Bayes–adatfeldolgozás III

Kálvin Sándor

2014 november 12

Prior eloszlás meghatározása – szimmetria megfontolás

Legyen λ egy spektrumban a csúcs helye, és ne legyen előzetes ismeretünk ezen paraméterről, azaz egy $[a, b]$ intervallumban bárhol elhelyezkedhet.
Ebben az esetben, ha a koordináta-rendszer kezdőpontját λ_0-val eltoljuk, akkor ez nem változtathatja meg a csúcs helyének prior valószínűségét, azaz

$$p(\lambda|I)d\lambda = p(\lambda + \lambda_0|I)d(\lambda + \lambda_0)$$

Mivel λ_0 konstans $d(\lambda + \lambda_0) = d(\lambda)$, amiből következik, hogy

$$p(\lambda|I) = \text{konstans}$$

Azaz hiányos ismeretünket egy helyparaméterről egyenletes prior eloszlással írhatjuk le.
A prior valószínűségi sűrűségfüggvény, így normáltnak kell lennie.

$$\int_a^b p(\lambda|I) = 1 \quad \rightarrow \quad p(\lambda|I) = \frac{1}{b-a}$$

Prior eloszlás meghatározása – szimmetria megfontolás

Hasonlóan, szeretnénk meghatározni a csúcs A amplitúdójának prior valószínűségi eloszlását, ha nem ismerjük az amplitúdó nagyságrendjét.
Tekintsük a következő esetet. A csúcs amplitúdóját az [1,1000] intervallumban várjuk. Használjunk egyenletes priort azaz $p(A|I) = 1/999$. Annak a valószínűsége, hogy A 1 és 10 közé esik $p(1 \leq A \leq 10|I) = \int_1^{10} p(A|I) = 9/999$.
Annak a valószínűsége, hogy A 100 és 1000 közé esik $p(100 \leq A \leq 1000|I) = \int_{100}^{1000} p(A|I) = 899/999$
Azaz annak a prior valószínűsége, hogy az amplitúdó a felső dekádba esik százszor valószínűbb mint, hogy az alsó dekádba. Ebben az esetben az egyenletes prior a nagyobb amplitúdókat preferálja, elvárásunkkal ellentétben.

Szimmetria megfontolás

Keressünk olyan valószínűségi eloszlását ami dekádonként azonos valószínűségeket szolgáltat.
Az amplitúdó skálájának megváltoztatása (α-val való szorzása), nem változtathatja meg az amplitúdó prior valószínűségét:

$$p(A|I)dA = p(\alpha A|I)d(\alpha A) = \alpha p(\alpha A|I)dA$$

Ezen feltétel abban az esetben teljesűl, ha $p(A|I) \propto 1/A$.
Figyelembe véve, hogy az eloszlásfüggvény normált

$$\int_a^b p(A|I)dA = 1 \quad \rightarrow \quad p(A|I) = \frac{1}{A\ln(b/a)}$$

A pozitív skálaparaméterekre alkalmazható prior valószínűségi hozzárendelést Jeffreys priornak nevezzük.
(A prior intervallum alsó határa nem lehet nulla.)

Csúcs helyének és amplitúdójának meghatározása

A (plazma)fizikában gyakran felmerülő feladat, hogy meg kell határozni zajjal terhelt mérésekből egy csúcs helyét, amplitúdóját, helyét, a hátteret (alakját, nagyságát) ...

Tekintsük a következő egyszerű feladatot. A hátteret levontuk és ismerjük a csúcs alakját.
Legyen a jel (D) adott w szélességű Gauss csúcs, azaz

$$D(x, A, x_0) = A \exp\left[-\frac{(x-x_0)^2}{2w^2}\right],$$

ahol x a mérési változó,
A és x_0 a meghatározandó paraméterek, a csúcs amplitúdója és helye.
Méréseinket végezzük az $\{x_k\}$ helyeken, a mérési adatok legyenek $\{d_k\}$.

Csúcs helyének és amplitúdójának meghatározása

A Bayes tétel alkalmazásához meg kell adni a likelihood eloszlást és a meghatározandó paraméterek prior valószínűségi eloszlását.

A likelihood meghatározása

Az ideális (zaj nélküli) mérési értékek:

$$D_k(x_k, A, x_0) = A \exp\left[-\frac{(x_k - x_0)^2}{2w^2}\right]$$

Zajjal terhelt mérés esetén $d_k = D_k + \epsilon$. Tételezzük fel, hogy az ϵ mérési zaj Gauss eloszlású σ szórással és 0 várható értékkel.

$$p(d_k|A, x_0, I) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(d_k - D_k)^2}{2\sigma^2}\right].$$

Csúcs helyének és amplitúdójának meghatározása

Ha a mérések függetlenek akkor a likelihood függvényt az egyes mérések valószínűségeinek szorzata adja:

$$p(\{d_k\}|A, x_0, I) = \prod_{k=1}^{n} p(d_k|A, x_0, I)$$

likelihood

$$p(\{d_k\}|A, x_0, I) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\sum_{k=1}^{n} \frac{(d_k - D_k)^2}{2\sigma^2}\right]$$

A prior eloszlás meghatározása

$$p(A, x_0|I) = p(A|x_0, I)p(x_0|I)$$

Ha a csúcs amplitúdója független a helyétől, akkor

$$p(A, x_0|I) = p(A|I)p(x_0|I)$$

Csúcs helyének és amplitúdójának meghatározása

A csúcs pozíciója mivel helyparaméter, prior eloszlása legyen egyenletes.

$$p(x_0|I) = \begin{cases} \frac{1}{x_2 - x_1} & \text{ha} \quad x_1 < x_0 < x_2 \\ 0 & \text{máshol} \end{cases}$$

Az amplitúdó prior eloszlására két lehetőséget fogunk vizsgálni. Az elő esetben használjunk egyenletes priort.

$$p(A|I) = \begin{cases} \frac{1}{A_1 - A_1} & \text{ha} \quad A_1 < A < A_2 \\ 0 & \text{máshol} \end{cases}$$

A másik esetben vegyük figyelembe, hogy az amplitúdó pozitív skálaparaméter. Alkalmazzunk Jeffreys prior.

$$p(A|I) = \begin{cases} \frac{1}{A \ln(A_2/A_1)} & \text{ha} \quad A_1 < A < A_2 \\ 0 & \text{máshol} \end{cases}$$

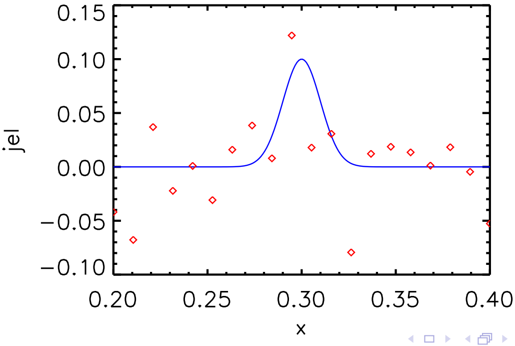
Csúcs helyének és amplitúdójának meghatározása

Bayes-tétel szerint következtetésünket a meghatározandó paraméterekről a poszterior eloszlás írja le.

$$p(A, x_0 | \{d_k\}, I) \propto p(\{d_k\} | A, x_0, I) p(A, x_0 | I)$$

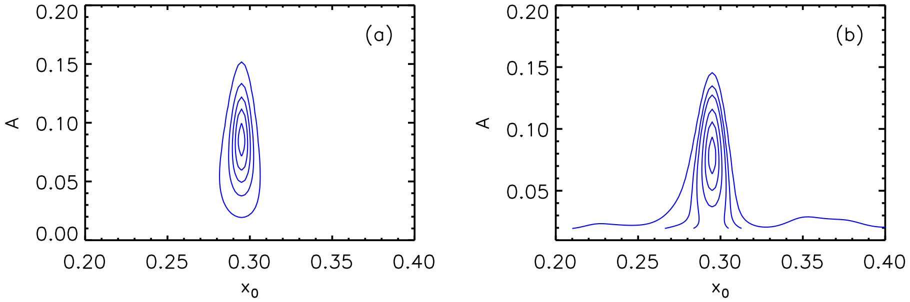
Csúcs helyének és amplitúdójának meghatározása

A kiértékelést az alábbi példán szemléltetjük. Véletlenszerűen generáltunk mérési adatokat amelyek paraméterei a következőek
a mérési adatok száma 20
amplitúdó $A = 0,1$
hely $x_0 = 0,3$
zaj szórása $\sigma = 0,04$.



Csúcs helyének és amplitúdójának meghatározása

(a mérési adatok száma 20 , zaj szórása $\sigma = 0,04$)
A mérési adatokból meghatározott $p(A, x_0|\{d_k\}, I)$ poszterior eloszlások



(a) A $p(A, x_0|\{d_k\}, I)$ poszterior eloszlás egyenletes prior esetén.
(b) A $p(A, x_0|\{d_k\}, I)$ poszterior eloszlás Jeffreys prior esetén.

Csúcs helyének és amplitúdójának meghatározása

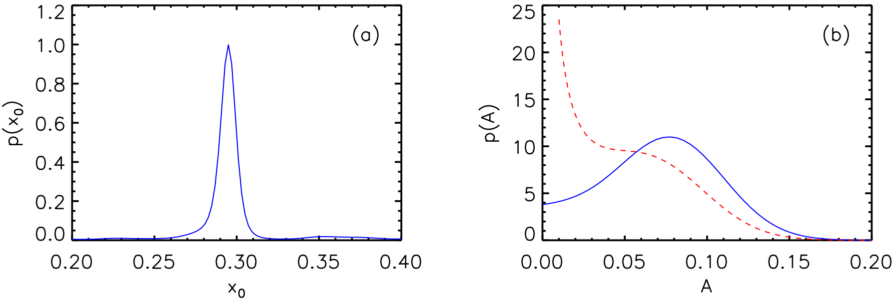
Az $p(A, x_0|\{d_k\}, I)$ poszterior eloszlás teljesen leírja következtetésünket a csúcs helyéről és amplitúdójáról.
Az amplitúdó és hely marginális eloszlásának meghatározása.

$$p(A|\{d_k\}, I) = \int p(A, x_0|\{d_k\}, I) dx_0 \tag{1}$$

$$p(x_0|\{d_k\}, I) = \int p(A, x_0|\{d_k\}, I) dA \tag{2}$$

Csúcs helyének és amplitúdójának meghatározása

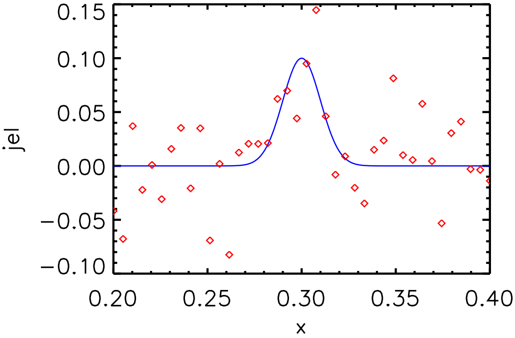
(a mérési adatok száma 20 , zaj szórása $\sigma = 0,04$)



(a) A csúcs helyének $p(x_0|\{d_k\}, I)$ marginális eloszlása.
(b) A csúcs amplitúdójának $p(A|\{d_k\}, I)$ marginális eloszlása.
folytonos vonal: egyenletes prior – szaggatott vonal: Jeffreys prior

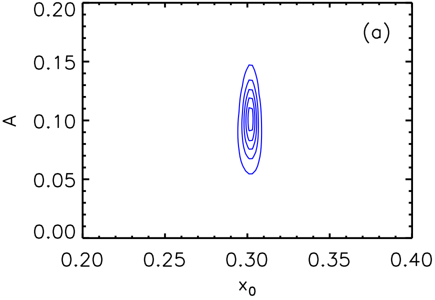
Csúcs helyének és amplitúdójának meghatározása

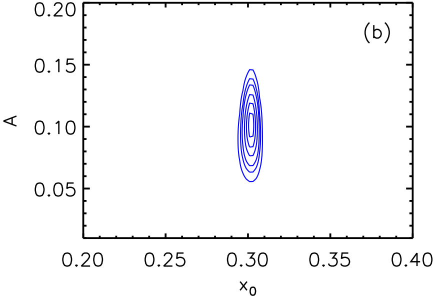
a mérési adatok száma 40
amplitúdó $A = 0,1$
hely $x_0 = 0,3$
zaj szórása $\sigma = 0,04$.



Csúcs helyének és amplitúdójának meghatározása

(a mérési adatok száma 40 , zaj szórása $\sigma = 0,04$)
A mérési adatokból meghatározott $p(A, x_0|\{d_k\}, I)$ poszterior eloszlások

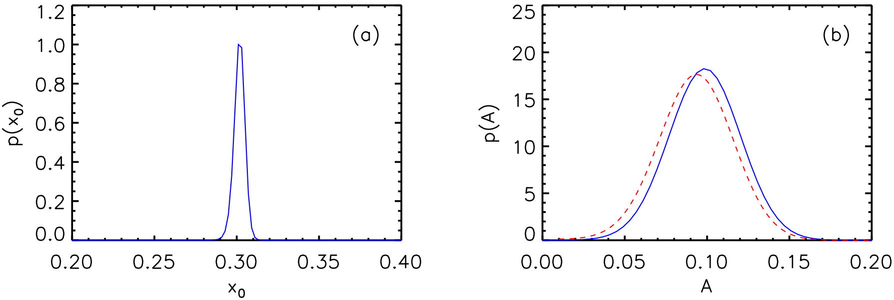




(a) A $p(A, x_0|\{d_k\}, I)$ poszterior eloszlás egyenletes prior esetén.
(b) A $p(A, x_0|\{d_k\}, I)$ poszterior eloszlás Jeffreys prior esetén.

Csúcs helyének és amplitúdójának meghatározása

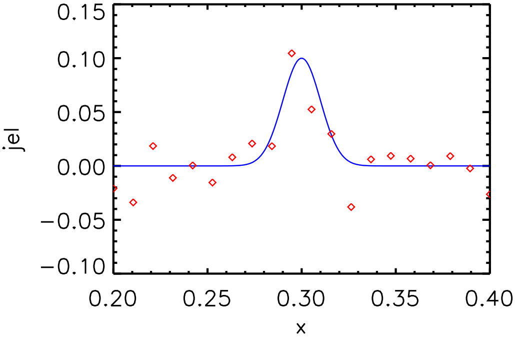
(a mérési adatok száma 40 , zaj szórása $\sigma = 0,04$)



(a) A csúcs helyének $p(x_0|\{d_k\}, I)$ marginális eloszlása.
(b) A csúcs amplitúdójának $p(A|\{d_k\}, I)$ marginális eloszlása.
folytonos vonal: egyenletes prior – szaggatott vonal: Jeffreys prior

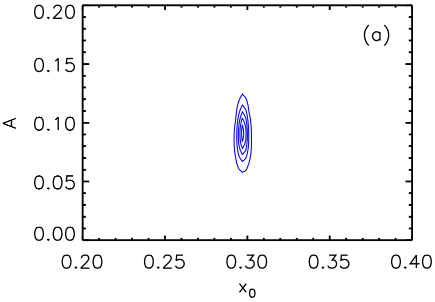
Csúcs helyének és amplitúdójának meghatározása

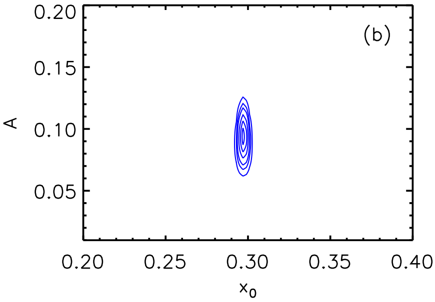
a mérési adatok száma 20
amplitúdó $A = 0,1$
hely $x_0 = 0,3$
zaj szórása $\sigma = 0,02$.



Csúcs helyének és amplitúdójának meghatározása

(a mérési adatok száma 20 , zaj szórása $\sigma = 0,02$)
A mérési adatokból meghatározott $p(A, x_0|\{d_k\}, I)$ poszterior eloszlások

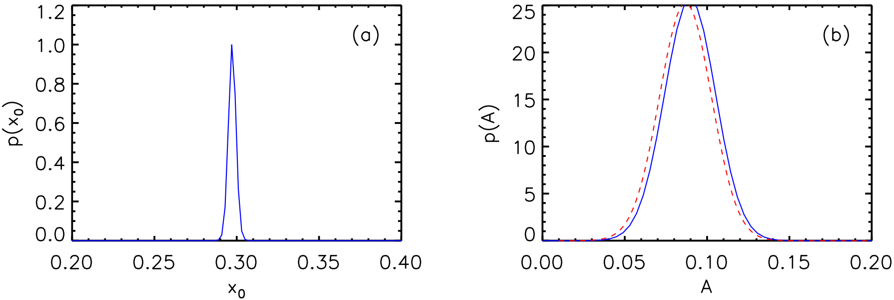




(a) A $p(A, x_0|\{d_k\}, I)$ poszterior eloszlás egyenletes prior esetén.
(b) A $p(A, x_0|\{d_k\}, I)$ poszterior eloszlás Jeffreys prior esetén.

Csúcs helyének és amplitúdójának meghatározása

(a mérési adatok száma 20 , zaj szórása $\sigma = 0,02$)



(a) A csúcs helyének $p(x_0|\{d_k\}, I)$ marginális eloszlása.
(b) A csúcs amplitúdójának $p(A|\{d_k\}, I)$ marginális eloszlása.
folytonos vonal: egyenletes prior – szaggatott vonal: Jeffreys prior

Csúcs helyének és amplitúdójának meghatározása

Az amplitúdó valószínűsége negatív értékek esetén nulla, az alkalmazott priornak megfelelően.
Megfelelően választott prior eloszlással a fizikailag lehetetlen tartományok kizárhatóak.
Mivel az amplitúdó pozitív skálaparaméter, a Jeffreys prior alkalmazása reálisabb eredményt szolgáltat. Az amplitúdó ebben az esetben kis értékekre valószínűbb, mint egyenletes prior alkalmazása esetén.
Több mérési adat esetén, vagy ha a mérések hibája csökken a poszterior eloszlás egyre inkább független a prior megválasztásától.

Valószínűségek meghatározása

Bayes-tétel

$$p(A|B,I) = \frac{p(B|A,I)p(A|I)}{p(B|I)}$$

Bayes elmélet kapcsolatot teremt a valószínűségek között.
A likelihood a fizikai modell és a mérési eljárás ismeretében meghatározható, deduktív következtetés.
A prior valószínűség meghatározásához a rendelkezésünkre álló előzetes információkat kell figyelembe venni.

Általános elv, hogy olyan valószínűségi eloszlást kell meghatározni, ami leginkább kifejezi hiányos ismereteinket az adott eseményekkel, paraméterrel kapcsolatban.

Az elégtelen ok elve(EOE)

Legyen egy N elemű teljes egymást kölcsönösen kizáró H_i eseményrendszer. Ha egyik esemény sem kitüntetett akkor mindegyikhez ugyanolyan valószínűséget rendelek.
Az eseményrendszer normált, azaz

$$\sum_{1}^{N} p(H_i|I) = 1$$

$$\rightarrow \qquad p(H_i|I) = \frac{1}{N}$$

Szabályos kocka esetén annak a valószínűsége, hogy 6-ost dobok $1/6$.
Szabályos érme esetén az írás valószínűsége $1/2$.

Az elégtelen ok elve

Tekintsük a következő problémát.
Egy dobozban legyen P piros és N-P kék golyó.
Véletlenszerűen húzunk egy golyót. Mi a valószínűsége, hogy elsőre pirosat húzunk?

$$\text{Az jól ismert válasz:} \qquad p(P_1|I) = \frac{P}{N}$$

Hogyan számíthatjuk ki ezt a valószínűséget a EOE és a valószínűségi kalkulus segítségével.

Az elégtelen ok elve

Címkézzük meg a golyókat és definiáljunk egy teljes EKK eseményrendszert.
H_i = 'Ezen a golyón i címke van'

Használjuk a EOE a valószínűség meghatározására.

$$p(H_i|I) = \frac{1}{N}$$

Definiáljunk egy teljes EKK eseményrendszert.

H_P = 'Ez a golyó piros'
H_K = 'Ez a golyó kék'

Az elégtelen ok elve

Használjuk a felbontás szabályt

$$p(H_P|I) = \sum_{i=1}^{N} p(H_P, H_i|I) = \sum_{i=1}^{N} p(H_P|H_i, I)p(H_i|I)$$

Használjuk ki, hogy

$$p(H_P|H_i, I) = \begin{cases} 1 & \text{ha az } i. \text{ golyó piros} \\ 0 & \text{különben} \end{cases}$$

$$p(P_1|I) = p(H_P|I) = \frac{1}{N} \sum_{i=1}^{N} p(H_P|H_i, I) = \frac{P}{N}$$

(A felbontás–szabály jól használható valószínűségek meghatározására.)

Az elégtelen ok elve

Mi a valószínűsége, hogy a második golyó amit húzunk piros, ha visszatesszük a golyót? Mivel a doboz tartalma nem változott

$$p(P_2|P_1, I) = \frac{P}{N}$$

Mi a valószínűsége, hogy a második golyó amit húzunk piros, ha nem tesszük vissza a golyót, de tudjuk, hogy az első húzás eredménye piros? A doboz tartalma megváltozott, így

$$p(P_2|P_1, I) = \frac{P-1}{N-1}$$

$$p(P_2|K_1, I) = \frac{P}{N-1}$$

azaz a második húzás valószínűsége függ az elő húzás eredményétől.

Az elégtelen ok elve

Mi valószínűsége, hogy a második golyó amit húzunk piros, ha nem tesszük vissza a golyót, de nem nézzük meg, hogy mi az első húzás eredménye?
A felbontás szabályt alkalmazva

$$p(P_2|I) = p(P_2, P_1|I) + p(P_2, K_1|I)$$

$$= p(P_2|P_1, I)p(P_1|I) + p(P_2|K_1, I)p(K_1|I)$$

$$= \frac{P-1}{N-1}\frac{P}{N} + \frac{P}{N-1}\frac{K}{N} = \frac{P}{N}$$

Az elégtelen ok elve

Nem tesszük vissza az első golyót, nem nézzük meg, hogy mi az első húzás eredménye. A második húzásnak az eredménye piros. Mi annak a valószínűsége, hogy az elő húzás piros volt.
A klasszikus elméletben P/N mivel $p(P_1|I)$ a P_1 esemény tulajdonsága.
Használjuk a Bayes tételt.

$$p(P_1|P_2I) = \frac{p(P_2|P_1I)p(P_1|I)}{p(P_2|P_1I)p(P_1|I) + p(P_2|K_1I)p(K_1|I)}$$

$$= \frac{P-1}{N-1}$$

A Bayes elméletben a valószínűség nem okozati hanem logikai kapcsolat.

Az elégtelen ok elve

Az előző példák megmutatják hogyan lehet a valószínűségeket mintavételi eseményekre meghatározni.
Sok eloszlás egyértelműen meghatározható ezzel a módszerrel.
(Diszkrét eloszlások – Binomiális – Multimodális Folytonos eloszlások is – Poission – a binomiális eloszlás folytonos határesete)
Deduktív következtetés – matematika
Ha nincs elég információnk más elvek alapján kell a valószínűségeket meghatározni.

Olyan valószínűségi eloszlást kell megadni ami leginkább kifejezi hiányos ismereteinket.

Maximum entrópia elve

Hogyan határozhatjuk meg a valószínűségi eloszlásokat, ha az eloszlásról csak bizonyos megkötéseink vannak.
Ha tudjuk, hogy a kocka szabályos akkor $1/6$ annak a valószínűsége, hogy hatost dobunk, stb. Hogyan határozzuk meg a különböző dobások valószínűségét, ha csak annyit tudunk, hogy a dobások átlaga 4. (Egy szabályos kocka esetén az átlag 3.5) Végtelen sok valószínűségi eloszlás lehetséges.
Maximum entrópia elve (Jaynes) szerint a legkevesebb információt tartalmazó valószínűség eloszlást az eloszlás (információs) entrópiájának maximalizálásával határozzuk meg figyelembe véve a valószínűségi eloszlásra vonatkozó kikötéseket.

Maximum entrópia elve

Egy valószínűségi eloszlás információs entrópiájának definíciója

$$S(\{p_i\}) = -\sum_{i=1} p_i \ln(p_i/m_i)$$

Folytonos esetben

$$S(p(a|I)) = -\int p(a|I) \ln[p(a|I)/m(a)] da$$

$$\sum_i p_i = 1 \qquad \int m(a) = 1$$

m_i és $m(a)$ Lebesque mérték. Mivel ezek a fázistér ill. a paraméter skálájának változtatásakor a valószínűségekkel egyformán transzformálódnak az entrópia független lesz a változók megválasztásától.

Maximum entrópia elve – Példa I

Határozzuk meg a valószínűség eloszlást ha semmilyen információval sem rendelkezünk. Csak valószínűségekre alkalmazandó normálási feltételt használhatjuk ki,

$$\sum_i p_i = 1$$

Lagrange szorzók módszerével oldjuk meg az optimalizációt.

$$W = -\sum_i p_i \ln(p_i/m_i) + \lambda(1 - \sum_i p_i)$$

$$\frac{\partial W}{\partial p_j} = -1 - \ln(p_j/m_i) - \lambda = 0$$

minden j-re.

$$\frac{\partial W}{\partial \lambda} = 1 - \sum_i p_i = 0$$

$$p_j = m_i \exp[-(1+\lambda)]$$

Maximum entrópia elve – Példa I

$$\sum p_i = \sum_i m_i \exp[-(1+\lambda)] = \exp[-(1+\lambda)] = 1$$

$$p_i = m_i$$

Folytonos eseményrendszer esetén

$$p(x) = m(x) \tag{3}$$

Ezek alapján m_i és $m(x)$ a legkevesebb információval rendelkező valószínűség eloszlás, ha a valószínűségekről semmilyen információnk nincs. Meghatározásához szimmetria megfontolásokat vagy egyéb racionális megfontolásokat vehetünk figyelembe.

Maximum entrópia elve – Példa II

Az eloszlás szórásnégyzete véges. Legyen $m_i = C$.

$$\sum_i p_i = 1 \qquad \text{és} \qquad \sum_i (x_i - \mu)^2 p_i = \sigma^2$$

$$W = -\sum_{i=1} p_i \ln(p_i) + \lambda_0 (1 - \sum_i p_i) + \lambda_1 (\sigma^2 - \sum_i (x_i - \mu)^2 p_i)$$

λ_0 és λ_1 ismeretlen Lagrange szorzók

$$\frac{\partial W}{\partial p_j} = -1 - \ln(p_j) - \lambda_0 - \lambda_1 (x_j - \mu)^2 = 0$$

$$p_j = e^{-(1+\lambda_0)} e^{-\lambda_1 (x_j - \mu)^2}$$

Maximum entrópia elve – Példa II

Folytonos esetet tekintve

$$p(x) = e^{-(1+\lambda_0)} e^{-\lambda_1 (x-\mu)^2}$$

A normalizációt figyelembe véve

$$p(x) = \sqrt{\frac{\lambda_1}{\pi}} e^{-\lambda_1 (x-\mu)^2}$$

A második feltételt felhasználva

$$\sigma^2 = \sqrt{\frac{\lambda_1}{\pi}} \int (x-\mu)^2 e^{-\lambda_1 (x-\mu)^2} = \frac{1}{2\lambda_1}$$

Folytonos eseményrendszer esetén Gauss valószínűségi eloszlást kapunk:

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

Maximum entrópia elve – Példa II

Abban az esetben, ha csak az az információ áll rendelkezésünkre, hogy az eloszlás jellemezhető átlagértékkel és véges szórásnégyzettel akkor a normális eloszlás az amely legjobban kifejezi tudatlanságunkat az adott változóról. Ezen eredmény indokolja a mérési adatok bizonytalanságának Gauss eloszlással való leírását, ha a zaj forrásait nem ismerjük.

Maximum entrópia elve – Példa III

Az eloszlás várható értéke adott.

$$\sum_i p_i = 1 \qquad \text{és} \qquad \sum_i x_i p_i = \mu$$

$$W = -\sum_{i=1} p_i \ln(p_i/m_i) + \lambda_0(1 - \sum_i p_i) + \lambda_1(\mu - \sum_i x_i p_i)$$

λ_0 és λ_1 ismeretlen Lagrange szorzók

$$\frac{\partial W}{\partial p_j} = -1 - \ln(p_j/m_i) - \lambda_0 - \lambda_1 x_j = 0$$

$$p_j = m_i e^{-(1+\lambda_0)} e^{-\lambda_1 x_j}$$

Maximum entrópia elve – Példa III

A feltételeket kihasználva

$$\sum_i p_i = 1 = e^{-(1+\lambda_0)} \sum_i m_i e^{-\lambda_1 x_i}$$

$$e^{-(1+\lambda_0)} = \frac{1}{\sum_i m_i e^{-\lambda_1 x_i}}$$

$$\sum_i x_i p_i = \mu = \frac{\sum_i x_i m_i e^{-\lambda_1 x_i}}{\sum_i m_i e^{-\lambda_1 x_i}}$$

$$\mu \sum_i m_i e^{-\lambda_1 x_i} = \sum_i x_i m_i e^{-\lambda_1 x_i}$$

Adott μ esetén a fenti nemlineáris egyenlet λ_1-re megoldható és a p_i valószínűségek meghatározhatóak.

Maximum entrópia elve – Példa III

Határozzuk meg kockadobás esetén a valószínűségeket, ha csak annyit tudunk, hogy a dobások átlaga 4.
Ebben az esetben μ=4 , x_i =i, m_i=1/6. A fenti kifejezéseket felhasználva meghatározhatjuk a valószínűségeket.