Az előbbiekben az a priori eloszlást az ismeretlen μ paraméterre választottuk. Jeffreys szerint azonban jobb, ha az úgynevezett standard hatásnagyságot *(standard effect size)* vizsgáljuk. Jelöljük ezt δ-val, amelyre

$$\delta = \frac{\mu - \mu_0}{\sigma}.$$

Ekkor az a priorit nem a μ-re, hanem a δ-ra választjuk meg.

$$\delta | H_2 \sim N\left(\frac{\mu_0 - \mu_0}{\sigma}, \left(\frac{1}{\sigma}\right)^2 \frac{\sigma^2}{n_0}\right) = N\left(0, \frac{1}{n_0}\right)$$

A 10. ábra azt szemlélteti, hogy egy fix várható érték mellett (az ábrán ez a várható érték 0) amennyiben az n_0 értékét megváltoztatjuk, azzal változik a δ az a priori eloszlása is H_2 mellett. Minél "laposabb" a haranggörbénk, annál kevésbé informatív az adott a priorink.



10. ábra. Normális eloszlású prior a standard hatás nagyságra

Tehát ha a normális eloszlású, 0 várható értékű a priorik közül választanánk egyet δ-nak, és kevésbé informatív a priorit szeretnénk, akkor n_0 értékét kicsinek kell megválasztanunk. Hiszen minél kisebb n_0 értéke, annál kevesebb információt tartalmaz az a priorink. Nézzük az előbbiekben kiszámított Bayes-faktorunkat.

$$B[H_1 : H_2] = \left(\frac{n + n_0}{n_0}\right)^{1/2} exp\left(-\frac{1}{2}\frac{n}{n + n_0}Z^2\right)$$

Ha n_0-val 0-hoz tartunk, akkor a Bayes-faktor első szorzótényezője végtelenhez tart, míg a második szorzótényezője egy konstanshoz. Így a Bayes-faktorunk függetlenül a mintánktól a H_1 hipotézisünket fogja támogatni.

8.2. Bayesi t-próba egymintás esetben[13]

Legyen az y mintánk normális eloszlású, melyre μ és σ^2 ismeretlen paraméterek.

$$Y_1, Y_2, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$$

Az egymintás t-próba során azt vizsgáljuk, hogy egy adott mintában a várható érték szignifikánsan különbözik-e egy adott μ_0 értéktől. Tehát a hipotéziseink a következőképpen írhatók fel

$$H_1 : \mu = \mu_0$$
$$H_2 : \mu \neq \mu_0$$

Ahhoz, hogy el tudjuk dönteni, hogy melyik hipotézisünket támogatjuk, szükséges a Bayes-faktor kiszámítása. Ehhez meg kell határoznunk a paramétereink a priori eloszlását is, μ-ét a H_2 hipotézis mellett (a H_1 hipotézis mellett $\mu = \mu_0$ 1 valószínűséggel), illetve σ^2 a prioriját a H_1 és H_2 hipotézisek mellett. A következőkben μ helyett a standardizált hatásnagysággal (δ) számolunk tovább.

$$\delta = \frac{\mu - \mu_0}{\sigma}$$

Válasszuk az ismeretlen σ^2 paraméter a priorijának H_1 és H_2 hipotézisek mellett Jeffreys priorját.

$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

illetve a H_2 hipotézis mellett válasszuk δ a priorijának a standard Cauchy eloszlást. Ezeket az a priori választásokat a szakirodalom **Jeffreys-Zellner-Siow** priornak, röviden JZS priornak nevezi.

Így a Bayes-faktor definíció szerint a következőképpen számolható:

$$\begin{aligned} BF[H_1 : H_2] &= \frac{p(y|H_1)}{p(y|H_2)} \\ &= \frac{\int p(y|\delta = 0, \sigma^2) p(\sigma^2|H_1)\, d\sigma^2}{\int \int p(y|\delta, \sigma^2) p(\delta, \sigma^2|H_2)\, d\delta\, d\sigma^2}. \end{aligned}$$

Behelyettesítve a megfelelő a priori eloszlásokat és kiszámolva az integrálokat a Bayes-faktor értéke az alábbi képletre egyszerűsödik.

$$BF[H_1 : H_2] = \frac{\left(1 + \frac{t^2}{v}\right)^{-\frac{(v+1)}{2}}}{\int\limits_0^\infty (1 + ng)^{-\frac{1}{2}} \left(1 + \frac{t^2}{(1+ng)v}\right)^{-\frac{v+1}{2}} (2\pi)^{-\frac{1}{2}} g^{-\frac{3}{2}} e^{-\frac{1}{2g}}\, dg},$$

[13] A fejezet a [9] és [10] alapján készült.

$ahol$

n : a megfigyelések száma

v : a szabadsági fok $v = n - 1$ és

t : a klasszikus egymintás t-próba próbastatisztikája, azaz:

$$t = \frac{\bar{y} - \mu_0}{s}\sqrt{n}$$

$ahol$

$\bar{y}$ a minta átlagát,

s a minta korrigált tapasztalati szórását, és

μ_0 az előre megadott értéket jelöli.

Habár a képlet bonyolultnak tűnik, a kiszámolása mégis egyszerű, hiszen a statisztikusnak elég a t értékét és a minta méretét (n) megadni. Vagyis nem szükséges megadnunk magát a mintát. A konkrét JZS Bayes-faktor ismeretében már el tudjuk dönteni, hogy a Bayes-faktor melyik hipotézist támogatja az alábbi táblázat segítségével.

	A JZS Bayes-faktor értékei			
	H_1 hipotézist támogató		H_2 hipotézist támogató	
n	10	3	1/3	1/10
5	-	0.40	3.15	4.97
10	-	0.89	2.73	3.60
20	-	1.20	2.64	3.26
50	-	1.51	2.68	3.17
100	0.69	1.72	2.76	3.20
200	1.08	1.90	2.86	3.27
500	1.44	2.12	2.99	3.38

1. táblázat. Kritikus t értékek

A táblázatban a Bayes-faktor értékek a $10, 3, 1/3, 1/10$ felosztást követik, a minta nagyságát (n) figyelembe véve. Például, ha 100 megfigyelésből a t értékünk 3.3, akkor a táblázatból leolvasva ez az érték meghaladja a 3.20-at, azaz a kritikus értéket, ezért ennek a t értéknek megfelelő JZS Bayes-faktorunk kevesebb lesz, mint $\frac{1}{10}$. A 7. fejezetben megismert táblázat szerint, így a JZS Bayes-faktorunk a H_2 hipotézist támogatja.

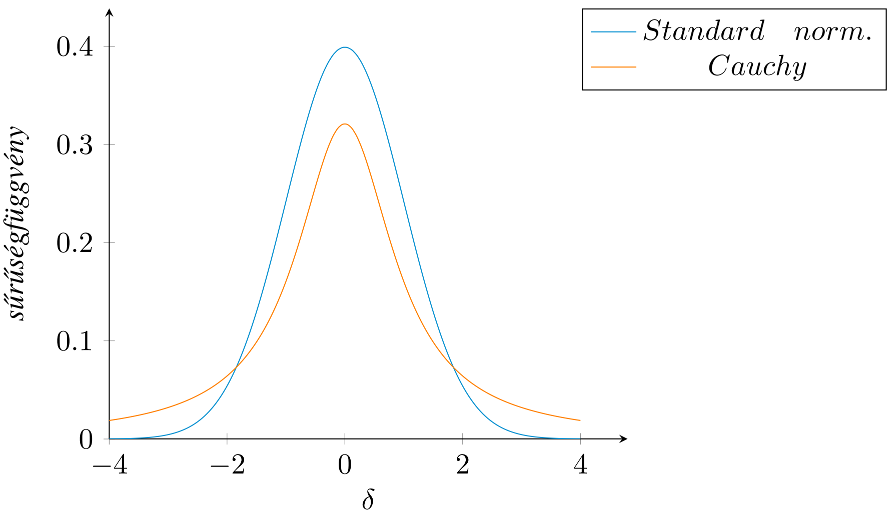
Miért éppen a Cauchy eloszlás?

A δ a priorijának az előbbiekben a JZS prior szerint a standard Cauchy eloszlást választottuk. Felmerülhet a kérdés, hogy vajon ez az eloszlás miért jobb, mint például a normális eloszlás? A következőkben nézzük meg, hogy mi változik, ha a δ a priorijának a standard normális eloszlást választjuk a H_2 hipotézis mellett.

$$\delta|H_2 \sim N(0,1)$$

J.N. Rouder szerint (2009) a JZS prior kevesebb információt hordoz, mint a standard normális eloszlású a priori. Ez valóban így van, mert a hatásnagyság a priori eloszlásának a szórásnégyzetét egy eloszlásnak tekinti (nem pedig egy $\sigma^2 = 1$ konstansnak). Azaz a szórásnégyzet értékét illető bizonytalanságunkat az a prior eloszlásban fogalmazzuk meg. Ez az eloszlás az egy szabadságfokú inverz-khínégyzet eloszlás, azaz az inverz gamma eloszlás $\alpha = \frac{1}{2}$ és $\beta = \frac{1}{2}$ paraméterekkel. Az inverz gamma eloszlást kiintegrálva kapjuk, hogy a marginális eloszlás, a δ Cauchy eloszlású. Zellner és Siow a Cauchy eloszlást 1 paraméterrel választotta. Az így kapott eloszlást standard Cauchy eloszlásnak nevezzük.

A standard Cauchy a priori első ránézésre olyan, mint a standard normális eloszlás sűrűségfüggvénye. Azonban mégis különböznek egymástól.



11. ábra. Standard normális és standard Cauchy a priori a standard hatásnagyságra

A 11. ábrán látható, hogy a standard Cauchy eloszlás $-\infty$ és $+\infty$ között kevésbé gyorsan simul rá az x tengelyre, mint a normális eloszlás. Azaz a nagyobb hatásnagyságok gyakoribbak a Cauchy eloszlás esetén. A ZS prior szerint a H_2 hipotézis súlya nagy δ esetén lassabban oszlik el, mint a standard normális hipotézis esetén. Ha visszagondolunk a Bayes-faktor definíciójára :

$$B[H_1 : H_2] = \frac{p(y|H_1)}{p(y|H_2)} = \frac{\int p(y|\delta, H_1)\, d\delta}{\int p(y|\delta, H_2)\, d\delta} = \frac{\int p(y|\delta)p(\delta|H_1)\, d\delta}{\int p(y|\delta)p(\delta|H_2),\, d\delta}$$

akkor így a kisebb likelihood értékekhez ($p(y|\delta)$) nagyobb súlyozás tartozik ($p(\delta|H_2)$). Ez a nagyobb súlyozás pedig pontosan azt eredményezi, hogy a Bayes-faktorunk (amelyik a H_1 hipotézist támogatja) értéke nagyobb lesz.

8.2.1. A klasszikus és bayesi egymintás t-próba összevetése egy példán keresztül

Ebben az alfejezetben a klasszikus és bayesi egymintás t-próbát fogjuk összevetni egy példán keresztül. Tíz ember egy kísérletben vesz részt, amelyben két különböző altató hatását szeretnék összevetni. Minden ember mindkét gyógyszert kipróbálta. A feljegyzett adatokat az alábbi táblázat szemlélteti.

Tesztalanyok	1. típusú altató	2. típusú altató
1. alany	1.0	1.9
2. alany	-1.1	0.8
3. alany	-0.2	1.1
4. alany	-1.2	0.1
5. alany	-0.1	-0.1
6. alany	3.4	4.4
7. alany	3.7	5.5
8. alany	0.8	1.6
9. alany	0.0	4.6
10. alany	2.0	3.4

A táblázatban az egyes adatok azt jelentik, hogy az adott alany az adott típusú altató szedése mellett mennyivel aludt többet, illetve kevesebbet (órában), mint szokott.

A feladatunk az, hogy eldöntsük a két altató hatása megegyezik-e vagy sem? Legyen a továbbiakban α értéke a hipotézistesztelésben leggyakrabban használt terjedelem, 0.05.

A klasszikus egymintás t-próba kiszámítása

Írjuk fel a feladat szövege szerint a hipotéziseinket.

$$H_1 : \mu_1 = \mu_2 \iff \textit{A várható értékek megegyeznek}$$
$$H_2 : \mu_1 \neq \mu_2 \iff \textit{A várható értékek különböznek}$$

ahol μ_1 az 1. típusú altató adatainak várható értéke, míg μ_2 a 2. típusúé.

A fentiekben meghatározott feladatban szereplő két minta nem független egymástól, hiszen ugyanazokon a személyeken végezték el a kísérletet. Ezért a továbbiakban párosított t-próbát fogunk végezni, azaz a minták különbségére végzünk egymintás t-próbát. (Feltesszük, hogy a különbség normális eloszlású.)

Ekkor a hipotéziseink a következőképpen módosulnak.

$$H_1 : \mu = 0 \iff \textit{A várható értékek megegyeznek}$$

$$H_2 : \mu \neq 0 \iff \textit{A várható értékek különböznek},$$

ahol μ a két minta várható értékének különbségét jelöli.

A hipotézisteszteléshez az R programot fogjuk használni. Az alábbi kód segítségével kiszámoljuk a próbához tartozó p-értéket, valamint a konfidenciaintervallumot is.

```
extra <- c(1,-1.1,-0.2,-1.2,-0.1,3.4,4.7,0.8,0.0,2.0
                  ,1.9,0.8,1.1,0.1,-0.1,4.4,5.5,1.6,4.6,3.4)
group <- c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2)
ID <- c(1,2,3,4,5,6,7,8,9,10,1,2,3,4,5,6,7,8,9,10)
adat <- data.frame(extra = extra, group = group, ID = ID)
library(BayesFactor)
t.test(extra~group, data=adat, paired=TRUE)
```

Ekkor az alábbi eredményt kapjuk.

```
Paired t-test


data:  extra by group
t = -3.6015, df = 9, p-value = 0.005735
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.2793686 -0.5206314
sample estimates:
mean of the differences
                   -1.4
```

Innen leolvashatjuk, hogy a 0 nem eleme a (-2.2793686,-0.5206314) konfidencia intervallumnak, azaz az 1. típusú altató kevesebb alvást biztosít, mint a 2. típusú altató. Hasonló következtetésre juthatunk, ha a p-értéket és az α szignifikancia szintet vizsgáljuk. Mivel p értéke (jelentősen) kisebb, mint α értéke, ezért elvetjük a H_1 hipotézist.

A bayesi egymintás t-próba kiszámítása

Ahhoz, hogy meg tudjuk mondani, melyik hipotézisünket érdemesebb támogatni, ki kell számolnunk a Bayes-faktor értékét.

A számolást az R program segítségével végezzük el. Futassuk le a következő kódot.

```
library(BayesFactor)
ttestBF(x = adat$extra[1:10],y=adat$extra[11:20], paired=TRUE)
```

Ekkor az alábbi eredményre jutunk.

```
Bayes factor analysis
Alt., r=0.707 : 9.704538
Against denominator:
  Null, mu = 0
Bayes factor type: BFoneSample, JZS
```

Leolvasható, hogy a Bayes-faktor értéke, amely a H_2 hipotézist támogatja 9.704538. Ha a 7. fejezetben megismert táblázatot vizsgáljuk a 9.7 a 3-10 intervallumba esik, azaz a H_2 hipotézis mellett közepes bizonyítékot találtunk.

Ha ezt a megállapításunkat összevetjük a klasszikus esettel, akkor arra az eredményre juthatunk, hogy míg a klasszikus esetben a próba elég erősen a H_1 hipotézis elvetését támogatja (a H_2 hipotézist tartja valószínűbbnek), addig a bayesi próba esetén "csak" közepes bizonyítékunk van H_2 hipotézis mellett.

8.3. Bayesi t-próba kétmintás esetben[14]

A klasszikus kétmintás t-próbához hasonlóan a bayesi kétmintás t-próba is két független, normális eloszlású mintát használ, amelyekről tudjuk, hogy a szórásaik megegyeznek (nem szükséges ismernünk a közös szórás pontos értékét). Legyenek a mintáink a következők

$$X_i \overset{\text{i.i.d.}}{\sim} N(\mu_1, \sigma^2)$$

$$Y_i \overset{\text{i.i.d.}}{\sim} N(\mu_2, \sigma^2),$$

ahol $x = (X_1, \dots, X_n)$ és $y = (Y_1, \dots, Y_m)$

A hipotéziseink ekkor:

$$H_1 : \mu_1 = \mu_2 \iff \textit{A várható értékek megegyeznek}$$

$$H_2 : \mu_1 \neq \mu_2 \iff \textit{A várható értékek különböznek}$$

[14] A fejezet a [9] alapján készült.

A következőkben μ_1 és μ_2 helyett azt nézzük, hogy mennyire térnek el a "nagy átlagtól". Azaz ha a μ a "nagy átlag", és ha α a teljes hatás (*total effect*), akkor:

$$X_i \overset{\text{i.i.d.}}{\sim} N(\mu + \frac{\alpha}{2}, \sigma^2)$$

$$Y_i \overset{\text{i.i.d.}}{\sim} N(\mu - \frac{\alpha}{2}, \sigma^2)$$

Most számoljuk ki a két várható érték eltérését.

$$\mu_1 - \mu_2 = \left(\mu + \frac{\alpha}{2}\right) - \left(\mu - \frac{\alpha}{2}\right) = \alpha$$

Ekkor a hipotéziseink az alábbiak szerint módosulnak:

$$H_1 : \alpha = 0 \iff \textit{A várható értékek megegyeznek},$$

$$H_2 : \alpha \neq 0 \iff \textit{A várható értékek különböznek}.$$

A hipotézisvizsgálathoz szükséges, hogy meghatározzuk a paramétereink a priori eloszlását is, α-ét a H_2 hipotézis mellett (a H_1 hipotézis mellett $\alpha = 0$ 1 valószínűséggel), illetve μ és σ^2 a priorijait a H_1 és H_2 hipotézisek mellett. A következőkben α helyett a standardizált hatás nagysággal (δ) számolunk tovább. Válasszuk az ismeretlen paraméterek a priorijának a JZS priort, azaz H_1 és H_2 hipotézisek mellett használjuk μ, σ^2 Jeffreys priorját:

$$p(\sigma^2) \propto \frac{1}{\sigma^2},$$

$$p(\mu) \propto 1,$$

$$p(\mu, \sigma^2) \propto \frac{1}{\sigma^2},$$

illetve a H_2 hipotézis mellett válasszuk δ-nak a standard Cauchy eloszlást.

A következő lépésben ki kell számolnunk a Bayes-faktort, majd a kapott értéket behelyettesítve a táblázatba (7. fejezet) le tudjuk olvasni, hogy melyik hipotézisünket tartjuk esélyesebbnek.

$$BF[H_1 : H_2] = \frac{p(x, y|H_1)}{p(x, y|H_2)}$$

$$= \frac{\int\int p(x, y|\delta = 0, \mu, \sigma^2) p(\mu, \sigma^2|H_1)\, d\mu\, d\sigma^2}{\int\int\int p(x, y|\delta, \mu, \sigma^2) p(\delta|H_2) p(\mu, \sigma^2|H_2)\, d\mu\, d\sigma^2\, d\delta}$$

A korábbi a priori választásaink miatt ez a Bayes-faktor numerikusan kiszámolható az előbbi, egymintás t-próbához tartozó képlettel, csak a mostani kétmintás esetben némi változtatással. A képletben az eddigi egymintás t-próba próbastatisztikája helyett a kétmintás t-próba próbastatisztikáját kell írnunk, azaz:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)s_x^{*2} + (m-1)s_y^{*2}}} \sqrt{\frac{nm(n+m-2)}{n+m}},$$

ahol

$\bar{x}$: az x minta átlaga,

$\bar{y}$: az y minta átlaga,

s_x^{*2}: az x minta korrigált tapasztalati szórásnégyzete,

s_y^{*2}: az y minta korrigált tapasztalati szórásnégyzete,

n: az x minta mérete,

m: az y minta mérete.

A 8.2. fejezetben látott képletnél N helyébe $\frac{nm}{n+m}$ kerül, míg az eddigi szabadsági fok helyett $(n+m-2)$.

$$BF[H_1 : H_2] = \frac{\left(1 + \frac{t^2}{(n+m-2)}\right)^{-\frac{((n+m-2)+1)}{2}}}{\int\limits_0^\infty (1+\frac{nm}{n+m}g)^{-\frac{1}{2}}\left(1+\frac{t^2}{\left(1+\frac{nm}{n+m}g\right)(n+m-2)}\right)^{-\frac{(n+m-2)+1}{2}}(2\pi)^{-\frac{1}{2}}g^{-\frac{3}{2}}e^{-\frac{1}{2g}}\,dg}$$

8.3.1. A klasszikus és bayesi kétmintás t-próba összevetése egy példán keresztül

Ebben az alfejezetben a klasszikus és a fentiekben megismert bayesi t-próbát fogjuk alkalmazni egy példán keresztül.

A feladatunkban szereplő minta 2004-ből, Észak-Kaliforniából származik, és a észak-kaliforniai édesanyák súlygyarapodását méri várandóságuk ideje alatt (fontban). A két minta két különböző kategória értékeit tárolja, a fiatal és az idősebb édesanyákét. A kérdés az, hogy a minta alapján jogos-e azt mondani, hogy a fiatalabb édesanyák átlagos súlygyarapodása eltér az idősebb anyák átlagos súlygyarapodásától? A következőkben feltesszük, hogy a mintáink függetlenek és az adatok normális eloszlásúak, az alábbi paraméterekkel:

$$Y_{I,i} \overset{\text{i.i.d.}}{\sim} N(\mu_1, \sigma^2),$$

$$Y_{F,i} \overset{\text{i.i.d.}}{\sim} N(\mu_2, \sigma^2),$$

ahol Y_I az idősebb anyák adatait jelöli, míg Y_F pedig a fiatalabb anyukákét.

A hipotéziseink ekkor:

$$H_1 : \alpha = 0 \iff \textit{A várható értékek megegyeznek},$$

$$H_2 : \alpha \neq 0 \iff \textit{A várható értékek különböznek},$$

ahol α az előző fejezetben említett teljes hatás (*total effect*).

A bayesi kétmintás t-próba kiszámítása

Az előző fejezetekben láttuk, hogy a két hipotézisről következtetéseket levonni például a Bayes-faktor értékének ismeretében tudunk (vagy az a poszteriori valószínűségeket vizsgáljuk).

Tegyük fel, hogy

$$p(H_1) = p(H_2) = 0.5.$$

Ekkor nincsenek ismereteink a hipotézisekről, vagy nem szeretnénk, hogy az előzetes tudásunk befolyásolja a végeredményt. Ezen feltevés mellett a Bayes-faktor értéke pontosan az a poszteriori esélyhányados.

A Bayes-faktor kiszámításához most is az R programot fogjuk használni. Feltesszük, hogy az ismeretlen paramétereinkre a korábban megismert JZS priort használjuk. Azaz H_1 és H_2 hipotézisek mellett használjuk μ, σ^2-re Jeffreys priorját. A programba előre be vannak építve a feladatunk adatai (nc adathalmaz), így csak egy függvényt kell meghívnunk a megfelelő változókkal, hogy az értékeket ki tudjuk számítani.

```
library(statsr)
data(nc)
bayes_inference(y=gained, x=mature, data=nc,type='ht',
                statistic='mean',  alternative='twosided', null=0,
                prior='JZS', r=1, method='theo', show_summ=FALSE)
```

A kód lefuttatása után a következő eredményt kapjuk.

```
Hypotheses:
H1: mu_mature mom  = mu_younger mom
H2: mu_mature mom != mu_younger mom

Priors: P(H1) = 0.5  P(H2) = 0.5

Results:
BF[H1:H2] = 5.7162
P(H1|data) = 0.8511
P(H2|data) = 0.1489

Posterior summaries for under H2:
95% Cred. Int.: (-4.3243 , 0.854)
```

A minta alapján tehát annak a valószínűsége, hogy nincs különbség a fiatal és idősebb édesanyák átlagos súlygyarapodása között 0.8511. Míg annak a valószínűsége a minta alapján, hogy a két átlagos súlygyarapodás nem egyezik meg 0.1489. A program kiszámolja nekünk azt az intervallumot is (*credible interval*), amely azt mondja meg, hogy a $\mu_2 - \mu_1$ a poszteriori értéke 95%

valószínűséggel ebbe az intervallumba esik. Azt is leolvashatjuk, hogy a $BF[H_1 : H_2]$ Bayes-faktor értéke 5.7162. Ezt az értéket a korábbi táblázatunkba behelyettesítve (7. fejezet) közepes bizonyítékot kapunk a H_1 hipotézis mellett. Tehát a Bayes-faktor értéke alapján nincs okunk arra, hogy elvessük a H_1 hipotézist, azaz a fiatalabb és idősebb édesanyák átlagos súlygyarapodása szignifikánsan nem különbözik.

A klasszikus kétmintás t-próba kiszámítása

Mivel nem tudjuk, hogy a két minta szórása megegyezik-e, ezért a Welch-próbát fogjuk alkalmazni szintén az R program segítségével. Tegyük fel, hogy a terjedelem (α) 0.05. A program az alábbi kód segítségével kiszámolja a próbastatisztika értékét (t), majd kiszámítja a p-értéket is. Végül a kapott eredményeket kiírja.

```
library(statsr)
data(nc)
t.test(gained ~ mature, data = nc, conf.level = 0.95)
```

A kód lefuttatása után a következő eredményt kapjuk.

```
Welch Two Sample t-test


data:  gained by mature
t = -1.3765, df = 175.34, p-value = 0.1704
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.3071463  0.7676886
sample estimates:
 mean in group mature mom mean in group younger mom
                28.79070                   30.56043
```

Leolvasható, hogy a t értéke -1.3765, a p-érték pedig 0.1704, valamint a program kiszámolja a konfidencia intervallumot is.

Mivel a p-érték (a legnagyobb terjedelem, ami mellett elfogadjuk a H_1 hipotézist) nagyobb, mint a terjedelem (α), ezért a H_1 hipotézis elutasítására nem találtunk komoly bizonyítékot. Azaz, mint a bayesi t-próba esetén, a fiatalabb és az idősebb édesanyák átlagos súlygyarapodása a várandóságuk alatt szignifikáns különbséget nem mutat.

Hivatkozások

[1] David Hitchcock, University of South Carolina Columbia: STAT 535 (Introduction to Bayesian Data Analysis), Spring 2014

[2] Jarad Niemi, University of Iowa State: STAT 544 (Bayesian Statistics), Spring 2014

[3] Dr. Malcolm Farrow, Newcastle University: MAS3301 (Bayesian Statistics), Fall 2008

[4] Peter M. Lee: Bayesian Statistics - An Introduction (4th edition); Wiley Press, 2012

[5] Brian J. Reich, Sujit K. Ghosh: Bayesian Statistical Methods; CRC Press, 2019

[6] Merlise Clyde, Mine Cetinkaya-Rundel, Colin Rundel, David Banks, Christine Chai, Lizzy Huang: An Introduction to Bayesian Thinking; Published with bookdown, 2019

[7] John Cook, University of Texas: Conjugate prior relationships https://www.johndcook.com/blog/conjugate_prior_diagram/

[8] Hunyadi László: Bayesi gondolkodás a statisztikában; Statisztikai Szemle, 89. évfolyam 10–11. szám, 2011

[9] Jeffrey N. Rouder, Paul L. Speckman, Dongchu Sun, and Richard D. Morey: Bayesian t tests for accepting and rejecting the null hypothesis; Psychonomic Bulletin & Review, 16 (2), 225-237; 2009

[10] Thom Baguley: Serious Stats - A Guide to Advanced Statistics for the Behavioral Science; Red Globe Press, 2012

[11] Michael I. Jordan, University of California, Berkeley: Stat 260/CS 294 (Bayesian Modeling and Inference), Spring 2010