Bayes–adatfeldolgozás II

Kálvin Sándor

2014 november 10

Előző előadás

- A Bayes statisztika a $p(A|I)$ valószínűséget úgy definiálja mint a hihetőségét annak, hogy az A esemény igaz, ha az I ismeretek birtokában vagyunk.
- Minden eseményhez, elmélethez valószínűséget tudunk rendelni, nem csak valószínűségi változókhoz.
- A Bayes elmélet a megismerés matematikai leírása, tudásunkat egy hipotézisről hogyan változtatja meg a mérés.

Az alkalmazandó szabály a Bayes tétel

$$p(H|D,I) = \frac{p(D|H,I)p(H|I)}{p(D|I)}$$

Előző előadás

Teljes egymást kölcsönösen kizáró eseményrendszer esetén a Bayes tétel

$$p(H_i|D,I) = \frac{p(D|H_i,I)p(H_i|I)}{\sum_i p(D|H_i,I)p(H_i|I)}$$

az eseményrendszer valószínűségi eloszlását határozza meg.

Marginalizáció
$\sum_i p(A,H_i|I) = p(A,\sum_i(H_i)|I) = p(A,H|I) = p(A|I)$
Eliminálhatóak az u.n. érdektelen paraméterek.

Felbontás
$P(A|I) = \sum_i p(A|H_i,I)P(H_i|I)$
Összetett esemény valószínűsége egyszerűbb események valószínűségéből határozható meg.

Előző előadás

- Ahhoz, hogy a Bayes valószínűségi elméletet alkalmazzuk a probléma egyértelmű megfogalmazása szükséges.
  - Definiálni kell egy teljes egymást kölcsönösen kizáró eseményrendszert.
  - Meg kell határozni, hogy ez események hogyan határozzák meg a mért adatokat – likelihood – minden paramétert figyelembe kell venni.
  - Meg kell adni a prior valószínűségeket.
  - A Bayes tétel használatával meg kell határozni a poszterior valószínűséget.
  - A számunkra érdektelenek paraméterek marginalizációval eliminálhatóak.
- A rendelkezésre álló információkat egyértelműen kell megadni, hogy azok meghatározzák a valószínűségeket amelyekre szükségünk van.

Paraméterbecslés

Legyen egy H hipotézisünk – egy fizikai modell– amit igaznak tételezünk fel. Legyen a hipotézisnek egy vagy több paramétere $H(\alpha, \beta, ...)$ (összetett hipotézis).

A mérési eredményekből ezekről a paraméterekről akarunk következtetést levonni.

A Bayes–paraméterbecslés nem egy adott értéket szolgáltat a paraméter értékéről, hanem annak valószínűségi eloszlását. A modell paramétere konkrét érték, a paraméter valószínűségi eloszlása jellemzi ismereteink hiányos voltát.

Paraméterbecslés

A modell paramétereit két csoportra osztjuk:
A – ezen paraméterek értékéről akarunk információt szerezni
B – érdektelen paraméterek (nuisance paraméterek) amelyek szükségesek a fizikai modell leírásához, de különben értékei számunkra érdektelenek.
Legyen $p(\mathbf{D}|\mathbf{A}, \mathbf{B}, I)$ a likelihood, $p(\mathbf{A}, \mathbf{B}|I)$ a prior.

Bayes tétel

$$p(\mathbf{A}, \mathbf{B}|\mathbf{D}, I) = \frac{p(\mathbf{D}|\mathbf{A}, \mathbf{B}, I)p(\mathbf{A}, \mathbf{B}|I)}{\iint p(\mathbf{D}|\mathbf{A}, \mathbf{B}, I)p(\mathbf{A}, \mathbf{B}|I)d\mathbf{A}d\mathbf{B}}$$

$$p(\mathbf{A}, \mathbf{B}|\mathbf{D}, I) \propto p(\mathbf{D}|\mathbf{A}, \mathbf{B}, I)p(\mathbf{A}, \mathbf{B}|I)$$

A nevező nem függ a paraméterektől,normálási állandó.

Paraméterbecslés

A számunkra érdekes **A** paraméterek poszterior eloszlását marginalizációval kaphatjuk meg:

$$p(\mathbf{A}|\mathbf{D}, I) = \int p(\mathbf{A}, \mathbf{B}|\mathbf{D}, I) d\mathbf{B}.$$

A fenti egyenletek a Bayes–paraméterbecslést teljesen meghatározzák.
A marginális eloszlások kiszámítása sokszor nem triviális.
Többdimenziós esetben az integrálások (a marginalizációk elvégzése) nagy mennyiségű numerikus számolást igényelnek.
Egyszerűsítések lehetségesek, ha
a mérések függetlenek,
a likelihood Gauss eloszlást követ,
a modell a paraméterekben lineáris.

A poszterior közelítése

Egy paraméter poszterior eloszlása következtetésünket a paraméter értékéről teljesen meghatározza.
Szeretnénk a poszterior eloszlást néhány értékkel jellemezni.
Legyen $p(\alpha|\{d_k\}, I)$, az α paraméter valószínűségi sűrűségfüggvénye.
$\{d_k\}$ jelöli a mért adatokat.

Módusz – a poszterior legvalószínűbb értéke
Módusz – $\alpha_0 = argmax_\alpha(p(\alpha|\{d_k\}, I))$
Átlagértek – $\mu =< \alpha >= \int \alpha p(\alpha|\{d_k\}, I) d\alpha$
Szóras – $\sigma^2 =< (\alpha - \mu)^2 >= \int (\alpha - \mu)^2 p(\alpha|\{d_k\}, I) d\alpha$

Nem minden eloszlásnak van átlagértéke és szórása.

$$p(\alpha) = \frac{1}{\pi} \frac{1}{1 + \alpha^2}$$

Szórás helyett használhatjuk a félértékszélességet .

A poszterior közelítése

Fejtsük Taylor sorba a poszterior eloszlás $L(\alpha) = \ln\left[p(\alpha|\{d_k\}, I)\right]$ logaritmusát az α_0 pont környezetében:

$$L(\alpha) = L(\alpha_0) + \frac{dL(\alpha_0)}{d\alpha}(\alpha - \alpha_0) + \frac{1}{2}\frac{d^2L(\alpha_0)}{d\alpha^2}(\alpha - \alpha_0)^2 + \cdots$$

A módusz az eloszlásfüggvény maximuma. Mivel L monoton függvénye p-nek, így L maximuma p-nek is maximuma

$$\frac{dL(\alpha_0)}{d\alpha} = 0 \quad \rightarrow \alpha_0$$

A log-poszterior exponenciálisát véve

$$p(\alpha|\{d_k\}, I) \propto \exp\left[\frac{1}{2}\frac{d^2L(\alpha_0)}{d\alpha^2}(\alpha - \alpha_0)^2\right].$$

A poszterior közelítése

A poszterior eloszlást Gauss eloszlással közelítettük.

$$p(\alpha|\{d_k\}, I) \propto \exp\left[-\frac{(\alpha-\alpha_0)^2}{2\sigma^2}\right]$$

$$\sigma = \left[-\frac{d^2 L(\alpha_0)}{d\alpha^2}\right]^{-1/2}$$

A σ paraméter az eloszlás szélességével arányos, így a paraméter meghatározásának bizonytalanságát jellemzi.
Következtetésünket az α paraméterről tömören a következőképpen foglalhatjuk össze:

$$\alpha = \alpha_0 \pm \sigma$$

A poszterior közelítése

A fenti közelítés akkor alkalmazható, ha

- a poszterior egy maximummal rendelkezik és
- közelítőleg szimmetrikus és
- gyorsan lecseng.

Más esetekben több adatok kell megadni, úgymint
magasabb momentumokat,
konfidencia intervallumot,

vagy a teljes poszterior eloszlást.

Klasszikus paraméterbecslés

Maximum likelihood becslés.
Legkisebb négyzetek módszere.
A klasszikus paraméterbecslés esetében nem vehetünk figyelembe előzetes információkat.
A Bayes tételben egyenletes priort kell használni.

$$p(\mathbf{A}|\mathbf{D}, I) \propto p(\mathbf{D}|\mathbf{A}, I)$$

A poszterior a likelihood-dal arányos.
Az érdektelen paramétereket nem tudjuk figyelembe venni.

$$\mathcal{L}(\mathbf{A}) = p(\mathbf{D}|\mathbf{A}, I)$$

A likelihood függvény maximalizálásával kaphatjuk az **A** paraméterek legjobb becslését. A Maximum likelihood becslés a Bayes poszterior eloszlás módusza egyenletes priort feltételezve.

Klasszikus paraméterbecslés

Tekintsük a következő fizikai modellt. A várt mérési adatokat a $D = f(x, \mathbf{A})$ függvénnyel lehet jellemezni. Az $\{x_i\}$ mintavételi helyeken a mérési adatok $\mathbf{D} = \{d_i\}$.
Az adatok legyenek leírhatóak a $p_i(d_i|\mathbf{A}, I)$ eloszlással és függetlenek.
A likelihood:

$$\mathcal{L}(\mathbf{A}) = p(\mathbf{D}|\mathbf{A}, I) = \prod_{i=1} p_i(d_i|\mathbf{A}, I)$$

Ha a mérések bizonytalansága Gauss eloszlással írható le:

$$p_i(d_i|\mathbf{A}, I) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(d_i - f(x_i, \mathbf{A}))^2}{2\sigma_i^2}\right]$$

A likelihood függvény logaritmusa a következő:

$$\ln\left[\mathcal{L}(\mathbf{A})\right] \propto -\frac{1}{2}\chi^2 = -\frac{1}{2}\sum_{i=1}\left[\frac{d_i - f(x_i, \mathbf{A})}{\sigma_i}\right]^2$$

Klasszikus paraméterbecslés

A log-likelihood függvénynek ott van maximuma ahol χ^2-nek minimuma, így a paraméterek legjobb becslését a χ^2 minimalizációjából határozhatjuk meg. Ebben az esetben χ^2 minimalizálásról vagy legkisebb négyzetek módszeréről beszélünk.

A maximum likelihood és a legkisebb négyzetek módszere a Bayes–paraméterbecslés speciális esete.
Mindkét esetben egyenletes prior feltételezéssel éltünk.
A legkisebb négyzetek módszerénél a mérési adatok függetlenségét és a mérések zajának Gauss eloszlását is fel kellett használnunk.

Átlagolás Gauss zaj esetén

A normál eloszlást gyakran használjuk a mérési adatok bizonytalanságának jellemzésere.
Legyen $\mathbf{D} = \{d_i\}$ a független mérési adatok halmaza.
$d_i = \mu + \varepsilon_i$, ahol az ε_i zaj σ paraméterrel jellemezhető Gauss eloszlás

$$p(d_i|\mu, \sigma, I) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(d_i - \mu)^2}{2\sigma^2}\right]$$

Határozzuk meg a μ paraméter legjobb becslését, és a becslés hibáját abban az esetben amikor σ értéke ismert.
A poszterior eloszlás a következő

$$p(\mu|\mathbf{D}, \sigma, I) \propto p(\mathbf{D}|\mu, \sigma, I) p(\mu|\sigma, I)$$

Átlagolás Gauss zaj esetén

Ha a mérési adatok függetlenek, a likelihood:

$$p(\mathbf{D}|\mu,\sigma,I) = \prod_{i=1}^{n} p(d_i|\mu,\sigma,I) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\sum_{i=1}^{n}\frac{(d_i-\mu)^2}{2\sigma^2}\right]$$

Legyen egyenletes prior eloszlást, azaz $p(\mu|\sigma,I) = p(\mu) = \mathrm{konstans}$. A poszterior logaritmusa a következő:

$$L = \ln\left[p(\mu|\mathbf{D},\sigma)\right] = \mathrm{konstans} - \sum_{i=1}^{n}\frac{(d_i-\mu)^2}{2\sigma^2}$$

μ legjobb becslését a

$$\frac{dL(\mu_0)}{d\mu} = \sum_{i=1}^{n}\frac{d_i-\mu_0}{\sigma^2} = 0$$

feltétel szolgáltatja.

Átlagolás Gauss zaj esetén

$$\mu_0 = \frac{1}{n}\sum_{i=1}^{n} d_i.$$

A becslés hibáját a L második deriváltjából kaphatjuk meg:

$$\frac{d^2L(\mu_0)}{d\mu^2} = -\sum_{i=1}^{n}\frac{1}{\sigma^2} = -\frac{n}{\sigma^2}$$

A következtetésünket μ-ről tömören összefoglalva:

$$\mu = \mu_0 \pm \frac{\sigma}{\sqrt{n}}$$

A jól ismert eredményt kapjuk, a legjobb becslés a mért adatok átlaga, a becslés hibája a mérések számának négyzetgyökével fordítottan arányos.

Átlagolás Gauss zaj esetén I

$d_i = \mu + \varepsilon_i$, ahol az ε_i zaj σ_i paraméterrel jellemezhető Gauss eloszlás. A mérések hibája nem azonos. Legyen $w_i = 1/\sigma_i^2$

$$\mu_0 = \frac{\sum_{i=1}^{n} w_i d_i.}{\sum_{i=1}^{n} w_i}$$

A következtetésünket μ-ről tömören összefoglalva:

$$\mu = \mu_0 \pm \frac{1}{\sqrt{\sum_{i=1}^{n} w_i}}$$

A súlyozott átlagolás kifejezéseit kaptuk.

Átlagolás Gauss zaj esetén

Átlagolás, ha a mérési hibákat nem ismerjük.
Legyen a mérések hibája (σ) minden adatra azonos, de ismeretlen paraméter.
μ poszterior eloszlását marginalizációval kaphatjuk:

$$p(\mu|\mathbf{D}, I) = \int p(\mu, \sigma|\mathbf{D}, I) d\sigma$$

$$p(\mu, \sigma|\mathbf{D}, I) \propto p(\mathbf{D}|\mu, \sigma, I) p(\mu, \sigma|I)$$

Prior:

$$p(\mu, \sigma|I) = \begin{cases} \text{konstans} & \text{ha} \quad \sigma > 0 \\ 0 & \text{máshol} \end{cases} .$$

Átlagolás Gauss zaj esetén

$$p(\mu|\mathbf{D}, I) \propto \left[\sum_{i=1}^{n}(d_i - \mu^2)\right]^{-(n-1)/2}$$

$$\frac{dL(\mu_0)}{d\mu} = \frac{(n-1)\sum(d_i - \mu_0)}{\sum(d_i - \mu_0)^2} = 0$$

$$\mu_0 = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$$\frac{d^2L(\mu_0)}{d\mu^2} = -\frac{(n-1)n}{\sum(d_i - \mu_0)^2}$$

$$\mu = \mu_0 \pm \frac{S}{\sqrt{n}} \qquad S = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \mu_0)^2}{n-1}}$$

Átlagolás Gauss zaj esetén

Marginalizációval meghatározhatjuk a Gauss folyamat σ szórásának poszterior eloszlását is:

$$p(\sigma|\mathbf{D}, I) = \int p(\mu, \sigma|\mathbf{D}, I) d\mu$$

A marginalizáció eredménye:

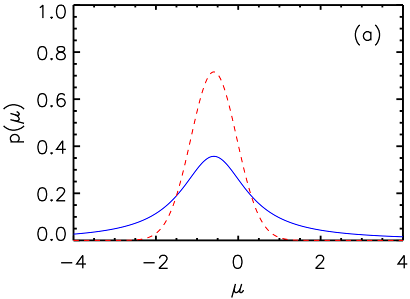
$$p(\sigma|\mathbf{D}, I) \propto \sigma^{1-n} \exp\left(-\frac{\sum_{i=1}^{n}(d_i - \mu_0)^2}{2\sigma^2}\right)$$

Ezen kifejezés felhasználásával megkaphatjuk σ legjobb becslését és hibáját:

$$\sigma = \sigma_0 \pm \frac{\sigma_0}{\sqrt{2(n-1)}} \quad \text{ahol} \quad \sigma_0 = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \mu_0)^2}{(n-1)}}$$

Átlagolás Gauss zaj esetén

Az $n = 4$ mérési adatot véletlenszerűen generáltuk nulla várható értékkel és egységnyi szórással.



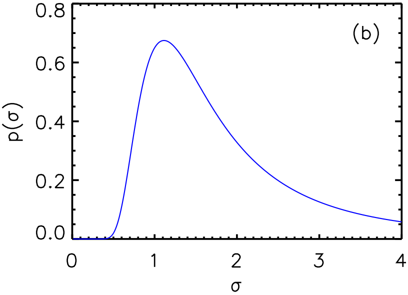
kék görbe – μ marginális eloszlása
piros görbe – a marginális eloszlás közelítése Gauss eloszlással

$$p(\mu|\mathbf{D}, I) \propto \left[\sum_{i=1}^{n}(d_i - \mu^2)\right]^{-(n-1)/2}$$

Az átlagérték meghatározásának bizonytalanságát növeli, ha az adatok hibáját a mérésből kell meghatározni.

Átlagolás Gauss zaj esetén

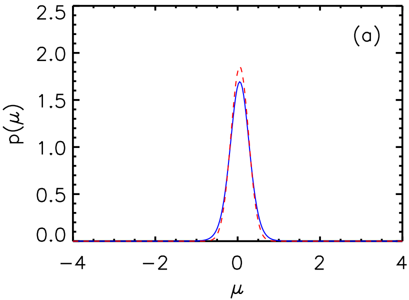
Az előző példában a szórás poszterior eloszlása



$$p(\sigma|\mathbf{D}, I) \propto \sigma^{1-n} \exp\left(-\frac{\sum_{i=1}^{n}(d_i - \mu_0)^2}{2\sigma^2}\right)$$

Átlagolás Gauss zaj esetén

Az $n = 16$ mérési adatot véletlenszerűen generáltuk nulla várható értékkel és egységnyi szórással.



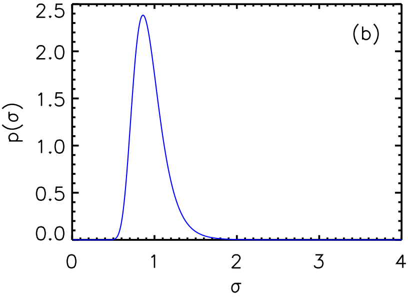
kék görbe – μ marginális eloszlása
piros görbe – a marginális eloszlás közelítése Gauss eloszlással

$$p(\mu|\mathbf{D}, I) \propto \left[\sum_{i=1}^{n}(d_i - \mu^2)\right]^{-(n-1)/2}$$

A marginális eloszlás Gauss eloszláshoz közelít.

Átlagolás Gauss zaj esetén

Az $n = 16$ mérési adatot véletlenszerűen generáltuk nulla várható értékkel és egységnyi szórással.



$$p(\sigma|\mathbf{D}, I) \propto \sigma^{1-n} \exp\left(-\frac{\sum_{i=1}^{n}(d_i - \mu_0)^2}{2\sigma^2}\right)$$

négyszeres adatmennyiség mintegy felére csökkentette a poszterior eloszlások szélességét (a becslés hatékonysága $\propto 1/\sqrt{n}$).